Abstract. Background/Aim: The diagnostic scores (DSs) for patients with non-specific abdominal pain (NSAP) have been rarely evaluated. Patients and Methods: In the NSAP study group there were 614 patients (268 females and 346 males) versus 719 patients in the non-NSAP group including 368 females and 351 males. The clinical symptoms (n=22), signs and tests (n=14) and laboratory analyses (n=3) were recorded in each patient. Meta-analytical techniques were used to detect the summary sensitivity (Se) and specificity (Sp) estimates for each data set (symptoms, signs and tests as well as DS models). Results: In receiver operating characteristic (ROC) analysis, the area under curve (AUC) values for i) symptoms ii) signs and tests and iii) DS were as following: i) AUC=0.542 (95% CI=0.512-0.572); ii) AUC=0.625 (95% CI=0.550-0.700), and iii) AUC=0.874 (95% CI=0.850-0.898). The differences between these AUC values are as following: between i and ii, p=0.097; between i and iii, p<0.0001 and between ii and iii, p<0.0001. Conclusion: This is the first study to provide evidence that DS may help in the difficult diagnosis of NSAP.

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 and +358 400969444, Fax: +358 17172611, e-mail: matti.eskelinen@kuh.fi

Key Words: Non-specific abdominal pain, symptoms, signs, tests, diagnostic score, HSROC, diagnostic accuracy.
Table I. The clinical history of non-specific abdominal pain (NSAP) versus any other cause of acute abdominal pain.

| Clinical history variable | Positive endpoint | Negative endpoint | TP | FN | FP | TN |
|---------------------------|-------------------|------------------|----|----|----|----|
| 1. Location of initial pain | Upper, central or lower midline | Other | 343 | 271 | 369 | 350 |
| 2. Location of pain at diagnosis | Upper, central or lower midline | Other | 260 | 354 | 211 | 508 |
| 3. Duration of pain at diagnosis | ≤12 hours | >12 hours | 217 | 397 | 241 | 475 |
| 4. Intensity of abdominal pain | Subjectively moderate or weak pain | Intolerable pain | 537 | 77 | 580 | 139 |
| 5. Progression of pain from onset to diagnosis | Weaker or subjectively same pain than at the onset | Worse pain | 478 | 136 | 468 | 251 |
| 6. Type of pain | Steady pain | Colicky or intermitted pain | 313 | 301 | 418 | 301 |
| 7. Aggravating factors | No aggravating factors | Movement, coughing, respiration, food or other | 204 | 410 | 152 | 567 |
| 8. Relieving factors | No | Yes | 233 | 381 | 201 | 518 |
| 9. Previous similar pain | Yes | No | 210 | 397 | 237 | 475 |
| 10. Vertigo | No | Yes | 588 | 23 | 701 | 17 |
| 11. Nausea | Yes | No | 326 | 288 | 440 | 279 |
| 12. Vomiting | No | Yes | 403 | 211 | 355 | 364 |
| 13. Appetite | Normal appetite | No appetite | 207 | 407 | 149 | 570 |
| 14. Previous indigestion | No | Yes | 504 | 108 | 548 | 171 |
| 15. Jaundice | No | Yes | 609 | 5 | 691 | 28 |
| 16. Bowels | Normal | Diarrhea, constipation, blood, mucus or white stools | 472 | 142 | 543 | 176 |
| 17. Micturition | Normal | Abnormal | 581 | 33 | 666 | 53 |
| 18. Drugs for abdominal pain | No | Yes | 589 | 25 | 689 | 29 |
| 19. Previous abdominal surgery | No | Yes | 477 | 137 | 522 | 196 |
| 20. Previous abdominal diseases | No | Yes | 516 | 98 | 582 | 136 |
| 21. Use of alcohol | No | Yes | 581 | 33 | 684 | 34 |
| 22. Gender | Female | Male | 268 | 346 | 368 | 351 |

FN: False negative; FP: false positive; TN: true negative; TP: true positive.

relative risk of a patient with a given symptom and sign and test to have NSAP.

