Gender-specific Performance of a Diagnostic Score in Acute Appendicitis

JANNICA MEKLIN1*, MAARET ESKELINEN1*, KARI SYRJÄNEN2,3 and MATTI ESKELINEN1

1Department of Surgery, Kuopio University Hospital and School of Medicine, University of Eastern Finland, Kuopio, Finland;
2Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, Brazil;
3SMW Consultants, Ltd., Kaarina, Finland

Abstract. Background/Aim: Although a negative appendectomy in female patients with acute abdominal pain (AAP) can be twice as frequent as in male patients, the accuracy of diagnostic scores (DSs) in acute appendicitis (AA) is rarely considered among patients with AAP. The aim was to study the gender-specific performance of a DS in AA. Patients and Methods: As an extension of the World Organisation of Gastro-Enterology Research Committee (OMGE) AAP study, 1,333 patients presenting with AAP were included in the study. The clinical history and diagnostic symptoms (n=22), signs (n=14) and laboratory tests (n=3) were recorded in each patient. Results: The most significant diagnostic predictors were used to construct DS formulas for AA diagnosis, separately for both genders. The formulas were tested at 6 different cut-off levels to find the best diagnostic performance for AA in females and males. The highest specificities of the DS LC– [DS without leucocyte count (LC)] and DS LC+ (DS with LC) scores in detecting AA were 98% (95% CI=97-99%) and 98% (95% CI=96-99%), respectively. In the ROC comparison test, there was no statistically significant difference in the performance of DS LC– and DS LC+ in female and male patients. Conclusion: Our gender-specific DS reached very high AUC values for AA (0.948-0.956) in both genders, and there was no statistically significant difference in the AUC values of DS LC– and DS LC+ between women and men with AAP.

This article is freely accessible online.

*These Authors contributed equally to this study.

Correspondence to: Matti Eskelinen, MD, Ph.D., School of Medicine, University of Eastern Finland, P.O. Box 100, FI-70029 KYS, Finland. Tel: +358 17173311, Fax: +358 17172611, GSM: +358 400969444, e-mail: matti.eskelinen@kuh.fi

Key Words: Acute appendicitis, acute abdominal pain, diagnostic score, female, male, leucocyte count, ROC, HSROC, diagnostic accuracy.
Identifying the DS models. In the computation of the diagnostic score (DS), a multivariate logistic (stepwise) regression analysis (SPSS Statistics 26.0.0.1; IBM, NY, USA) was used to disclose the variables with an independent predictive value. All the variables presented in Tables I and II were included in the analysis as binary data, e.g. AA=1 and other diagnosis of AAP=0. Using the coefficients of the regression model, a DS was built and its predictive value for AA was studied. The coefficient of the multivariate analysis shows the relative risk (RR=e^β, n=β) of a patient with a given symptom or sign to have an AA.

The formula with LC (DS_LC+) in women. The formula with LC, showing the highest diagnostic performance for AA in women is as follows DS_LC+ =3.17×tenderness (positive endpoint=1, negative endpoint=0)+2.39×rigidity (positive endpoint=1, negative endpoint=0)+2.00×guarding (positive endpoint=1, negative endpoint=0)+1.63×LC+1.45×pain at diagnosis (positive endpoint=1, negative endpoint=0)+0.77×renal tenderness (positive endpoint=1, negative endpoint=0)+7.80. The mean (SD) of DS_LC+ values for AA in women (n=111) were 1.32 (1.81) and DS_LC+ mean (SD) values for all women with AAP (n=575) were −3.10 (3.21) (Table IV).

The formula without LC (DS_LC−) in women. The formula without LC, showing the highest diagnostic performance for AA in women is as follows DS_LC− =2.98×tenderness (positive endpoint=1, negative endpoint=0)+2.45×rigidity (positive endpoint=1, negative endpoint=0)+2.08×guarding (positive endpoint=1, negative endpoint=0)+1.33×pain at diagnosis (positive endpoint=1, negative endpoint=0)+0.88×renal tenderness (positive endpoint=1, negative endpoint=0)+7.22. The mean (SD) of DS_LC− values for AA in women (n=121) were 0.61 (1.81) and DS_LC− mean (SD) values for all female patients with AAP (n=697) were −3.41 (2.84) (Table III).

