

Does Serum Bilirubin Add Value in the Diagnosis of Appendicitis? A Multi-Institution Analysis of 2024 Patients

Andrew Phillip Maurice (✉ andrew.maurice@uqconnect.edu.au)

Royal Brisbane and Women's Hospital

Benjamin David Swinson

Royal Brisbane and Women's Hospital

Eduardo Miguel Apellaniz

The University of Queensland

Sarah Emily King

The University of Queensland

Thomas Grant Mackay

Logan Hospital

Matthew John Burstow

Logan Hospital

Terence Chuan Chua

Logan Hospital

Bernard Mark Smithers

The University of Queensland

Harsheet Sethi

The University of Queensland

Research Article

Keywords: Negative appendicectomies, complicated appendicitis, Hyperbilirubinaemia, neutrophil

Posted Date: October 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-885174/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

BACKGROUND

Even with modern diagnostics, appendicitis can be difficult to diagnose accurately. Negative appendicectomies (NA) and delayed diagnosis of complicated appendicitis (CA, i.e. perforation or abscess) remain common. Serum bilirubin has been proposed as an additional biomarker to assist with the diagnosis of appendicitis. In this large series, we assessed the value of bilirubin in the diagnosis of appendicitis.

METHODS

A retrospective review of patients with suspected appendicitis at three hospitals over a three year period was performed. All consecutive patients with appendicectomy were included. In addition, a “discharged” sub-group of consecutive patients who were admitted with suspected appendicitis but discharged without an operation was also identified.

Demographic data, presence of fever, tachycardia, total white cell count (WCC), neutrophil count, total serum bilirubin, operative findings and final histology were recorded. Multivariate logistic regression was performed to determine independent predictors of appendicitis and CA. Receiver-operator analysis was performed to compare bilirubin to WCC and neutrophil count.

RESULTS

There were 2024 patients: 1167 had uncomplicated appendicitis, 355 had CA and 303 underwent NA. 200 non-surgical “discharged” patients were included for comparison. Compared to those without appendicitis (NA and discharged groups), increased serum bilirubin was associated with an increased likelihood of appendicitis (OR 1.030 (1.013, 1.048), $p < 0.0001$) and increased likelihood of CA (OR 1.035, 95% CI (1.021, 1.050), $p < 0.001$). These results remained significant when the discharged group, NA group and uncomplicated appendicitis groups were analyzed separately.

The sensitivity and specificity of bilirubin was inferior to neutrophil count for the diagnosis of appendicitis (AUC 0.657 versus 0.725, $p < 0.0001$). Bilirubin, WCC and neutrophils each were all relatively insensitive and non-specific over a variety of cut-off values and combinations did not improve their accuracy.

CONCLUSION

Hyperbilirubinaemia was independently associated with an increased likelihood of both uncomplicated and complicated appendicitis, however had similar sensitivity and specificity when compared to WCC or neutrophils. Bilirubin, neutrophil count and WCC alone are not discriminatory enough to be used in isolation but may be incrementally useful adjuncts in pre-operative assessment of patients with suspected appendicitis.

Introduction

Appendicitis is a common condition, with a lifetime incidence of 7-8% in developed countries(1, 2). In the modern era, appendicitis can still be difficult to diagnose accurately. In Australia, despite the use of imaging in up to 70% of patients undergoing appendicectomy, 19% of all appendicectomy specimens are “negative appendicectomies” that do not demonstrate inflammation(3). National audits demonstrate similar negative appendicectomy rates in other developed countries(4). In addition, a delay to diagnosis has been correlated with the development of complicated appendicitis (perforation, abscess and phlegmon) with associated increased morbidity(5).

Routine laboratory investigations in the work up of patients, for suspected appendicitis, include white blood cell count (WCC) and neutrophil count. In addition, many other serum and urinary biomarkers have been assessed, aiming to improve the diagnostic accuracy and assessment of severity of appendicitis, including CA-125(6), procalcitonin(7), bilirubin and calprotectin(8). Hyperbilirubinaemia is observed in many inflammatory states and bilirubin is increasingly used as a cheap, readily available and ubiquitous marker for illness severity in many hospital settings(9, 10) with its role in the diagnosis of appendicitis, having been examined in several series and a systematic review(11, 12).

Most previous studies that have assessed the role of bilirubin in diagnosing appendicitis have been small, with limited multivariate statistical analysis to account for relevant confounding factors. Most series have focused on the use of bilirubin to differentiate between uncomplicated and complicated appendicitis, excluding negative appendicectomy and those without appendicitis. The aim of this study was to assess the independent diagnostic value of serum bilirubin to other established diagnostic criteria for both complicated and uncomplicated appendicitis.