The DS formula for NSAP was: 0.22 × Gender (female=1, male=0) - 0.02 × Age (years) - 0.47 × Location of initial pain (PE=1, NE=0) + 0.53 × Location of pain at diagnosis (PE=1, NE=0) + 0.45 × Progression of pain (PE=1, NE=0) + 0.29 × Relieving factors (PE=1, NE=0) - 0.38 × Previous similar pain (PE=1, NE=0) + 0.70 × Vertigo (PE=1, NE=0) + 1.63 × Jaundice (PE=1, NE=0) + 0.47 × Mood (PE=1, NE=0) + 0.99 × Distension (PE=1, NE=0) + 2.34 × Mass (PE=1, NE=0) + 0.55 × Rebound (PE=1, NE=0) + 0.76 × Guarding (PE=1, NE=0) + 2.90 × Rigidity (PE=1, NE=0) + 1.49 × Murphy (PE=1, NE=0) + 0.59 × Bowel sounds (PE=1, NE=0) + 1.02 × Leucocyte count (PE=1, NE=0) + 3.36 × Urine (PE=1, NE=0) - 13.77. PE-positive endpoint and NE-negative endpoint (Table III).

Statistical analysis. STATA/SE version 16.1 (StataCorp, College Station, TX, USA) was used for analysis. The statistical tests presented were two-sided, and p-Values <0.05 were considered statistically significant. Using 2×2 tables, sensitivity (Se) and specificity (Sp) with 95% confidence intervals (95% CI) for each clinical history-taking variable, finding or test was determined. A meta-analytical technique (metaprop) was used to create separate forest plots for Se and Sp for each set of data, including each diagnostic variable. We calculated the summary estimates of Se and Sp, positive (LR+) and negative likelihood ratio (LR−) and diagnostic odds ratio, using a random effects bivariate model and fitted the summary hierarchical receiving operating characteristic (HSROC) curves using the NSAP endpoint. Roccomp test (STATA) was used to compare the AUC values of HSROC tests between the 3 diagnostic sets (history-taking, clinical signs, DSs).

Results

Patient data of the study. In the NSAP study group, 614 patients (268 females and 346 males) were included, and in the non-NSAP group, there were 719 patients (368 females and 351 males) with the following AAP diagnoses: acute appendicitis (n=271), acute cholecystitis (n=124), acute renal colic (n=59), acute small bowel obstruction (n=53), non-organic dyspepsia (n=50) and other AAP patients (n=160), with the mean (SD) age of 37.5 (21.7) years.

The clinical symptoms in NSAP. The overall sensitivity of the clinical symptoms for detecting NSAP was 69% (95% CI=58-80%) (Figure 1). The Se was higher than 69% for 11 of the symptoms. The five most sensitive clinical history-taking variables (vertigo, jaundice, micturition, drugs for abdominal pain and use of alcohol) showed 95-99% Se in

in vivo 35: 2729-2738 (2021)
diagnosis of NSAP (Figure 1). The overall specificity of the history-taking for detecting NSAP was only 35% (95% CI=24-48%) (Figure 2). Altogether, 11 symptoms showed Sp higher than 35%. The five most specific symptoms of NSAP (location of pain at diagnosis, aggravating factors, relieving factors, previous similar pain and appetite) showed Sp values of 67-79% (Figure 2).

The clinical signs and tests in NSAP. The overall sensitivity of the signs and tests for NSAP was 86% (95% CI=76-93%) (Figure 3), and 9 signs and tests had Se exceeding 86%. The six most accurate signs and tests (abdominal movement, distension, mass, rigidity, Murphy’s positive and urine) showed 97-100% Se (Figure 3). The overall specificity of the signs and tests was only 31% (95% CI=20-43%) (Figure 4), while 7 signs and tests showed Sp higher than 31%. The five most specific signs and tests (tenderness, rebound, guarding, body temperature and leucocyte count), however, showed 47-81% Sp (Figure 4).

The $DS$ in NSAP. The most significant predictors of NSAP in multivariate analysis were gender, age (years), location of...
initial pain, location of pain at diagnosis, progression of pain, relieving factors, previous similar pain, vertigo, jaundice, mood, distension, mass, rebound, guarding, rigidity, Murphy, bowel sounds, leucocyte count and urine. The best diagnostic level for DS model [DS IV; Se=81%, Sp=80%, efficiency (Eff)=81%] was reached at a cut-off level of 0.54 for DS (Figures 5 and 6). The DS model was tested at six different cut-off levels to disclose the highest diagnostic accuracy (Figures 5 and 6). The overall Se and Sp of these six DS models were 82% (95% CI=80-84%) and 79% (95% CI=77-81%), respectively (Figures 5 and 6). Three of these models showed Se ≥82% and four models had Sp ≥79%.