Table I. The clinical history of the patients with acute appendicitis versus other diagnosis of acute abdominal pain.

| Clinical history variable | Positive endpoint | Negative endpoint | TP Gender | FN Gender | FP Gender | TN Gender |
|--------------------------|------------------|------------------|-----------|-----------|-----------|-----------|
|                          |                  |                  | F  M      | F  M      | F  M      | F  M      |
| 1. Location of initial | Upper quadrant | Other quadrants | 4  2      | 117 147   | 94 52     | 482 435   |
| pain of abdomen         | of abdomen       | of abdomen       |           |           |           |           |
| 2. Location of pain     | Right lower     | Other quadrants  | 94 113    | 27 36     | 156 93    | 420 394   |
| at diagnosis            | quadrant of     | of abdomen       |           |           |           |           |
|                          | abdomen          |                 |           |           |           |           |
| 3. Intensity of         | Subjectively moderate/ | Weak pain | 77 102    | 44 47     | 381 313   | 195 174   |
| abdominal pain          | intolerable pain |                 |           |           |           |           |
| 4. Progression of pain  | Subjectively same | Weaker pain than | 86 112    | 35 37     | 388 302   | 188 185   |
| from onset to           | or worse pain    | at the onset     |           |           |           |           |
| diagnosis               |                  |                 |           |           |           |           |
| 5. Type of pain         | Subjectively steady pain | Colicky or intermittent pain | 109 81    | 40 40     | 256 285   | 231 291   |
| 6. Aggravating factors  | Movement, coughing, | No aggravating factors | 112 132   | 9 17      | 409 324   | 167 163   |
|                          | respiration, food or other | |           |           |           |           |
| 7. Relieving factors    | No relieving factors | Vomiting, lying still, food, antacids or no relieving factors | 24 28     | 97 121    | 192 190   | 384 297   |
| 8. Previous similar pain | No               | Yes              | 103 126   | 16 19     | 335 308   | 236 176   |
| 9. Vertigo              | No               | Yes              | 121 149   | 0 0       | 547 472   | 28 12     |
| 10. Nausea              | Yes              | No               | 74 80     | 47 69     | 345 272   | 231 215   |
| 11. Vomiting            | Yes              | No               | 58 71     | 63 78     | 230 216   | 346 271   |
| 12. Appetite            | No appetite      | Normal appetite  | 100 119   | 21 30     | 421 337   | 155 150   |
| 13. Previous indigestion| No               | Yes              | 111 129   | 10 20     | 444 368   | 131 118   |
| 14. Jaundice            | No               | Yes              | 121 148   | 0 1       | 557 476   | 19 11     |
| 15. Bowels              | Normal           | Constipation, diarrhea, blood, mucus, white or normal stools | 102 124   | 19 25     | 434 355   | 142 132   |
| 16. Micturition         | Normal           | Abnormal         | 115 148   | 6 1       | 535 449   | 41 38     |
| 17. Drugs for abdominal | No               | Yes              | 120 148   | 1 1       | 552 458   | 24 28     |
| pain                    |                  |                 |           |           |           |           |
| 18. Previous abdominal | No               | Yes              | 101 146   | 20 3      | 584 368   | 192 118   |
| surgery                 |                  |                 |           |           |           |           |
| 19. Previous abdominal  | No               | Yes              | 111 138   | 10 11     | 463 387   | 113 99    |
| diseases                |                  |                 |           |           |           |           |
| 20. Use of alcohol      | No               | Yes              | 121 149   | 0 0       | 567 463   | 9 23      |

TP, true positive; FN, false negative; FP, false positive; TN, true negative.
The formula without LC ($D_{SC-}$) in men. The formula without LC, showing the highest diagnostic performance for AA in HSROC analysis is as follows $D_{SC-} = 1.97 \times \text{tenderness (positive endpoint=1, negative endpoint=0)} + 1.88 \times \text{previous abdominal surgery (positive endpoint=1, negative endpoint=0)} + 1.61 \times \text{rebound (positive endpoint=1, negative endpoint=0)} + 1.43 \times \text{rigidity (positive endpoint=1, negative endpoint=0)} + 1.30 \times \text{pain at diagnosis (positive endpoint=1, negative endpoint=0)} + 1.14 \times \text{guarding (positive endpoint=1, negative endpoint=0)} + 1.05 \times \text{body temperature (positive endpoint=1, negative endpoint=0)} - 7.69$. The mean (SD) $D_{SC-}$ values for AA in males (n=149) were –1.13 (1.74) and $D_{SC-}$ mean (SD) values for all male patients with AAP (n=636) were –2.63 (3.05) (Table V).

The formula with LC ($D_{SC+}$) in men. The formula with LC, showing the highest diagnostic performance for AA in men is as follows $D_{SC+} = 2.51 \times \text{previous abdominal surgery (positive endpoint=1, negative endpoint=0)} + 2.18 \times \text{LC (positive endpoint=1, negative endpoint=0)} + 1.58 \times \text{pain at diagnosis (positive endpoint=1, negative endpoint=0)} + 1.41 \times \text{tenderness (positive endpoint=1, negative endpoint=0)} + 1.30 \times \text{pain at diagnosis (positive endpoint=1, negative endpoint=0)} + 1.14 \times \text{guarding (positive endpoint=1, negative endpoint=0)} + 1.05 \times \text{body temperature (positive endpoint=1, negative endpoint=0)} - 7.69$. The mean (SD) $D_{SC+}$ values for AA in males (n=149) were –1.13 (1.74) and $D_{SC+}$ mean (SD) values for all male patients with AAP (n=636) were –2.33 (2.21) (Table VI).