Methods

A retrospective review was performed of patients with suspected appendicitis at three large tertiary hospitals with 24-hour acute surgical services over a three-year period from January 2016 to December 2019. All consecutive operative cases were identified through searching operative theatre software records, assessing for all cases of laparoscopy or appendicectomy, in addition to searching surgical team daily patient lists for all patients with appendicitis. Non-operative patients were identified through searching surgical team daily patient lists. Ethical approval for this study was sought and obtained from the RBWH Human Research Ethics Committee.

Data was extracted from the electronic medical records including patient demographics, pregnancy status (urine or serum β -hCG), observations, laboratory findings at time of admission (specifically white blood cell count, neutrophil count, total serum bilirubin levels), radiological findings, comorbidities, operation notes and histological findings of the removed appendix. The highest recorded heart rate at admission and any documented temperature of 37.8°C or above were noted. Due to limitations in recording and software, heart rate and fever were recorded as dichotomized categorical variables

(tachycardia was defined as heart rate above 100, fever was defined as a temperature of 37.8°C or above). Non-operative patients with uncomplicated and complicated appendicitis were included provided there was clear radiological evidence of the diagnosis on Computed Tomography (CT). Laboratory investigations of interest were those taken at presentation of the patient to the emergency department, prior to surgery.

Patients who underwent surgery were divided into three groups: those with no histological evidence of appendicitis (negative appendicectomy), those with complicated appendicitis and those with uncomplicated appendicitis. Complicated appendicitis was defined as perforation or abscess cavity encountered intra-operatively, or as clearly described in the histological report. For those with non-operative management, complicated appendicitis was defined as CT findings consistent with complicated appendicitis (such as pneumoperitoneum or abscess). Surgical patients were excluded from analysis if histology demonstrated alternate pathology other than acute appendicitis or negative appendicectomy (e.g. parasitic worms, tumours/adenomas) or if the patients were pregnant, had Gilbert's disease or any known liver disease.

Although those with negative appendicectomy by definition do not have appendicitis, these patients may differ from those ultimately discharged without an operation, as the pre-surgery probability of having appendicitis was sufficiently high enough for the surgeon to offer surgery. Thus, surgical admission lists were searched to identify a fourth group of consecutive patients (referred to as the "discharged" group) who had been referred and admitted under the surgical team with suspected appendicitis but were subsequently discharged without an operation. Based on the number of surgical patients identified in this study, to detect a one-sided, mean difference in bilirubin of 3mmol/L or higher with 80 percent power, only 60 discharged patients would be required, however a larger number of 200 was chosen to ensure sufficient power for multivariate analysis(13). These were patients aged 14 and older and included those who subsequently underwent radiological investigations that were found to be negative or underwent clinical observation only and did not undergo appendicectomy within a year of this admission. These patients were only included if there was clear documentation from the surgical staff that appendicitis was suspected at the time of admission.

Three major comparisons were performed: the first comparison to determine the risk factors associated with appendicitis (both complicated and uncomplicated) compared to patients without appendicitis (combining negative appendicectomy with the discharged group). To ensure the inclusion of complicated appendicitis did not skew the results, a sub-analysis was performed excluding the complicated appendicitis patients.

The second major comparison was between those with appendicitis, to those who were discharged (i.e. excluding negative appendicectomy). Furthermore, differences in baseline characteristics were evaluated between the negative appendicectomy and discharged groups as these groups might differ.

The third major comparison was within the subgroup of those with histologically proven appendicitis only, determining the risk factors associated with complicated appendicitis versus uncomplicated

appendicitis.

Statistical analysis was performed in *STATA* version 15.1. Univariate analysis was performed to ascertain differences in the means of the continuous covariates (as these were approximately normally distributed, a two-sample T-test for equivalent means was used) and proportions for categorical variables (using the Chi-squared test) for each of the two comparisons listed. Multivariate logistic regression analyses were performed to ascertain the independent predictors of a negative appendicectomy and of complicated appendicitis, specifically adjusting for demographics (age and gender), fever and tachycardia.

The diagnostic accuracy of bilirubin, WCC and neutrophils were evaluated using receiver operator curve (ROC) analysis and area under the curves (AUC) were ascertained and compared statistically using the *roccomp* program in *STATA*. To aid with interpretation of this analysis, we assessed the sensitivity, specificity, positive and negative predictive values at different cutoff values. The lower limit for detection of bilirubin was 4 and for WCC and neutrophils it was 0.

To evaluate if combinations of biomarkers improved results, combinations of cut-off points for biomarkers were evaluated for sensitivity and specificity.

Neutrophil and white cell counts are given as count x 10⁹/liter. Bilirubin level refers to the total bilirubin in micromoles per liter (μmol/L).

