International Surgery Journal
Fagan PVB et al. Int Surg J. 2020 Nov;7(11):3540-3545
http://www.ijsurgery.com
pISSN 2349-3305 | eISSN 2349-2902

DOI: https://dx.doi.org/10.18203/2349-2902.isj20204655

Original Research Article

Validation of the appendicitis inflammatory response score in a provincial population

Paul V. B. Fagan[1]*, Brad Stanfield[1], Olga Korduke[1], Nigel Henderson[1], Karl Kodeda[1,2]

[1]Department of General Surgery, Taranaki DHB, Taranaki, New Zealand
[2]Sahlgrenska Academy, Gothenburg, Sweden

Received: 21 August 2020
Revised: 13 October 2020
Accepted: 15 October 2020

***Correspondence:**
Dr. Paul V. B. Fagan,
E-mail: pvbfagan@gmail.com

ABSTRACT

Background: The appendicitis inflammatory response (AIR) score is a high performing, and easy to use clinical prediction tool for the evaluation of appendicitis, but its efficacy has not been studied in the provincial setting. This retrospective, single centre study aims to validate the AIR score, estimate the effect AIR score-based risk stratification would have on the negative appendicectomy rate and compare it against other well-known clinical prediction tools for appendicitis.
Methods: 425 patients treated with appendicectomy or laparoscopy between 1st January 2015 and December 31st 2017 were retrospectively provided with an AIR score. This score was compared against the final macroscopic and histological results to determine its accuracy in the local population.
Results: The AIR score did not perform as well as in other published series, with an area under the receiver operating characteristic curve (AUC) of 0.836. The AIR score performed favourably in comparison to the Alvarado score (0.761), APPEND score (0.747) and adult appendicitis score (AAS) (0.828).
Conclusions: This study showed that the AIR score has a high accuracy, and validates its use in a provincial setting. AIR score-based management of appendicitis would be expected to reduce non-therapeutic explorations by a minimum of 50%.

Keywords: Clinical prediction rules, Appendicitis, AIR score, Alvarado, APPEND, Adult appendicitis score

INTRODUCTION

Appendicitis can be a challenging diagnosis. With the advent of laparoscopic surgery and improved access to computerised tomography (CT) scans, there is a temptation to overuse these tools to remove diagnostic doubt. Unfortunately, laparoscopic surgery and CT imaging are not without risks. Serious complications are not commonly associated with the removal of a normal appendix, but there is a financial cost, as well as exposure to a general anaesthetic and rarer complications associated with laparoscopy as well as long-term risks of adhesions.[1-6] Routine CT scanning is not feasible in younger patients, due to the risks associated with radiation exposure, and routine ultrasound (USS) does not significantly reduce non-therapeutic operations.[7-9] A number of clinical prediction tools for appendicitis exist as adjuncts to this diagnostic challenge, but none of these were utilised at our centre.[10-13] Despite published negative appendicectomy rates in New Zealand being relatively high from 19-29% only one centre in New Zealand has published on the use of clinical prediction rules (the APPEND score) in the diagnosis of appendicitis.[10,14-16] The AIR score, which was created in 2008 is a high performing, and easy to use clinical prediction tool.[9,11,17] This is a score separating patients with suspected appendicitis into a low probability zone (1-4), which suggests discharge and outpatient management; an

indeterminate zone (5-8) which suggests imaging, observation or diagnostic laparoscopy; and a high probability zone (9-12) which suggests surgical intervention.

METHODS

Study design

This single centre retrospective analysis included all consecutive patients admitted under general surgery at Taranaki base hospital, New Zealand, who received surgery for the evaluation or treatment of appendicitis or right iliac fossa (RIF) pain, in which appendicitis was a suspected diagnosis. The eligible population was identified from a hospital record database search using ICD-10 codes and included all patients who were coded as having received laparoscopy and laparoscopic or open appendicectomy during the study period. All variables of the AIR score were recorded, as well as age, duration of symptoms, gender and preoperative CT and Ultrasound (USS) imaging. Data was also collected from electronic and paper records for the purposes of documenting alternate appendicitis scoring tools for comparison (AAS, Alvarado and APPEND). Once all of the eligible patients were identified, each patient was retrospectively assigned an AIR score on admission, which was then compared to the final histological assessment and macroscopic intraoperative assessment to determine the AIR scores' accuracy in the local population. Appendicitis was defined histologically as the finding of neutrophils extending into the muscularis propria, and not just the lamina propria.[18] The study period was between January 1st 2015 to December 31st 2017 and no clinical prediction rules were in use during this time. Furthermore, no set protocols on the use of preoperative imaging were in use and any use of preoperative imaging with USS or CT prior to diagnostic laparoscopy or appendicectomy was surgeon dependant. Taranaki Base hospital is a provincial hospital with a catchment area covering a population of approximately 110 000. The closest alternate acute surgical services are at least 3 hours by road and thus the study cohort can be considered population based.