Statistical analysis. The other statistical analyses were performed using STATA/SE version 16.1 (StataCorp, College Station, TX, USA). Statistical tests presented were two-sided, and $P$-value $<0.05$ was considered statistically significant. Using 2x2 tables, we calculated sensitivity (Se) and specificity (Sp) with 95% confidence intervals (95% CI) for each symptom, sign or test, and created separate forest plots for showing each set of data, separately for each diagnostic variable. We calculated the summary estimates of sensitivity (Se) and specificity (Sp), positive (LR+) and negative likelihood ratio (LR–) and diagnostic odds ratio (DOR), using a random effect bivariate model and fitted the summary hierarchical receiving operating characteristic (HSROC) curves, including all diagnostic variables in the $D_{SC-}$ and $D_{SC+}$ models, using the AA endpoint.

Using the STATA's predict tool, we also made posterior predictions [Empirical Bayes (EB) estimates] of the Se and Sp in each variables in both female and male AA patients in $D_{SC-}$ and $D_{SC+}$. Analogous to its use in meta-analysis, EB estimates here give the best estimates of the true Se and Sp for each diagnostic variable, the variable-specific point estimates usually shrinking toward the summary point of the HSROC. We explored the statistical heterogeneity between diagnostic variables and DS models through visual examination of the forest plots and the HSROC curves. To study the potential bias, we used the Cook’s distance to check for the particularly influential variables, together with a scatter plot of the standardised (level 2) residuals to find out the variables that are distinct outliers.

Results

Diagnostic performance of the symptoms. The pooled overall gender-specific (F vs. M) Se of the diagnostic symptoms for detecting AA was 80% (95% CI=67%-90%) and 81% (95% CI=82-94%), respectively (Figures 1 and 2). In women Se was higher than 80% for 12 diagnostic symptoms, while in men Se was higher than 81% for 11 diagnostic symptoms.
The five best diagnostic symptoms in women (vertigo, jaundice, micturition, drugs for abdominal pain and use of alcohol) showed 95-99% Se, whereas in men (vertigo, jaundice, micturition, drugs for abdominal pain and use of alcohol) showed 99-100% Se in diagnosis of AA (Figure 2). The pooled overall Sp of the diagnostic symptoms for detecting AA was 30% (95% CI=19%-42%) and 31% (95% CI=20-43%) for women and men, respectively (Figures 3 and 4). In women 9 diagnostic symptoms showed Sp higher than 30%, while in men Sp exceeded 31% for 10 diagnostic symptoms. The five best diagnostic symptoms in diagnosis of AA among women (location of initial pain, location of pain at diagnosis, type of pain, relieving factors, vomiting) showed 60-84% Sp while those (location of initial pain, in vivo 34: 3687-3703 (2020)
location of pain at diagnosis, type of pain, relieving factors, vomiting) in men showed Sp of 51-89% (Figure 4).

Diagnostic performance of the signs and tests. The pooled overall Se of the diagnostic signs and tests for detecting AA was 86% (95% CI=79%-92%) and 88% (95% CI=82-94%), for women and men, respectively (Figures 5 and 6). In women 10 diagnostic signs and tests had Se exceeding 86%, while in men Se was higher than 88% for 10 diagnostic signs and tests. In diagnosis of AA the five best diagnostic signs and tests in women (distension, tenderness, mass, Murphy’s sign, urine) showed 93-99% Se whereas those (scar, distension, mass, Murphy’s sign, urine) in men showed 95-99% Se (Figure 6). The pooled overall Sp of the signs and
tests was 34% (95% CI=20%-50%) and 34% (95% CI=20-51%) for women and men, respectively (Figures 7 and 8). In women 8 diagnostic signs and tests showed Sp higher than 34%, whereas in men Sp was 34% for 7 diagnostic signs and tests. The seven best diagnostic signs and tests in women showed 59-90% Sp whereas those in men showed 53-86% Sp in diagnosis of AA (Figure 8).

Diagnostic performance of the DS without leucocytes (DS_{LC-}) in women. The most important predictors of AA in women without LC (n=697) were tenderness, rigidity, guarding, location of pain at diagnosis, and renal tenderness. The significant predictors were used to construct the DS_{LC-} formula for AA diagnosis. In practice, the use of the DS formula is relatively simple as shown by the following:

Table V. Diagnostic score of male patients (n=636) without leucocyte count (DS_{LC-}) shown as six different cut-off levels of symptoms, signs and test.