Results

Of 2121 patients identified, 2024 were included for analysis and 97 were excluded. Of the patients included for analysis, 1167 had uncomplicated appendicitis, 355 had complicated appendicitis, 303 underwent a negative appendicectomy and an additional 200 patients were in the comparison discharge group. Of the 97 excluded patients, 19 patients were excluded with tumors or adenomas of the appendix. 36 pregnant patients were excluded. 42 patients with alternative diagnoses were excluded (e.g. intestinal worms).

Demographics, clinical findings and serum biomarker levels are given for all patients as well as those with and without appendicitis (i.e. negative appendicectomy and discharged group combined) in Table 1, in addition to the results of the univariate and multivariate analysis.

Table 1
Appendicitis compared to non-Appendicitis in patients who had an appendectomy

	All patients (n = 2024)	Appendicitis (n = 1521)	Non- Appendicitis (n = 503)	Univariate p-value	OR, OR 95% CI, multivariate p-value
Age (mean years)	31.2 ± 16.0	32.2 ± 16.6	28.1 ± 13.6	<0.0001	1.011, (1.003, 1.020), p = 0.005
Female Gender	1096 (54%)	733 (48%)	363 (72%)	<0.0001	0.431 (0.339, 0.550), p<0.0001
Fever	365 (18%)	307 (20%)	58 (12%)	<0.0001	1.16 (0.828, 1.165), p = 0.375
Tachycardia (above 100)	455 (23%)	369 (25%)	86 (17%)	0.001	1.015 (0.849, 1.56), p = 0.365
Bilirubin	16.9 ± 8.9	17.9 ± 9.2	13.7 ± 7.0	<0.0001	1.030 (1.013, 1.048), p<0.0001
WCC	12.2 ± 4.6	13.0 ± 4.6	9.9 ± 4.0	<0.0001	0.960 (0.876, 1.053), p = 0.390
Neutrophils	9.4 ± 4.6	10.2 ± 4.5	6.8 ± 4.0	<0.0001	1.233 (1.120, 1.358), p < 0.0001

Caption: Demographics, clinical characteristics and mean serum biomarker levels for all patients, those admitted with suspected appendicitis but discharged without an operation. Means with ± one standard deviation are listed for continuous variables. Number of patients and proportion in brackets are given for dichotomous variables. For the multivariate analysis, odds-ratios greater than 1 indicates increased risk of appendicitis. WCC White Cell Count.

Appendicitis versus Non-Appendicitis

Patients with appendicitis were compared to those without appendicitis (i.e., a combination of the negative appendectomy and discharged without surgery groups). In the univariate analysis, compared to patients with negative appendectomy, patients with appendicitis were more likely to be male (p<0.0001), older (p<0.0001), febrile (p <0.0001), tachycardic (p = 0.001) and have higher levels of bilirubin (p<0.0001), WCC (p<0.0001) and neutrophils (p<0.0001). In the multivariate analysis, only male gender (OR 0.431, p<0.0001), older age (OR 1.011, p<0.0001) and increased neutrophils (OR 1.233, p<0.0001) and bilirubin (OR 1.030, p>0.0001) were independent predictors of appendicitis.

To ensure the inclusion of complicated appendicitis patients did not skew the above results, the multivariate analysis was repeated comparing non-appendicitis (n = 503) to uncomplicated appendicitis alone (n=1167), excluding complicated appendicitis. Compared to the non-appendicitis group, the independent predictors of uncomplicated appendicitis were male gender (OR 0.4495, 95% CI (0.350,

0.574), $p < 0.0001$), increased neutrophils (OR 1.200, 95% CI (1.089, 1.323), $p < 0.001$) and increased bilirubin level (OR 1.025, 95% CI (1.008, 1.043), $p = 0.005$). Age was no longer significant ($p = 0.302$). These results show that hyperbilirubinaemia is also an independent predictor of uncomplicated appendicitis.

Receiver operator curve analysis (Figure 1) demonstrated increased neutrophil count (AUC 0.725) to be the most accurate biomarker, for appendicitis, compared to increased WCC (AUC 0.708, $p < 0.0001$) and increased bilirubin (AUC 0.657, $p = 0.0001$). Bilirubin and WCC did not have significantly different AUC ($p = 0.28$). Therefore, all biomarkers are correlated with an increased likelihood of appendicitis, however increased neutrophil count was the most sensitive and specific test to differentiate appendicitis from those without appendicitis.

Appendicitis versus Discharged Group (Excluding Negative Appendicectomy)

Compared to the group of patients discharged without an operation, the negative appendicectomy group were younger (mean age 25.8 versus 31.5, $p < 0.0001$), febrile at presentation (15% versus 6%, $p = 0.002$) and tachycardic (21% versus 12%, $p = 0.013$), There were no significant differences in the neutrophils, WCC, bilirubin or proportion of patients who were female.