**HSROC and AUC values.** HSROC curves were used to visualise the pooled overall accuracy of the symptoms (Figure 7), signs and tests (Figure 8) and different DS models (Figure 9) in detecting NSAP. In SROC analysis, the AUC values for i) symptoms ii) signs and tests as well as iii) DS were as follows: i) AUC=0.542 (95% CI=0.512-0.572); ii) AUC=0.625 (95% CI=0.550-0.700), and iii) AUC=0.874 (95% CI=0.850-0.898). The differences between these AUC values (roccomp analysis) are as following: between i and ii, p=0.097; between i and iii, p<0.0001 and between ii and iii, p<0.0001.

**Discussion**

Prompted by the difficulty of NSAP diagnosis among the AAP patients and the lack of diagnostic accuracy studies on DS with HSROC analysis, we designed the present study to assess the diagnostic performance of i) symptoms, ii) signs and tests, as well as iii) the DS in confirming NSAP.
The clinical symptoms and signs and tests investigated in the AAP patients follow the diagnostic criteria shown in the Research Committee of the World Organization of Gastroenterology (OMGE) (9-16). Here we refer the most important clinical features in making the distinction between NSAP and AA. Nausea and vomiting are usually reported to be in favour of AA, however, the diagnostic accuracy of these symptoms has not been considered in detail before. In our study, 53% of the NSAP patients had nausea and 34% had vomiting. Special attention is paid to the location of pain, which in AA moves from midline to right lower quadrant (RLQ). Instead, in NSAP the pain is diffuse or remains in RLQ. At physical examination, it is necessary to record the abdominal tenderness, rebound tenderness, guarding, abdominal rigidity and Murphy’s sign. Location of tenderness in AA is mostly focal RLQ tenderness, whereas in NSAP, the location of tenderness is usually described to localize at midline or being more diffuse. In our study, 30% of the patients with NSAP had abdominal tenderness at the midline of the abdomen (Se of 0.30 and Sp of 0.81). Rebound tenderness and guarding are usually reported to be negative in NSAP patients, but the diagnostic accuracy of these tests has rarely been assessed in NSAP. In the present series, 33% of the NSAP patients had a negative guarding test (Se of 0.67 and Sp of 0.70) and 29% had positive rebound tenderness (Se of 0.71 and Sp of 0.63). Both clinical tests, when absent and correctly assessed, exclude intra-abdominal inflammation and peritoneal irritation. The abdominal rigidity test is usually shown to be negative in NSAP patients and in our cohort, 98% (11/613) of the NSAP patients had a negative abdominal rigidity test result (Se of 0.98 and Sp of 0.39).

Figure 2. Specificity values of history-taking in non-specific abdominal pain (NSAP) (random-effects model). ES: Estimated specificity; CI: confidence interval.
Acute appendicitis (AA) is a reason of similar symptoms and signs than that in NSAP and therefore the AA is an important differential diagnostic disease in confirming NSAP. Meklin et al. (8) reported the overall Se of the symptoms in AA; 80% (95% CI=67-90%), which was higher than that in NSAP patients in this study; 69% (95% CI=58-80 %). However, the Sp of the symptoms in NSAP in this study was slightly higher than that in AA patients in Meklin et al. (8) study; 35% (95% CI=24-48%) versus 30% (95% CI=19-42%). The overall Se of the signs and tests in detecting NSAP in this study was slightly higher than that in AA patients in Meklin et al. (8) study; 86% (95% CI=76-93%), which was similar to that of the AA patients in Meklin et al. (8) study; 86% (95% CI=79-92%). However, the pooled Sp of the signs & tests in detecting NSAP in this study was slightly lower than that of the AA patients in Meklin et al. (8) study; 31% (95% CI=20-43%) versus 34% (95% CI=20-50%).