| Diagnostic score (DS) | Positive endpoint | Negative endpoint | TP  | FN  | FP  | TN  |
|-----------------------|-------------------|-------------------|-----|-----|-----|-----|
| 1. Logistic model without leucocyte count DS_{LC-} I | Appendicitis | Other diagnosis of acute abdominal pain | 142 | 4   | 174 | 261 |
| 2. Logistic model without leucocyte count DS_{LC-} II | Appendicitis | Other diagnosis of acute abdominal pain | 140 | 6   | 144 | 291 |
| 3. Logistic model without leucocyte count DS_{LC-} III | Appendicitis | Other diagnosis of acute abdominal pain | 139 | 7   | 101 | 334 |
| 4. Logistic model without leucocyte count DS_{LC-} IV | Appendicitis | Other diagnosis of acute abdominal pain | 131 | 15  | 66  | 369 |
| 5. Logistic model without leucocyte count DS_{LC-} V | Appendicitis | Other diagnosis of acute abdominal pain | 125 | 21  | 40  | 395 |
| 6. Logistic model without leucocyte count DS_{LC-} VI | Appendicitis | Other diagnosis of acute abdominal pain | 125 | 7   | 40  | 334 |

TP, True positive; FN, false negative; FP, false positive; TN, true negative. Cut-off levels: DS I=-3.1, DS II=-2.6, DS III=-2.0, DS IV=-1.4, DS V=-0.48, VI=-2.00--0.48 (follow-up patients n=75).

Figure 3. Pooled specificities of the clinical symptoms in acute appendicitis (random-effects model) in women. ES: Estimated specificity; CI: Confidence interval.
A female patient is admitted to the emergency room with abdominal pain; at diagnosis the pain was localized at RLQ \(1 \times 1.33\); clinical examination showed RLQ tenderness \(1 \times 2.98\), rigidity \(1 \times 2.45\), guarding \(1 \times 2.08\) and the renal tenderness test was positive \(1 \times 0.88\). The best diagnostic performance level for DS LC– formula in females \((\text{Se}=93\%, \text{Sp}=92\%)\) in AA diagnosis was reached when the patients with a DS LC– value between \(-2.03\) and \(-0.49\) were considered as “grey area” patients=follow-up required before the decision to operate \(n=123\). The formula was tested at six different cut-off levels to disclose the best diagnostic performance in women \(\text{Se}=88\% \text{ (95\% CI}=83-92\%)\) and \(89\%\).
Three of these formulas showed Se >88% and four formulas had Sp >89%. At the best diagnostic DS$_{LC-}$ formula in women (formula DS VI, Figure 9 and 10) showed Se of 93% (95% CI =87-97%) and Sp of 92% (95% CI =89-94%).

Diagnostic performance of the DS without leucocytes in women (DS$_{LC-}$). Similar as for the DS$_{LC-}$ formulas, the significant independent predictors were used to build up the six different DS$_{LC+}$ formulas. The pooled overall Se and Sp of these six DS$_{LC+}$ models in women was 90% (95% CI =85-95%) and 85% (95% CI =74-94%), respectively (Figures 11 and 12). Four formulas showed Se >90% and four formulas Sp over 85%. The DS$_{LC+}$ formula (formula DS XII, Figures 11 and 12) showed Se of 93% (95% CI =87-97%) and Sp of 91% (95% CI =88-94%) (Figures 11 and 12).

Diagnostic performance of the DS with leucocytes in men (DS$_{LC+}$). The most important predictors of AA in male patients without LC (n=636) were location of pain at diagnosis, previous abdominal surgery, tenderness, rebound, rigidity, guarding and body temperature (DS$_{LC-}$ formula is shown in patients and methods chapter). In male patients DS$_{LC-}$ formula model reached Se of 95% with Sp of 89% when the male patients with DS value between −2.00 and −0.48 were considered as “grey area” patients=follow-up required before the decision to operate (n=75). The DS$_{LC-}$ formula was tested at six different cut-off levels to find the best diagnostic performance for AA in men (Figures 13 and 14). The pooled overall Se and Sp of these six DS$_{LC-}$ formulas were 94% (95% CI =90-96%) and 79% (95% CI =68-88%), respectively (Figures 13 and 14). Four of these formulas showed Se >94% and three formulas had Sp >79%. At the best diagnostic DS$_{LC-}$ in men (formula DS XI, Figures 13 and 14) showed Se of 95% (95% CI =89-98%) and Sp of 89% (95% CI =86-92%).

Diagnostic performance of the DS with leucocytes in men (DS$_{LC+}$). The pooled overall Se and Sp of the six DS$_{LC+}$ formulas in men was 93% (95% CI =88-96%) and 84% (95% CI =74-92%), respectively (Figures 15 and 16). Four formulas showed Se >93% and three formulas had Sp >84%. The best diagnostic DS$_{LC+}$ formula in men (formula DS XII, Figures 15 and 16) showed Se of 93% (95% CI =87-97%) and Sp of 93% (95% CI =90-96%) (Figures 15 and 16).
HSROC analyses and empirical Bayes (EB) estimates in both genders. STATA (metandiplot algorithm) was used to draw the HSROC curves and EB estimates to visualise the comparison of the pooled overall diagnostic performance of the different DS formulas in detecting AA in women (Figures 17 and 18) and men (Figures 19 and 20). In the HSROC analysis in women, there is no statistically significant difference between the DS LC– and DS LC+ formulas, with AUC=0.949 (95% CI=0.921-0.968) and AUC=0.953 (95% CI=0.923-0.969) (p=0.631, ROC comparison test). The same is true with the HSROC analysis in men, with no difference between the DS LC– and DS LC+ formulas, with AUC=0.948 (95% CI=0.920-0.964) and AUC=0.956 (95% CI=0.930-0.969) (p=0.321, ROC comparison test).