Univariate and multivariate analyses were performed comparing patients with appendicitis with the discharged group, excluding the negative appendicectomy patients (Table 2). On univariate analysis, the bilirubin was higher in the appendicitis group, (mean 17.9 versus 14.4, $p = 0.0002$) however, this was not significant in the multivariate analysis (Table 2).

Table 2

Patients with pathologically confirmed appendicitis compared to patients discharged without surgery

	Appendicitis (n = 1521)	Discharged (n = 200)	Univariate p-value	OR, OR 95% CI, multivariate p-value
Age (mean years)	32.2 ± 16.6	31.5 ± 14.1	0.575	0.997, (0.987, 1.001), p = 0.586
Female Gender	733 (48%)	151 (76%)	<0.0001	0.353 (0.247, 0.505), p<0.0001
Fever	307 (20%)	12 (6%)	<0.0001	2.59 (1.38, 4.86), p < 0.0001
Tachycardia (above 100)	369 (25%)	24 (12%)	<0.0001	1.403 (0.870, 2.263), p = 0.165
Bilirubin	17.9 ± 9.2	14.4 ± 7.3	<0.0001	1.013 (0.991, 1.037), p = 0.241
WCC	13.0 ± 4.6	9.8 ± 4.0	<0.0001	*
Neutrophils	10.2 ± 4.5	6.9 ± 3.9	<0.0001	1.187 (1.135, 1.241) p < 0.0001

Caption: The negative appendicectomy group were more likely to be tachycardic, febrile and younger than the discharged group, so analysis was repeated with these patients excluded. WCC White Cell Count.

* Omitted from logistic regression model due to co-linearity with neutrophils.

Uncomplicated versus Complicated Appendicitis

The results of the univariate and multivariate comparison of uncomplicated versus complicated appendicitis are provided in Table 3. In the univariate analysis, all covariates were significant predictors of complicated appendicitis. The multivariate analysis demonstrated increased age (OR 1.036, 95% CI [1.028, 1.045], p<0.001), presence of fever (OR 1.685, 95% CI [1.234, 2.300], p<0.001), tachycardia (OR 1.611, 95% CI [1.178, 2.201], p<0.003), increased bilirubin (OR 1.035, 95% CI [1.021, 1.050], p<0.001) and increased neutrophils (OR 1.148, 95% CI (1.036, 1.272), p=0.008) were independent predictors of complicated appendicitis.

Table 3
Uncomplicated Compared to Complicated Appendicitis

	Uncomplicated (n= 1168)	Complicated (n=354)	Univariate p-value	Odds Ratio (OR), Multivariate p- value and OR 95% CI
Mean age in years	29.9 ± 0.9	39.9 ± 2.1	<0.0001	1.036 (1.028, 1.045), p<0.001
Female Gender	584 (50.0%)	149 (42.1%)	0.009	0.804 (0.611, 1.059), p = 0.120
Fever	194 (16.7%)	113 (32.9%)	<0.0001	1.685 (1.234, 2.300), p < 0.001
Tachycardia	248 (21.4%)	121 (35.2%)	<0.0001	1.611 (1.178, 2.201), p < 0.003
Serum total bilirubin	16.74 ± 0.47	21.75 ± 1.19	<0.0001	1.035 (1.021, 1.050), p < 0.001
WCC	12.5 ± 0.25	14.61 ± 0.51	<0.0001	0.958 (0.868, 1.058), p = 0.400
Neutrophils	9.23 ± 0.64	12.05 ± 0.5	<0.0001	1.148 (1.036, 1.272), p = 0.008

Caption: Characteristics of those with uncomplicated versus complicated appendicitis, with univariate and multi-variate comparison. Baseline characteristics for all patients with appendicitis (n = 1521) have already been given in the second column of Table 1. Means with ± one standard deviation given for continuous variables. Number of patients and proportion in brackets are given for dichotomous variables. An odds ratio of greater than 1 indicated increased odds of complicated appendicitis.

Receiver operator analysis (Figure 2) demonstrated that the AUC for bilirubin (0.6451) and neutrophils (0.6455) were higher compared to WCC (0.621, p<0.001), however the AUC for neutrophils was not statistically significantly different to the AUC for bilirubin (p=0.97). Thus, the sensitivity and specificity of bilirubin for the detection of complicated appendicitis was no better or worse than neutrophils.

Diagnostic Accuracy and Cut-Off Values

Table 4 and Table 5 provides a list of sensitivities, specificities, positive predictive values (PPVs) and negative predictive values (NPVs) for each of the three biomarkers at different cut-off levels for appendicitis versus non-appendicitis and uncomplicated versus complicated appendicitis. The sensitivity, specificity, PPVs and NPVs were relatively poor over a range of values, mirroring the results from the ROC analysis.