Score assignment

The investigators used the admission form to record all variables of the AIR score on admission (Table. 1).

The AIR score has only one subjective variable, 'abdominal defence'.[11] This variable was standardised between the investigators to avoid confusion after discussion with the surgical department. Definitions were agreed on to determine what examination findings would be classed as none, light, moderate and strong for the purposes of assigning a score.

Score zero (or none) was defined as soft abdominal tenderness that did not limit deep and adequate palpation, or tenderness or guarding that was variable or not reproducible during the exam.

Score 1 (or mild) was defined as tenderness which prevents deep/adequate palpation or voluntary guarding, but minimal or no percussion tenderness.

Score 2 (or medium) was defined as moderate involuntary guarding, some percussion or rebound tenderness, but not as severe as described in score 3.

Score 3 (or strong) was defined as any rigidity, any localized or general peritonism, severe tenderness with strong guarding preventing deep palpation, any gross percussion tenderness or gross rebound tenderness.

Two of the investigators separately assigned scores for each patient based on the admission documents during which time they were blinded to the final histology. The scores were compared and any discrepancy between the scores was settled by a third investigator who was also blinded to the final histology.

Total score is given out of 12.

Table 1: AIR score.

AIR score	Score/12
Vomiting	1
Pain in right iliac fossa (RIF)	1
Muscular defence or rebound tenderness	
Light	1
Medium	2
Strong	3
Body temperature >38.5^0 C	1
Neutrophil percentage (%)	
70-84.9	1
≥85	2
White cell count (×10^9/l)	
10-14.9	1
≥15	2
CRP (g/l)	
10-49.9	1
≥50	2

Exclusions

Laparoscopy that was not for the purposes of treatment of appendicitis or evaluation of RIF pain was excluded, such as staging laparoscopy, and any elective laparoscopic procedure, as well as elective appendicectomy. Patients in which insufficient documentation was available to create the AIR score were also excluded (Figure 1).

Statistical analysis

Data collection and analysis were performed in Microsoft excel MSO 2016 and further statistical analysis was performed by MedCalc statistical software version 18.3

(MedCalc Software, Ostend, Belgium; https://www.medcalc.org) for receiver operating characteristic curves as well as sensitivity, specificity and predictive values.

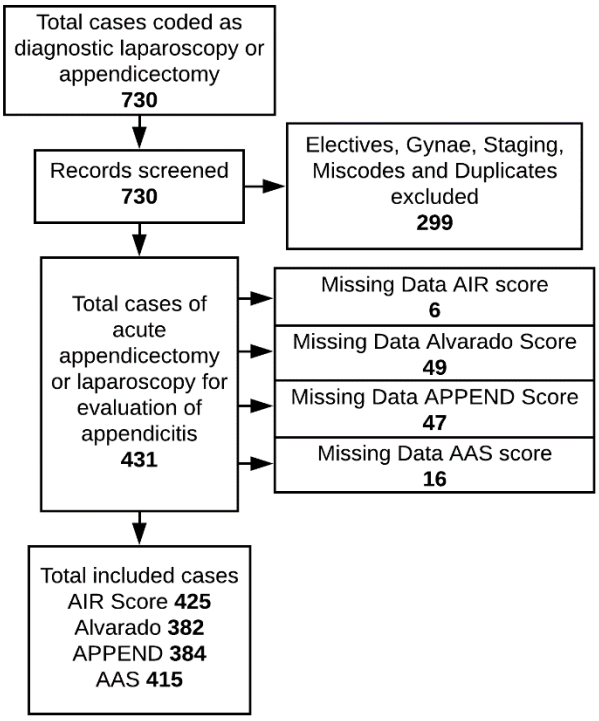

Figure 1: Exclusions for missing data.

RESULTS

425 patients who received appendicectomy or diagnostic laparoscopy under general surgery over three years from 1st January 2015 to 31st December 2017 were included. There were 212 males and 213 females. The age range was from 2 to 92 years, with a median age of 27 (Q_1 18, Q_3 43).