Discussion

We studied patients presenting with AAP as a part of the survey by the OMGE Committee (4-8) and estimated the diagnostic accuracy of a combined history-taking, clinical examination and laboratory testing in verified AA (5), NSAP (5), acute small bowel obstruction (7) and in acute renal stone disease (8). Although there are several different DS systems designed for AAP diagnosis (5, 9-15) and the international guidelines recommend routine diagnostic scoring to improve the diagnosis of AA (16, 17), a debate continues on the shortcomings of the specific DS models in women and men with AAP. Thus, it was appropriate to compare the performance of our gender-specific DS models in both genders, using DSs with and without LC.

Comparison of the symptoms, signs and laboratory tests in women vs. men. There was no gender-specific difference in the clinical symptoms, since the five diagnostic symptoms with highest diagnostic accuracy were identical in women and men (vertigo, jaundice, micturition, drugs for abdominal pain and use of alcohol), showing 95-100% Se in diagnosis of AA. The same applies to gender-specific difference in Sp, the five most relevant symptoms being identical in both genders (location of initial pain, location of pain at diagnosis, type of pain, relieving factors and vomiting) presenting with 51-89% Sp in diagnosis of AA.

Similarly, there was no significant difference in signs and laboratory test results between women and men with confirmed...
AA in their pooled Se, because the five best diagnostic signs and tests were very similar in both genders; women (distension, tenderness, mass, Murphy’s sign, urine) and men (scar, distension, mass, Murphy’s sign, urine) showing 93-99% Se in diagnosis of AA. Also, the pooled Sp of the signs and tests in AA detection was equal in both genders.

Female DS$_{LC-}$ and DS$_{LC+}$. Of interest was to assess, whether the addition of LC would give any added value to our DSs, herein performed by comparing the diagnostic accuracy of DS$_{LC-}$ and DS$_{LC+}$ scores. The present analysis suggests that female patients with DS$_{LC-}$ below –2.03 should not be operated while women with DS$_{LC-}$ falling between –2.03 and –0.49 should be followed-up before the final decision. According to our data, only the AAP-women with the DS$_{LC-}$ values exceeding –0.49 should be operated without delay. In women whose LC was calculable (n=575) the important predictors of AA were the same as in women without LC, but LC is added to DS$_{LC+}$ formula (LC≥10,000 μl) (see the methods for formula details). In women, the highest diagnostic accuracy for DS$_{LC+}$ formula (Se=93%, Sp=91%) in AA diagnosis was reached when the patients with DS$_{LC+}$ values falling between –2.33 and –0.41 were considered as “grey zone” patients, for whom follow-up was appropriate before the decision to operate (n=77). Taken together, i) the female AAP patients with DS$_{LC+}$ value below –2.33 should not be operated, ii) those with DS$_{LC+}$ value between –2.33 and –0.41 should be followed-up and iii) all AAP-women with the DS$_{LC+}$ value higher than –0.41 should be operated without delay.

Male DS$_{LC-}$ and DS$_{LC+}$. The same considerations can be made among male patients with AAP. Our present data suggest that male patients with a DS$_{LC-}$ value below –2.00 should not be operated. Those men with DS$_{LC-}$ values between –2.00 and –0.48 should be followed-up, whereas all those with DS$_{LC-}$ values above –0.48 should be operated with no delay. In males with LC available (n=476), the AA predictors were the same as earlier but added with the LC and rectal digital examination. When the men with a DS$_{LC-}$ value between –1.74 and –0.14 were considered equivocal (n=67, follow-up required), the Se of this DS$_{LC-}$ in AA was 93%, with a Sp of 93% and an efficiency of 93%. As to the males and DS$_{LC+}$ the present analysis implicates that the