Table 4
Diagnostic Accuracy of Biomarkers for Appendicitis versus Non-Appendicitis

Appendicitis versus Non-Appendicitis					
	Cut-off value	Sensitivity	Specificity	PPV	NPV
Total Serum Bilirubin	4	0.999	0.0041	0.752	0.576
	5	0.989	0.018	0.752	0.352
	10	0.87	0.270	0.782	0.408
	15	0.591	0.651	0.836	0.345
	20	0.331	0.850	0.869	0.296
	25	0.174	0.935	0.890	0.273
	30	0.1	0.968	0.904	0.263
	35	0.053	0.982	0.899	0.256
WCC	1.5	1	0.002	0.751	1.000
	5	0.984	0.047	0.757	0.493
	10	0.728	0.577	0.838	0.413
	12.5	0.535	0.778	0.879	0.357
	15	0.314	0.900	0.904	0.303
	17.5	0.159	0.947	0.900	0.272
	20	0.070	0.978	0.906	0.259
	25	0.011	0.994	0.847	0.250
Neutrophils	0	1	0	0.751	
	2	0.993	0.020	0.753	0.486
	5	0.860	0.397	0.811	0.485
	7.5	0.709	0.667	0.865	0.432
	10	0.512	0.820	0.896	0.358
	12.5	0.290	0.900	0.897	0.296
	15	0.143	0.960	0.915	0.271
	20	0.025	0.992	0.904	0.252

Caption: Sensitivity and specificity for biomarkers for a representative selection of different cut-off values. The lowest biomarker levels listed were the lowest recorded in the data-set. PPV and NPV are based on prevalence of 75.1% in the sample. PPV Positive Predictive Value, NPV negative predictive value.

Table 5
Diagnostic Accuracy of Biomarkers for Uncomplicated vs Complicated Appendicitis

Uncomplicated vs. Complicated Appendicitis					
	Cut-off value	Sensitivity	Specificity	PPV	NPV
Total Serum Bilirubin	4	1.000	0.001	0.233	1.000
	5	0.997	0.013	0.234	0.936
	10	0.928	0.147	0.248	0.871
	15	0.731	0.452	0.288	0.847
	20	0.497	0.720	0.350	0.825
	25	0.304	0.865	0.406	0.804
	30	0.185	0.926	0.431	0.789
	35	0.119	0.967	0.522	0.784
WCC	2.1	1.000	0.000	0.233	1.000
	5	0.986	0.016	0.233	0.794
	10	0.841	0.306	0.269	0.864
	12.5	0.655	0.501	0.285	0.827
	15	0.430	0.722	0.319	0.807
	17.5	0.254	0.867	0.367	0.793
	20	0.143	0.952	0.474	0.786
	25	0.026	0.993	0.530	0.771
Neutrophils	0	1.000	0.000	0.233	1.000
	2	0.997	0.009	0.234	0.905
	5	0.946	0.166	0.256	0.910
	7.5	0.843	0.332	0.277	0.875
	10	0.670	0.536	0.304	0.843
	12.5	0.425	0.751	0.341	0.812
	15	0.262	0.894	0.428	0.800
	20	0.048	0.983	0.461	0.773

Caption: Sensitivity and specificity for biomarkers for a representative selection of different cut-off values. The lowest biomarker levels listed were the lowest recorded in the data-set. PPV and NPV are

based on prevalence of 30.3% complicated appendicitis. PPV Positive Predictive Value, NPV negative predictive value.

Combinations of biomarkers were assessed to ascertain if there was any improvement in diagnostic accuracy. As the ROC analysis demonstrated that WCC was inferior to bilirubin and neutrophils, only the combination of neutrophils and bilirubin has been performed. Table 6 provides sensitivity and specificity for combined cut-off values over a selected representative range of bilirubin and neutrophils. Compared to Table 4 and 5, combining neutrophils and bilirubin did not markedly improve the sensitivity or specificity.

Table 6
Combinations of Cut-Offs and Sensitivity / Specificity

Cut-Off values	Neutrophils	Appendicitis vs. Non-Appendicitis		Complicated vs. Uncomplicated Appendicitis	
		Sensitivity	Specificity	Sensitivity	Specificity
Bilirubin					
5	5	84.7	42.4	94.1	18.2
	7.5	70.1	68.5	83.9	34.1
	10	51.1	82.7	66.4	53.6
	15	14.7	95.4	26.0	88.7
	20	3.29	98.6	5.4	97.4
10	5	73.0	61.0	85.6	30.7
	7.5	61.0	86.7	76.6	43.7
	10	44.9	87.5	60.7	59.9
	15	13.4	96.7	24.9	90.2
	20	2.89	99.0	4.5	97.6
20	5	29.0	91.6	44.9	75.9
	7.5	25.7	97.0	42.1	79.3
	10	20.0	96.6	33.9	84.2
	15	6.8	98.6	15.3	95.8
	20	2.0	99.2	3.4	98.5
30	5	10.3	96.4	17.8	92.0
	7.5	8.8	99.0	16.7	93.6
	10	7.2	98.6	13.8	94.9
	15	3.0	99.0	6.5	98.1
	20	1.0	99.0	1.4	99.1

Caption: Selected combinations of cut-off values of bilirubin and neutrophils and corresponding sensitivity and specificity for the two comparisons. White cell count was not evaluated as it was the poorest biomarker in the initial analysis.