38/425 cases (8.9%) of appendicitis were perforated. 57/425 cases (13.4%) were necrotic. The overall negative appendicectomy rate over this period was 109/425 (25%). Data required to provide a score was already prospectively recorded at the time of the admission in almost all cases (Figure 1).

High probability zone (9-12)- this zone was highly specific, with an overall specificity of >96%.

Low probability zone (1-4)- this zone had a high sensitivity (Table 2), however it included an overall appendicitis rate of 19/87 (22%). The rate of appendicitis increased proportionally to the score (Figure 2). Nearly all of these cases were inflammatory appendicitis, but also included two cases of necrotic appendicitis and one case of perforated appendicitis over the 3-year period.

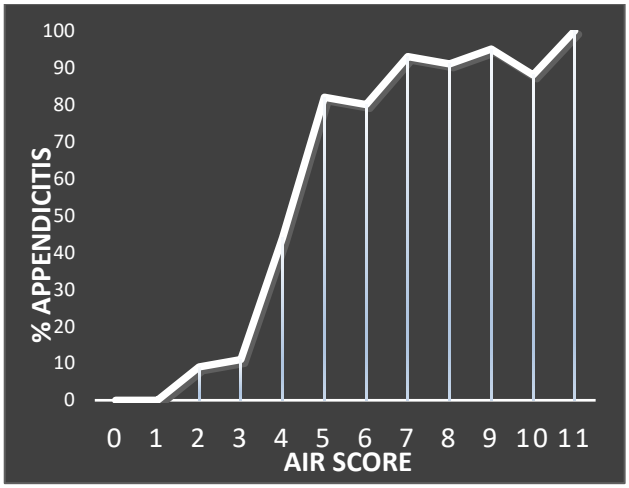

Figure 2: AIR score plotted against rate of appendicitis.

Secondary analysis was performed on the 87 patients in the low probability zone, with a repeat score calculated for these patients the day after admission. Patients who had same day surgery, patients who had insufficient data to re-score the following day, and patients treated with admission antibiotics were excluded from this secondary analysis.

The secondary analysis included 42 patients. The score reduced or remained the same in 36/37 (97%) of patients without appendicitis. In contrast, in 5/5 (100%) of cases with histologically confirmed appendicitis, the score increased the following day.

Analysis of patients who were excluded due to antibiotic treatment was carried out. In 4/4 (100%) of cases the score either remained the same or reduced. This included 2 cases of necrotic appendicitis and these cases became indistinguishable from non-appendicitis using the AIR score, due to the reduction of the score in this group.

Comparison analysis

The AIR score was compared against other clinical prediction tools, including the Alvarado score the APPEND score and the adult appendicitis score (AAS).[10,12,13] The excluded data was higher in these comparison analyses, with 49, 47 and 16 patients excluded respectively (due to missing or incomplete documentation). The area under the receiver operating characteristic curve (AUC) for the AIR score (Figure 3), was higher than the Alvarado and APPEND scores, and similar for the AAS (Table 3). The receiver operating characteristic (ROC) curve for the AIR score was also repeated after the data exclusions were applied, for direct comparison, and the AUC remained essentially unchanged (Table 3).

Table 2: AIR score sensitivities and specificities.

AIR˜ score	Appendicitis/Total (%)	Sensitivity	95% CI#	Specificity	95% CI#
≥0	0/2 (0)	100	98.8-100.0	0	0.0-3.3
>1	0/4 (0)	100	98.8-100.0	5.5	2.0-11.6
>2	2/22 (9)	99.4	97.7-99.9	23.9	16.2-33.0
>3	3/27 (11)	98.4	96.3-99.5	45.9	36.3-55.7
>4	14/32 (44)	94	90.7-96.3	62.4	52.6-71.5
>5	55/67 (82)	76.5	71.4-81.1	73.4	64.1-81.4
>6	57/71(80)	58.4	52.8-63.9	86.2	78.3-92.1
>7	64/69 (93)	38.1	32.7-43.7	90.8	83.8-95.5
>8	63/69 (91)	18.1	14.0-22.8	96.3	90.9-99.0
>9	39/41 (95)	5.7	3.4-8.9	98.2	93.5-99.8
>10	14/16 (88)	1.3	0.3-3.2	100	96.7-100
>11	4/4 (100)	0	0.0-1.2	100	96.7-100

#Confidence interval, ˜Appendicitis inflammatory response score.