| Study | ES (95% CI) | % Weight |
|-------|------------|----------|
| 1 (1. Mood) | 0.16 (0.13, 0.20) | 5.88 |
| 2 (2. Colour) | 0.12 (0.10, 0.15) | 5.88 |
| 3 (3. Abdominal movement) | 0.06 (0.04, 0.08) | 5.88 |
| 4 (4. Scar) | 0.35 (0.31, 0.39) | 5.88 |
| 5 (5. Distension) | 0.07 (0.05, 0.10) | 5.88 |
| 6 (6. Tenderness (OMGE)) | 0.70 (0.67, 0.74) | 5.88 |
| 7 (7. Mass) | 0.04 (0.03, 0.06) | 5.88 |
| 8 (8. Rebound) | 0.62 (0.58, 0.66) | 5.88 |
| 9 (9. Guarding) | 0.59 (0.55, 0.63) | 5.88 |
| 10 (10. Rigidity) | 0.90 (0.87, 0.92) | 5.88 |
| 11 (11. Murphy’s positive) | 0.13 (0.11, 0.16) | 5.88 |
| 12 (12. Bowel sounds) | 0.11 (0.09, 0.14) | 5.88 |
| 13 (13. Renal tenderness) | 0.29 (0.25, 0.33) | 5.88 |
| 14 (14. Rectal digital tenderness) | 0.74 (0.70, 0.77) | 5.88 |
| 15 (15. Body temperature (Temp)) | 0.62 (0.58, 0.66) | 5.88 |
| 16 (16. Leucocyte count (LC)) | 0.72 (0.68, 0.76) | 5.88 |
| 17 (17. Urine) | 0.06 (0.04, 0.09) | 5.88 |
| Overall (F² = 99.62%, \( p = 0.00 \)) | 0.34 (0.20, 0.50) | 100.00 |

Figure 7. Pooled specificities of the clinical signs and tests in acute appendicitis (random-effects model) in women. ES: Estimated sensitivity; CI: Confidence interval.
Figure 8. Pooled specificities of the clinical signs and tests in acute appendicitis (random-effects model) in men. ES: Estimated sensitivity; CI: Confidence interval.

Figure 9. Sensitivities of diagnostic scores without leucocyte count (DS\textsubscript{LC-}) in women at six different cut-off levels (DS I-VI).
patients with a DS\textsubscript{L}C\textsuperscript{+} value below –1.74 should not be operated, while those with DS\textsubscript{L}C\textsuperscript{+} values between –1.74 and –0.14 could be safely followed-up. This leaves only the male patients with DS\textsubscript{L}C\textsuperscript{+} values above –0.14, who should be immediately operated.

A right iliac fossa tenderness (RIFT) Study Group in UK suggested that female patients (272/964, 28.2%) with AAP were more than twice as likely as males to undergo surgery with histologically normal appendix as a result (i.e. negative appendectomy, FP) (120/993, 12.1%) (RR=2.33, 95% CI=1.92-2.84, \(p<0.001\)). Although several AA risk scorings (5, 9-15) demonstrate different predictive factors for AA diagnosis, as far as we know, only the RIPASA scoring provides gender-specific data (19). The present study is the
second to provide such data, while reporting the diagnostic performance of the DS<sub>LC−</sub> and DS<sub>LC+</sub> models in both genders: The aim was to elaborate the optimal combination of symptoms, signs and tests in the DS formulas with and without LC and using six different combinations (DS<sub>LC−</sub> and DS<sub>LC+</sub>) as diagnostic predictors of AA. Our DS is in line with APPEND score (20) that the LC testing was not an important predictor of AA diagnosis, while APPEND score does not account for the significance of gender in AA. Alvarado (9) and Appendicitis Inflammatory Response (21) scores emphasize LC analysis as a significant predictor of AA diagnosis and APPEND score identified neutrophil percent and neutrophil/lymphocyte ratio as important predictors of AA. AA diagnosis strategies in the future may include early markers of

### Figure 12. Specificities of diagnostic scores with leucocyte count (DSLC+) in women at six different cut-off levels (DS VII-XII).

| Study | ES (95% CI) | % Weight |
|-------|-------------|----------|
| 1 (1. Logistic model with leucocyte count DS VII) | 0.59 (0.54, 0.63) | 16.67 |
| 2 (2. Logistic model with leucocyte count DS VIII) | 0.77 (0.73, 0.81) | 16.67 |
| 3 (3. Logistic model with leucocyte count DS IX) | 0.87 (0.83, 0.90) | 16.67 |
| 4 (4. Logistic model with leucocyte count DS X) | 0.92 (0.90, 0.95) | 16.67 |
| 5 (5. Logistic model with leucocyte count DS XI) | 0.96 (0.93, 0.97) | 16.67 |
| 6 (6. Logistic model with leucocyte count DS XII) | 0.91 (0.88, 0.94) | 16.63 |
| Overall (I² = 98.28%, p = 0.00) | 0.85 (0.74, 0.94) | 100.00 |

### Figure 13. Sensitivities of diagnostic scores without leucocyte count (DSLC−) in men at six different cut-off levels (DS I-VI).

| Study | ES (95% CI) | % Weight |
|-------|-------------|----------|
| 1 (1. Logistic model without leucocyte count DS I) | 0.97 (0.93, 0.99) | 16.74 |
| 2 (2. Logistic model without leucocyte count DS II) | 0.95 (0.91, 0.98) | 16.74 |
| 3 (3. Logistic model without leucocyte count DS III) | 0.95 (0.90, 0.98) | 16.74 |
| 4 (4. Logistic model without leucocyte count DS IV) | 0.90 (0.84, 0.94) | 18.74 |
| 5 (5. Logistic model without leucocyte count DS V) | 0.86 (0.79, 0.91) | 18.74 |
| 6 (6. Logistic model without leucocyte count DS VI) | 0.85 (0.80, 0.90) | 16.30 |
| Overall (I² = 74.40%, p = 0.00) | 0.94 (0.90, 0.98) | 100.00 |
inflammation e.g. interleukin 6 (IL-6) blood levels of which were shown to increase even 3-fold from the reference levels in perforated AA (22), suggesting that IL-6 analysis may be useful in predicting AA complication risk. Although the IL-6 analysis is promising, the current antigen test method precludes its use as a rapid test in AA so far (23, 24).