Discussion

To our knowledge, this is the largest comprehensive series to examine the role of bilirubin in the diagnosis of appendicitis and the first large series to compare those with both complicated and uncomplicated appendicitis to those without appendicitis, including a group of patients admitted with suspected appendicitis who did not undergo surgery. Previous studies have concluded bilirubin is a useful marker in the setting of suspected complicated appendicitis. We have shown the utility of bilirubin in the diagnosis of appendicitis was limited due to poor sensitivity and specificity and it would be unlikely the bilirubin level will influence decision making in the majority of cases.

A 2017 comprehensive meta-analysis reviewed the sensitivity and specificity for the most common biomarkers in appendicitis and included studies until 2016(12). Five of these studies in the meta-analysis had more than 1000 patients (the largest had 1271 patients), with the majority being single-institution studies (14–18). Most results were from univariate analysis with one study (17) performing a limited multivariate analysis including clinical data, evaluating whether the bilirubin level was independently correlated with perforation risk. With respect to sensitivity and specificity, some studies demonstrated superior receiver-operator curves for bilirubin when compared to CRP and WCC; however, other studies demonstrated the opposite. Since this meta-analysis, several further studies have been published; however, the number of patients in these studies have been small(19, 20).

One study assessed the role of the bilirubin level comparing patients with histologic appendicitis to those with a negative appendectomy(14) however none of these studies included patients who were discharged without an operation after a period of observation. Any comparison that focuses on the negative appendectomy group needs to be treated with caution as we have shown that patients with a histologic negative appendectomy differed significantly from those who were discharged as they were more likely to be tachycardic, febrile and younger. This may account for the reason why the surgeon opted to perform appendectomy in these patients rather than discharge them as they had a higher pre-surgery probability of having appendicitis. This highlights that the correct comparison group should include those who were also discharged without surgery. To our knowledge, this is the first large series to examine the discharged group.

The larger series above evaluated the role of bilirubin in distinguishing perforated versus non-perforated appendicitis concluding that increased bilirubin was a marker of perforation. However, our data shows a correlation between an elevated bilirubin with the likelihood of both complicated and uncomplicated appendicitis, thus indicating a problem with specificity for this outcome.

Practical Interpretation of Bilirubin for the Clinician

From a practical perspective, the surgeon evaluating a patient with suspected appendicitis will be interested in what the bilirubin level means and what are optimal “cut-off” values for the diagnosis of appendicitis and complicated appendicitis. Results of ROC analysis unfortunately demonstrate that all biomarkers were relatively insensitive and non-specific over a variety of cut-off values and combinations.

The tables suggest that the biomarkers are only independently useful at extreme low or high values. For instance, a bilirubin level of 4 or less had a 100% negative predictive value for excluding complicated appendicitis. If the bilirubin level is over 30, this is 96.9% specific for appendicitis and 92.6% specific for complicated appendicitis: a high value such as this in a patient with equivocal clinical findings and normal WCC may guide the clinician towards observation, performing CT scan or diagnostic laparoscopy, rather than discharge.

For non-extreme values, although moderately increased levels of any biomarker statistically portends an increased likelihood of both uncomplicated and complicated appendicitis, there are no optimal “cut-off” values that are sufficiently sensitive or specific to use in isolation when evaluating the patient. Thus a moderately increased or normal bilirubin level will be interpreted similarly to a moderately increased or normal white cell count or neutrophil count, in context of other history and examination findings when evaluating a patient with right iliac fossa pain.

Given the quality of medical imaging in 2021, for suspected appendicitis, and the numerous other independent predictors of appendicitis and complicated appendicitis seen in this report, it is difficult to conceive a situation in which the bilirubin would play a pivotal role in the diagnosis of appendicitis in any patient. Although the above insights are interesting from a statistical and pathophysiological standpoint, in our view, the significance of the bilirubin level in suspected appendicitis should not be overemphasized.