Table 3: Comparison of clinical prediction scores.

Score	APPEND	Alvarado	AAS†	AIR˜
Total patients in analysis	384	382	415	425
Appendicitis (%)	286 (74.5)	284 (74.3)	309 (74.5)	318 (74.8)
Area under ROC* curve (95% CI#) (cf. AIR˜ Score)	0.747 (0.701-0.790) (0.834)	0.761 (0.715-0.803) (0.834)	0.828 (0.788-0.863) (0.831)	0.836 (0.797-0.869)

†Adult Appendicitis Score, *Receiver Operator Curve, #Confidence Interval, ˜Appendicitis Inflammatory Response Score

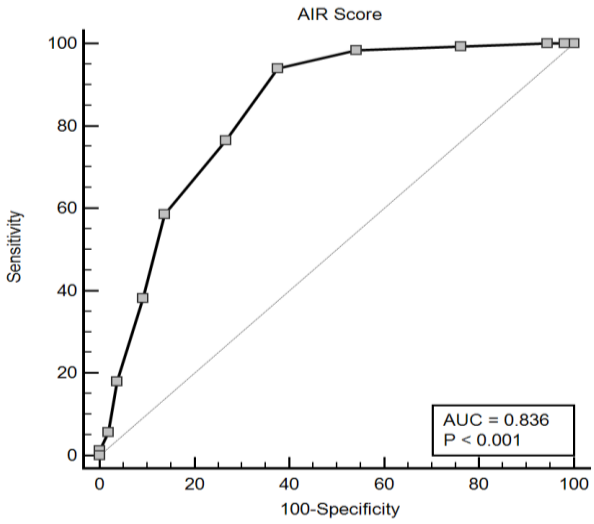

Figure: 3 AIR score ROC curve.

CT imaging

118 patients who underwent a pre-operative CT were found to have a histological diagnosis of appendicitis. 9 of these cases were reported as normal or equivocal on CT (92.4% sensitivity). The largest proportion of the 124 CTs requested over the study period were performed in the over 50 age group: 74/124 (60%). However, a significant proportion were performed in the under 50s group, with 50/124 receiving a CT (40%). 30 CTs (24%) were performed in patients under the age of 40 and 23 (19%) were performed in those under the age of 30. Only 5 patients in the high probability zone under the age of 50 received a pre-operative CT scan.

Ultrasound imaging

USS was performed preoperatively in 39 patients (34 female and 5 male). The appendix was positively identified in 7 patients (18%). The appendix was not seen in 30/39 (77%) of patients of which 15/30 (50%) of patients were subsequently found to have appendicitis.

DISCUSSION

The AIR score was particularly specific in the high probability zone (9-12) at 96%, which is similar to previously published data. [9,11,17,19] Furthermore, a higher indeterminate result was associated with a much higher risk of appendicitis (>90% for score of 7 and 8).

On average, the negative appendicectomy rate was >25% over our series, which is similar to published in data in NZ which varies from 19.5%-29%. [10,14–16,20]

If the negative appendicectomies from the low probability zone were excluded from the analysis, the resulting negative appendicectomy rate would be 42/338 (12.4%). Therefore, in future if we were to avoid operating on patients in the low probability zone, the negative appendicectomy rate would be expected to halve immediately. With the use of judicious imaging and observation within the intermediate probability zone, the

negative appendicectomy rate could potentially halve again.[9]

In contrast to other studies the AIR score did not perform as well, with an AUC of only 0.836. However, it performed favourably compared to other clinical prediction tools. [9,11,17,19]

In the low risk zone, the AIR score did not perform as well as expected and would be predicted to slightly delay diagnosis in roughly one in five cases of appendicitis. However the bulk of these cases were phlegmonous appendicitis and a short delay would be unlikely to change overall outcomes.[16,21–23] In the 1-4 zone, advanced appendicitis was rare, with only two cases of gangrenous appendicitis and only one case of perforated appendicitis in this series. These numbers were still higher than the investigators expected and as such, we would not expect this score to be used to reduce admissions in our centre, except in patients scoring 2 or less. In a centre that could support adequate outpatient management with repeat bloods and examination, the AIR score would likely perform much better at reducing admissions. At our centre however, we would advocate admission and observation for all patients with a high clinical suspicion for appendicitis who were scoring 3 or higher. Outpatient management can still be considered in the provincial setting in motivated patients or those who can be reassessed as an outpatient. Lastly, the AIR score is intended to be repeated every 4-8 hrs in the indeterminate zone, and if this same approach was applied at our centre to patients with a score of 3-4, secondary analysis suggests that the AIR score becomes highly accurate in this zone at 6-24 hours and can facilitate early discharge.[9]