**Conclusion**

In conclusion, our gender-specific DS reached very high AUC values (0.948-0.956) in both genders, and using the ROC comparison test, there was no statistically significant difference in the AUC values of DS LC− and DS LC+ between...
women and men. Kularatna et al. (15) reviewed available DS formulas in AA and showed between 84% and 96% AUC values in AA diagnosis. Although, weakness of the meta-analysis is heterogeneity and quality of included studies, it seems that the Tzanakis score (13) with ultrasound (US) and inflammatory markers reached the
highest diagnostic performance with 96% AUC in AA. This is equal with the AUC values obtained in the present study. However, the advantage of our DS is that this scoring does not need US or LC analysis to reach a high diagnostic accuracy in AA.

Conflicts of Interest

The Authors report no conflicts of interest or financial ties to disclose. The Authors alone are responsible for the content and writing of this article.

Authors’ Contributions

All Authors have met all of the following four criteria: 1. Substantial contributions to the conception or design of the work or the acquisition, analysis, or interpretation of data for the work. 2. Drafting the work or revising it critically for important intellectual content. 3. Final approval of the version to be published. 4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The study was funded by the Päivikki ja Sakari Sohlberg Foundation.

References

1. Ferris M, Quan S, Kaplan BS, Molodecky N, Ball CG, Chernoff GW, Bhala N, Ghosh S, Dixon E, Ng S and Kaplan GG: The global incidence of appendicitis: a systematic review of population-based studies. Ann Surg 266: 237-241, 2017. PMID: 28288060. DOI: 10.1097/SLA.0000000000002188
2. Meklin J, Eskelinen M, Syrjänen K and Eskelinen M: Leucocyte count does not improve the diagnostic performance of a diagnostic score (DS) in distinguishing acute appendicitis (AA) from non-specific abdominal pain (NSAP). In Vivo, 2020 In press.
3. A Bhangu, RIFT Study Group on behalf of the West Midlands Research Collaborative: Evaluation of appendicitis risk prediction models in adults with suspected appendicitis. Br J Surg 107: 73-86, 2020. PMID: 31797357. DOI: 10.1002/bjs.11440
4. Dombal FT de: The OMGE acute abdominal pain survey. Progress report 1986. Scand J Gastroenterol 23(Suppl. 144): 35-42, 1988. PMID: 3043646.
5. Eskelinen M, Ikonen J and Lipponen P: A computer-based diagnostic score to aid in diagnosis of acute appendicitis. Theor Surg 7: 86-90, 1992.
6. Eskelinen M and Lipponen P: Usefulness of history-taking in non-specific abdominal pain: a prospective study of 1333 patients with acute abdominal pain in Finland. In Vivo 26: 335-339, 2012. PMID: 22351680.
7. Eskelinen M, Ikonen J and Lipponen P: Contributions of history-taking, physical examination, and computer assistance to diagnosis of acute small-bowel obstruction. A prospective study.
of 1333 patients with acute abdominal pain. Scand J Gastroenterol 29: 715-721, 1994. PMID: 7973431. DOI: 10.3109/00365529409092499
8 Eskelinen M, Ikonen J and Lipponen P: Usefulness of history-taking, physical examination and diagnostic scoring in acute renal colic. Eur Urol 34: 467-473, 1998. PMID: 9831787. DOI: 10.1159/000019785
9 Alvarado A: A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 15: 557-564, 1986. PMID: 3963537. DOI: 10.1016/s0196-0644(86)80993-3
10 Ohmann C, Franke C and Yang Q: Clinical benefit of a prospective intervention study. German Study Group of Acute Abdominal Pain. Arch Surg 134: 993-996, 1999. PMID: 10487595. DOI: 10.1001/archsurg.134.9.993
11 Sitter H, Hoffmann S, Hassan I and Zielke A: diagnostic score in appendicitis. validation of a diagnostic score (Eskelinen score) in patients in whom acute appendicitis is suspected. Langenbecks Arch Surg 389: 213-218, 2004. PMID: 14624293. DOI: 10.1007/s00423-003-0436-9