When the NSAP patients in this study and the AA patients in Meklin et al. (8) are compared using the DS models, a similar trend can be seen. The overall Se of the DS models in NSAP is 82% (95% CI=80-84%) which is lower than that in AA patients (90%; 95% CI=85-95%). Although Se and Sp usually behave reciprocally, this was not the case with the overall Sp of the DS in NSAP patients, which was 79% (95% CI=77-81%), lower than that in AA (85%; 95% CI=74-94%).

ROC analysis has become popular to evaluate the diagnostic accuracy of various clinical methods and tests. The ROC analysis displays Se as a function of the false positive (FP) rate (1- Sp). Figure 7 shows the ROC analysis for clinical symptoms in NSAP detection and the curve closely parallels the diagonal reference line (AUC=0.500) with a low AUC value (AUC=0.542; 95% CI=0.512-0.572). The diagnostic accuracy of the signs and tests is slightly better than that of the clinical symptoms (AUC=0.625; 95% CI=0.550-0.700, Figure 8).
Figure 9 shows the ROC analysis for the DS, with the curve moved towards the upper left corner, showing significantly better diagnostic performance in the NSAP patients than that of the clinical examination. The ROC analysis can also be used for test optimization by selecting various cut-off points for DS. The value of the clinical test could then be expressed by the Se and Sp for this particular cut-off point in ROC analysis and not for hypothetical situation, where the cut-off point is continuously changing. In the present series, however, the diagnostic accuracy of the DS (AUC=0.874; 95% CI=0.850-0.898) was lower for the NSAP patients than that in AA patients, the major benefit of the DS is a possibility to avoid unnecessary laboratory analyses, endoscopy or radiological procedures to reach adequate diagnostic performance for NSAP.

**Conclusion**

Unfortunately, we could not perform direct comparisons to earlier clinical trials in NSAP, because this is the first study to provide evidence that DS could be used to assist in the difficult diagnosis of NSAP. Although the diagnostic accuracy of the DS is lower for the NSAP patients than that in AA patients, the major benefit of the DS is a possibility to avoid unnecessary laboratory analyses, endoscopy or radiological procedures to reach adequate diagnostic performance for NSAP.

**Conflicts of Interest**

The Authors report no conflicts of interest or financial ties.

**Authors’ Contributions**

All Authors contributed to the collection and analysis of data, drafting and revising the manuscript and read and approved the final article.

**Acknowledgements**

The study was funded by the Päivikki and Sakari Sohlberg Foundation.
### Figure 5. Sensitivity values of diagnostic scores at six different cut-off levels (DS I-VI). ES: Estimated sensitivity; CI: confidence interval.

| Study                         | ES (95% CI)   | Weight |
|-------------------------------|---------------|--------|
| 1. Logistic model DS I        | 0.76 (0.72, 0.79) | 16.67  |
| 2. Logistic model DS II       | 0.76 (0.73, 0.80) | 16.67  |
| 3. Logistic model DS III      | 0.79 (0.76, 0.82) | 16.67  |
| 4. Logistic model DS IV       | 0.80 (0.77, 0.83) | 16.67  |
| 5. Logistic model DS V        | 0.80 (0.77, 0.83) | 16.67  |
| 6. Logistic model DS VI       | 0.82 (0.79, 0.85) | 16.67  |
| Overall (I^2 = 59.70%, p = 0.03) | 0.79 (0.77, 0.81) | 100.00 |

### Figure 6. Specificity values of diagnostic scores at six different cut-off levels (DS I-VI). ES: Estimated specificity; CI: confidence interval.

| Study                         | ES (95% CI)   | Weight |
|-------------------------------|---------------|--------|
| 1. Logistic model DS I        | 0.85 (0.82, 0.88) | 16.67  |
| 2. Logistic model DS II       | 0.84 (0.81, 0.87) | 16.67  |
| 3. Logistic model DS III      | 0.82 (0.79, 0.85) | 16.67  |
| 4. Logistic model DS IV       | 0.81 (0.78, 0.84) | 16.67  |
| 5. Logistic model DS V        | 0.81 (0.77, 0.84) | 16.67  |
| 6. Logistic model DS VI       | 0.78 (0.75, 0.81) | 16.67  |
| Overall (I^2 = 61.02%, p = 0.03) | 0.82 (0.80, 0.84) | 100.00 |
References