It is important to acknowledge that the use of antibiotics potentially interferes with the accuracy of the AIR score in the low probability zone and we would advise avoiding antibiotics for AIR score-based management, unless a decision has been made to go to surgery or antibiotics are to be the primary treatment.

One limitation of this study is that a retrospective analysis was used to assess the validity of a prospective scoring tool.[11]A further limitation is that only patients who received surgery were included. We expect that this would critically skew the low probability zone comparative to prospective scoring, as a higher proportion of appendicitis would be present compared to all patients presenting with right iliac fossa pain. This likely explains the high relative percentage of appendicitis in the low probability zone, compared to what would be expected with prospective scoring. However, by virtue of focusing exclusively on post-operative cases and concentrating solely on patients in which appendicitis was unable to be excluded clinically, this study illustrates the real-world value of the AIR score as an adjunct to clinical examination in reducing non therapeutic operations. We also acknowledge that 84% of patients in the over 50 group received confirmatory imaging prior to appendicectomy. This is a reasonable management option in older patients and would explain the low rates of unexpected alternate pathology on laparoscopy in this series. However, for younger patients, the AIR score is a superior option, due to the higher risks of radiation attributable cancer in younger patients and with its specificity approaching that of CT imaging in the high probability zone.[7,24] Given that 40% of CTs were performed in the under 50 age group in this series, the AIR score would also be useful for reduction of unnecessary CT imaging and radiation exposure.

It must also be acknowledged that our hospital has significant limitations to out of hours ultrasound (on weekends and after 4 pm Monday to Friday) and often high competing demand for USS imaging and thus has the tendency to a high rate of diagnostic laparoscopy as the initial intervention. We accept that the large predicted reduction in the negative appendicectomy rate AIR score-based management may be exaggerated in a provincial centre with these sorts of limitations as opposed to a tertiary centre with 24-hour access to ultrasonography where the initial non therapeutic laparoscopy rate would be expected to be lower.

Despite the limitations above, this data validates the AIR score for use in a provincial setting, and allowing for the potential bias of listwise exclusion for missing values, the AIR score is not only significantly better than clinical diagnosis alone, but also compares favourably to the Alvarado score, the APPEND score and the AAS. The very low exclusion rate for missing data shows that the variables of the AIR score are already being prospectively collected, and this information just has to be put to good use. The scores can be assigned to patients without additional resources or any changes to the existing process, and given the established safety of in hospital delays in appendicitis, its use comes with minimal risk to patients.[16,21-23]

CONCLUSION

This study improves understanding on the use of clinical prediction rules in appendicitis and is the first study to directly compare the AAS, AIR, Alvarado and APPEND scores in the same patient population. The AIR score was found to be superior to the Alvarado and APPEND scores and although it was similar to the AAS in terms of accuracy, the AIR score is considerably less complex to calculate. This study adds to the growing body of evidence that the AIR score is an easy and effective adjunct to the clinical diagnosis of appendicitis and can be safely and simply applied at point of care, with no additional changes to the hospital infrastructure, staffing or admissions processes.

ACKNOWLEDGEMENTS

Authors would like to thank to professor Roland Andersson for his valued input into this study including