Should liver function tests and bilirubin be performed as a matter of routine for all patients with suspected appendicitis? At our institutions, liver function tests and bilirubin are ordered as routine for all patients admitted to the emergency department with abdominal pain, thus our physicians usually have the luxury of having this information. In institutions where ordering bilirubin is not routine, the cost of ordering additional investigations should be balanced against the likelihood that the investigation will change management: the analysis of this series would suggest that ordering bilirubin would not be cost-effective due to the poor diagnostic sensitivity and specificity. A satirical yet informative Australian study(21) demonstrated a marked reduction in ordering C-reactive protein (CRP) if the physician had to pay a small fine (Australian \$1). At our institutions, such a debate is particularly true for CRP, which is not universally ordered for many conditions, where the cost is weighed against the likelihood of management change. Such an issue is beyond the scope of this paper but was addressed in the meta-analysis(12), which concluded that common biomarkers (neutrophils, bilirubin etc.) were cheap but relatively inaccurate. More exotic biomarkers (IL-6, procalcitonin etc.) were more accurate; however, were significantly more expensive, not always available and took more time to process a result.

Strengths and Limitations

This study is strengthened by a multi-institution, large sample size and the most rigorous multivariate analysis on this topic to date with the inclusion of multiple objective clinical covariates to determine the independent predictive value of bilirubin. To our knowledge, this is the only large series to evaluate the role of bilirubin in suspected appendicitis, when compared to patients without appendicitis. Patients with

non-operative management were included and a comprehensive patient identification strategy was utilized to ensure all consecutive patients from each institution as included.

As would be expected, our data indicates that younger age and female sex are less likely to have appendicitis, a finding that has been mirrored in other studies(22, 23). A major strength of this study is the use of multivariate statistics to account for the effect of demographics, fever and presence of tachycardia.

Limitations include the retrospective design and no analysis of C-reactive protein (CRP). Ordering a CRP is not a uniform practice at our institutions thus not all patients had a CRP level, and it was felt performing analysis would be misleading.

Relevance to Current Research and Future Directions

Virtually all facets of appendicitis have come under investigation in contemporary research. Recent studies have investigated methods to reduce unnecessary imaging in the evaluation of right-iliac fossa pain(24), non-operative management of uncomplicated appendicitis(25), the monetary costs of negative appendectomy(26) and the need and timing for surgery in complicated appendicitis(27). All such contemporary issues are reliant on the rapid and reliable diagnosis of appendicitis, of which, an efficacious biomarker would immediately be of value.

In addition to being cheap, rapidly, and readily available, an ideal biomarker for the diagnosis of appendicitis would be sensitive and specific enough to reduce the negative appendectomy rate and increased the timeliness of diagnosis (and thus theoretically reduce the complicated appendicitis rate). Although this series has demonstrated several additional insights into the role of bilirubin in appendicitis, bilirubin unfortunately does not appear to fulfil all these criteria and it is unlikely a further large series would reveal any additional insights beyond what has already been discovered. Bilirubin will rarely be pivotal in surgical decision making for appendicitis but it does offer some incremental value.

Conclusions

A higher level of bilirubin is independently associated with an increased likelihood of both uncomplicated and complicated appendicitis. Although statistically significant as a marker of the presence of the disease and the degree of inflammation, it is not clinically useful because the accuracy is not high, and it remains a non-specific test used alone or in combination with other inflammatory markers. Although high levels are associated with a more complicated form of the disease, as an investigation it is unlikely that the bilirubin level would be pivotal in the evaluation and management of most patients with acute appendicitis.

Abbreviations

AUC – Area under the curve (of the ROC)

CA – Complicated Appendicitis

CRP – C-reactive protein

CT – Computed Tomography

NA – Negative Appendicectomy

NPV – Negative Predictive Value

OR – Odds Ratio (from Logistic Regression)

PPV – Positive Predictive Value

ROC – Receiver Operator Curve

WCC – White Cell Count

Declarations

- **Ethics approval and consent to participate:** This research was approved by the local hospital ethics committee (Royal Brisbane and Women's Hospital Human Research Ethics Committee, Approval 52003) and was conducted according to the declaration of Helsinki. As a retrospective review, there are no participants and the requirement for consent was waived by the local ethics committee.
- **Consent for publication:** Not applicable (no individual or identifying patient data are presented).
- **Availability of data and materials:** If the manuscript is accepted, data can be made available on reasonable request.
- **Competing Interests:** There are no competing interests and none of the authors have any relevant financial interests to declare.
- **Funding:** There was no funding for this study.
- **Acknowledgements:** Thank you Dr Donald Cameron (Townsville Hospital) for providing data for analysis.

Authors' contributions:

Author AM conceived the idea, collected data, performed the statistical analysis, and wrote the manuscript. BS, EA, SK, TM, MB and TC collected data. Manuscript writing and checking was provided by HS and MS.