1 Hastings RS and Powers RD: Abdominal pain in the ED: a 35 year retrospective. Am J Emerg Med 29(7): 711-716, 2011. PMID: 20825873. DOI: 10.1016/j.ajem.2010.01.045

2 Saaristo L, Ukkonen MT, Laukkarinen JM and Pauniaho SK: The rate of short-term revisits after diagnosis of non-specific abdominal pain is similar for surgeons and emergency physicians - results from a single tertiary hospital emergency department. Scand J Trauma Resusc Emerg Med 28(1): 63, 2020. PMID: 32611415. DOI: 10.1186/s13049-020-00751-8

3 Cervellin G, Mora R, Ticinesi A, Meschi T, Comelli I, Catena F and Lippi G: Epidemiology and outcomes of acute abdominal pain in a large urban Emergency Department: retrospective analysis of 5,340 cases. Ann Transl Med 4(19): 362, 2016. PMID: 27826565. DOI: 10.21037/atm.2016.09.10

4 Fagerström A, Paajanen P, Saarelainen H, Ahonen-Siirtola M, Ukkonen M, Miettinen P and Paajanen H: Non-specific abdominal pain remains as the most common reason for acute abdomen: 26-year retrospective audit in one emergency unit. Scand J Gastroenterol 52(10): 1072-1077, 2017. PMID: 28657380. DOI: 10.1080/00365521.2017.1342140

5 Andersson M, Kolodziej B and Andersson RE: Validation of the appendicitis inflammatory response (AIR) score. World J Surg 45(7): 2081-2091, 2021. PMID: 33825049. DOI: 10.1007/s00268-021-06042-2

6 Paajanen P, Lehtimäki TT, Fagerström A and Paajanen H: Diagnostic potential of magnetic resonance imaging in chronic...
7 Ravn-Christensen C, Qvist N, Bay-Nielsen M and Bisgaard T: Pathology is common in subsequent visits after admission for non-specific abdominal pain. Dan Med J 66(7): A5549, 2019. PMID: 31256781.

8 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 nonspecific abdominal pain (NSAP). In Vivo 34(6): 3327-3339, 2020. PMID: 33144440. DOI: 10.21873/invivo.12171

9 de Dombal FT: The OMGE acute abdominal pain survey. Progress report, 1986. Scand J Gastroenterol Suppl 144: 35-42, 1988. PMID: 3043646.

10 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(2): 335-339, 2012. PMID: 22351680.

11 Meklin J, Eskelinen M, Syrjänen K and Eskelinen M: Gender-specific performance of a diagnostic score in acute appendicitis. In Vivo 34(6): 3687-3703, 2020. PMID: 33144486. DOI: 10.21873/invivo.12217

12 Eskelinen M, Meklin J, Syrjänen K and Eskelinen M: Performance of a diagnostic score in confirming acute cholecystitis among patients with acute abdominal pain. Anticancer Res 40(12): 6947-6956, 2020. PMID: 33288589. DOI: 10.21873/anticancer.14719

13 Eskelinen M, Meklin J, Syrjänen K and Eskelinen M: Pediatric acute appendicitis score in children with acute abdominal pain (AAP). Anticancer Res 41(1): 297-306, 2021. PMID: 33419824. DOI: 10.21873/anticancer.14776

14 Eskelinen M, Meklin J, Syrjänen K and Eskelinen M: A diagnostic score (DS) is a powerful tool in diagnosis of acute appendicitis in elderly patients with acute abdominal pain. Anticancer Res 41(3): 1459-1469, 2021. PMID: 33788738. DOI: 10.21873/anticancer.14904

15 Eskelinen M, Meklin J, Syrjänen K and Eskelinen M: A diagnostic score for acute small bowel obstruction. Anticancer Res 41(4): 1959-1970, 2021. PMID: 33813402. DOI: 10.21873/anticancer.14963

16 Eskelinen M, Meklin J, Syrjänen K and Eskelinen M: A diagnostic score for reliable confirmation of acute renal colic among patients with acute abdominal pain. Anticancer Res 41(6): 3045-3054, 2021. PMID: 34083296. DOI: 10.21873/anticancer.15087

Received June 21, 2021
Revised July 8, 2021
Accepted July 12, 2021