advice on interpretation of the AIR score and clarification of the interpretation of abdominal defence, and for his relentless efforts to improve the care of patients with appendicitis. Many thanks are owed to Dr. Brittany Aldridge with her assistance with data collection.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Narayanan A, Sundararaman S, Varadhan L, Rajput R, Gupta V, ReayJones N. What is negative about negative appendicectomy rates? An experience from a district general hospital. Int Surg J. 2015;2(2):161.
2. Angenete E, Jacobsson A, Gellerstedt M, Haglind E. Effect of Laparoscopy on the Risk of Small-Bowel Obstruction. Arch Surg. 2012;147(4):359-64.
3. Swank HA, Mulder IM, La Chapelle CF, Reitsma JB, Lange JF, Bemelman WA. Systematic review of trocar-site hernia. Br J Surg. 2012;99(3):315-23.
4. Okabayashi K, Ashrafian H, Zacharakis E, Hasegawa H, Kitagawa Y, Athanasiou T et al. Adhesions after abdominal surgery: a systematic review of the incidence, distribution and severity. Surg Today. 2014;44(3):405-20.
5. Lee M, Paavana T, Mazari F, Wilson TR. The morbidity of negative appendicectomy. Ann R Coll Surg Engl. 2014;96(7):517-20.
6. De Castro SMM, Songun I, Dwars BJ. An unexpected severe complication after a negative laparoscopic appendectomy. Can J Surg. 2009;52(6):295-6.
7. Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med. 2009;169(22):2078-86.
8. Douglas CD, Macpherson NE, Davidson PM, Gani JS. Randomised controlled trial of ultrasonography in diagnosis of acute appendicitis, incorporating the Alvarado score. BMJ. 2000;321:919-22.
9. Andersson M, Kolodziej B, Andersson RE, STRAPPSCORE Study Group. Randomized clinical trial of Appendicitis Inflammatory Response score-based management of patients with suspected appendicitis. Br J Surg. 2017;104(11):1451-61.
10. Mikaere H, Zeng I, Lauti M, Kularatna M, MacCormick AD. Derivation and validation of the APPEND score: an acute appendicitis clinical prediction rule. ANZ J Surg. 2018;88(4):E303-7.
11. Andersson M, Andersson RE. The appendicitis inflammatory response score: A tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World J Surg. 2008;32(8):1843-9.
12. Sammalkorpi HE, Mentula P, Leppäniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis - a prospective study. BMC Gastroenterol. 2014;14(1):1-8.
13. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15(5):557-64.
14. de Burlet KJ, Crane G, Cullinane R, Larsen PD, Dennett ER. Review of appendicectomies over a decade in a tertiary hospital in New Zealand. ANZ J Surg. 2018;88(12):1253-7.
15. Pillai S, Hsee L, Pun A, Mathur S, Civil I. Comparison of appendicectomy outcomes: Acute surgical versus traditional pathway. ANZ J Surg. 2013;83(10):739-43.
16. Omundsen M, Dennett E. Delay to appendicectomy and associated morbidity: A retrospective review. ANZ J Surg. 2006;76(3):153-5.
17. Castro SMMD, Ünlü Ç, Steller EP, Van Wagensveld BA, Vrouenraets BC. Evaluation of the appendicitis inflammatory response score for patients with acute appendicitis. World J Surg. 2012;36(7):1540-5.
18. Carr NJ. The pathology of acute appendicitis. Ann Diagn Pathol. 2014;18(4):46-58.
19. Scott AJ, Mason SE, Arunakirinathan M, Reissis Y, Kinross JM, Smith JJ. Risk stratification by the Appendicitis Inflammatory Response score to guide decision-making in patients with suspected appendicitis. Br J Surg. 2015;102(5):563-72.
20. Scrimgeour DSG, Driver CP, Stoner RS, King SK, Beasley SW. When does ultrasonography influence management in suspected appendicitis? ANZ J Surg. 2014;84(5):331-4.
21. van Dijk ST, van Dijk AH, Dijkgraaf MG, Boermeester MA. Meta-analysis of in-hospital delay before surgery as a risk factor for complications in patients with acute appendicitis. Br J Surg. 2018;105(8):933-45.
22. Bhangu A, Singh P, Panagiotopoulou IG, Chatzizacharias N, Rana MM, Rollings Ket al. Safety of short, in-hospital delays before surgery for acute appendicitis: Multicentre cohort study, systematic review, and meta-analysis. Ann Surg. 2014;259(5):894-903.
23. Li J, Xu R, Hu DM, Zhang Y, Gong TP, Wu XL. Effect of Delay to Operation on Outcomes in Patients with Acute Appendicitis: a Systematic Review and Meta-analysis. J Gastrointest Surg. 2019;23(1):210-23.
24. Yun SJ, Ryu CW, Choi NY, Kim HC, Oh JY, Yang DM. Comparison of Low- and standard dose CT for the diagnosis of acute appendicitis: A meta-analysis. Am J Roentgenol. 2017;208(6):W198-207.

Cite this article as: Fagan PVB, Stanfield B, Korduke O, Henderson N, Kodeda K. Validation of the appendicitis inflammatory response score in a provincial population. Int Surg J 2020;7:3540-5.