12 Lintula H, Pesonen E, Kokki H, Vanamo K and Eskelinen M: A diagnostic score for children with suspected appendicitis. Langenbecks Arch Surg 390: 164-170, 2005. PMID: 15723233. DOI: 10.1007/s00423-005-0545-8
13 Tzanakis NE, Efstathiou SP, Danulidis K, Rallis GE, Tsioulos DI, Chatzivasilisou, Peros G and Nikiteas NI: A new approach to accurate diagnosis of acute appendicitis. World J Surg 29: 1151-1156, 2005. PMID: 16088420. DOI: 10.1007/s00268-005-7853-6
14 Frountzas M, Stergios K, Kopsini D, Schizas D, Kontzoglou K and Toutouzas K: Alvarado or RIPASA score for diagnosis of acute appendicitis? A meta-analysis of randomized trials. Int J Surg 56: 307-314, 2018. PMID: 30017607. DOI: 10.1016/j.ijsu.2018.07.003
15 Kularatna M, Lauti M, Haran C, MacFater W, Sheikh L, Huang Y, McCall J and MacCormick AD: Clinical prediction rules for appendicitis in adults: which is best? World J Surg 41: 1769-1781, 2017. PMID: 28258458. DOI: 10.1007/s00268-017-3926-6
16 Di Saverio S, Birindelli A, Kelly MD, Catena F, Weber DG, Sartelli M, Sugrue M, De Moya M, Gomes CA, Bhangu A, Agresta F, Moore EE, Soreide K, Griffihs E, De Castro S, Kashuk J, Kluger Y, Leppaniemi A, Ansaloni L, Andersson M, Coccollini F, Coimbra R, Rursamy KS, Campanile FC, Biffi W, Chiara O, Moore F, Peitzman AB, Fraga GP, Costa D, Maier RV, Rizoli S, Balogh ZJ, Bendinelii C, Cirocchi R, Tonini V, Piccinni A, Tugnoli G, Jovine E, Persiani R, Biondi A, Saclea A, Stahel P, Ivatury R, Velmahos G and Andersson R: WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. World J Emerg Surg 11: 34, 2016. PMID: 27437029. DOI: 10.1186/s13017-016-0090-5
17 Gorter RR, Eker HH, Gorter-Stam MA, Abis GS, Acharya A, Ankermits M, Antoniou SA, Arolfo S, Babic B, Boni L, Brunink M, van Dam DA, Defoort D, Deijen CL, DeLacy FB, Go PM, Harmsen AM, van den Helder RS, Iordache F, Ket JC, Myousses FE, Ozemen MM, Papoulos M, Rhodes M, Straatman J, Tenhagen M, Turrado V, Verczei A, Vilallonga R, Deelder JD and Bonjer J: Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. Surg Endosc 30: 4668-4690, 2016. PMID: 27660247. DOI: 10.1007/s00464-016-5245-7
18 Eskelinen M, Ikonen J and Lipponen P: Sex-specific diagnostic scores for acute appendicitis. Scand J Gastroenterol 29: 59-66, 1994. PMID: 8128177. DOI: 10.3109/00365529409090438
19 Butt MQ, Chatha SS, Ghuamman AQ and Farooq M: RIPASA score: a new diagnostic score for diagnosis of acute appendicitis. J Coll Physicians Surg Pak 24: 894-897, 2014. PMID: 25523723.
20 Mikaere H, Zeng I, Lauti M, Kularatna M and MacCormick AD: Derivation and validation of the APPEND score: an acute appendicitis clinical prediction rule. ANZ J Surg 88: E303-E307, 2018. PMID: 28512853. DOI: 10.1111/ans.14022
21 Andersson M and Andersson RE: The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World J Surg 32: 1843-1849, 2008. PMID: 18553045. DOI: 10.1007/s00268-008-9649-y
22 Anielski R, Kuśnierz-Cabala B and Szafraniec K: An evaluation of the utility of additional tests in the preoperative diagnostics of acute appendicitis. Langenbecks Arch Surg 395: 1061-1068, 2010. PMID: 19924356. DOI: 10.1007/s00268-009-0565-x
23 Aspinen S, Kinnunen M, Harju J, Juvonen P, Selander T, Holopainen A, Kokki H, Pulikki K and Eskelinen M: Inflammatory response to surgical trauma in patients with minimalabdominal cholecystectomy versus laparoscopic cholecystectomy: a randomised multicentre study. Scand J Gastroenterol 51: 739-744, 2016. PMID: 26758677. DOI: 10.3109/00365525.2015.1129436
24 Purdy M, Kokki M, Anttila M, Aspinen S, Juvonen P, Korhonen R, Selander T, Kokki H and Eskelinen M: Does the rectus sheath block analgesia reduce the inflammatory response biomarkers’ IL-1ra, IL-6, IL-8, IL-10 and IL-1β concentrations following surgery? A randomized clinical trial of patients with cancer and benign disease. Anticancer Res 36: 3005-3011, 2016. PMID: 27272818.

Received September 15, 2020
Revised September 30, 2020
Accepted October 2, 2020