References

1. Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol.* 1990;132(5):910-25.
2. Korner H, Sondenaa K, Soreide JA, Andersen E, Nysted A, Lende TH, et al. Incidence of acute nonperforated and perforated appendicitis: age-specific and sex-specific analysis. *World J Surg.* 1997;21(3):313-7.
3. Arthur T, Gartrell R, Manoharan B, Parker D, Collaboration Q. Emergency appendicectomy in Australia: findings from a multicentre, prospective study. *ANZ J Surg.* 2017;87(9):656-60.
4. van Rossem CC, Bolmers MD, Schreinemacher MH, van Geloven AA, Bemelman WA, Group SACS, et al. Prospective nationwide outcome audit of surgery for suspected acute appendicitis. *British Journal of Surgery.* 2016;103(1):144-51.
5. Papandria D, Goldstein SD, Rhee D, Salazar JH, Arlikar J, Gorgy A, et al. Risk of perforation increases with delay in recognition and surgery for acute appendicitis. *journal of surgical research.* 2013;184(2):723-9.
6. Berger Y, Nevler A, Shwaartz C, Lahat E, Zmora O, Gutman M, et al. Elevations of serum CA-125 predict severity of acute appendicitis in males. *ANZ journal of surgery.* 2016;86(4):260-3.
7. Kafetzis D, Velissariou I, Nikolaidis P, Sklavos M, Maktabi M, Spyridis G, et al. Procalcitonin as a predictor of severe appendicitis in children. *European Journal of Clinical Microbiology and Infectious Diseases.* 2005;24(7):484-7.
8. Kharbanda AB, Rai AJ, Cosme Y, Liu K, Dayan PS. Novel serum and urine markers for pediatric appendicitis. *Academic Emergency Medicine.* 2012;19(1):56-62.
9. Jenniskens M, Langouche L, Vanwijngaerden Y-M, Mesotten D, Van den Berghe G. Cholestatic liver (dys) function during sepsis and other critical illnesses. *Intensive care medicine.* 2016;42(1):16-27.
10. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama.* 2016;315(8):801-10.
11. Sandstrom A, Grieve DA. Hyperbilirubinaemia: its utility in non-perforated appendicitis. *ANZ journal of surgery.* 2017;87(7-8):587-90.
12. Acharya A, Markar SR, Ni M, Hanna GB. Biomarkers of acute appendicitis: systematic review and cost-benefit trade-off analysis. *Surg Endosc.* 2017;31(3):1022-31.

13. Hsieh F. Sample size tables for logistic regression. *Statistics in medicine*. 1989;8(7):795-802.
14. Panagiotopoulou I, Parashar D, Lin R, Antonowicz S, Wells A, Bajwa F, et al. The diagnostic value of white cell count, C-reactive protein and bilirubin in acute appendicitis and its complications. *The Annals of the Royal College of Surgeons of England*. 2013;95(3):215-21.
15. McGowan DR, Sims HM, Zia K, Uheba M, Shaikh IA. The value of biochemical markers in predicting a perforation in acute appendicitis. *ANZ journal of surgery*. 2013;83(1-2):79-83.
16. Farooqui W, Pommergaard H, Burcharth J, Eriksen J. The diagnostic value of a panel of serological markers in acute appendicitis. *Scandinavian Journal of Surgery*. 2015;104(2):72-8.
17. Hong YR, Chung C-W, Kim JW, Kwon CI, Ahn DH, Kwon SW, et al. Hyperbilirubinemia is a significant indicator for the severity of acute appendicitis. *Journal of the Korean Society of Coloproctology*. 2012;28(5):247.
18. Käser SA, Fankhauser G, Willi N, Maurer CA. C-reactive protein is superior to bilirubin for anticipation of perforation in acute appendicitis. *Scandinavian journal of gastroenterology*. 2010;45(7-8):885-92.
19. Parekh D, Jain D, Mohite S, Phalgune D. Comparison of Outer Diameter of Appendix, C-reactive Protein, and Serum Bilirubin Levels in Complicated Versus Uncomplicated Appendicitis. *Indian Journal of Surgery*. 2019:1-5.
20. Sushruth S, Vijayakumar C, Srinivasan K, Kumar NR, Balasubramaniyan G, Verma SK, et al. Role of C-reactive protein, white blood cell counts, bilirubin levels, and imaging in the diagnosis of acute appendicitis as a cause of right iliac fossa pain. *Cureus*. 2018;10(1).
21. Mallows JL. The effect of a gold coin fine on C-reactive protein test ordering in a tertiary referral emergency department. *Medical Journal of Australia*. 2013;199(11):813-4.
22. Güller U, Rosella L, McCall J, Brügger L, Candinas D. Negative appendectomy and perforation rates in patients undergoing laparoscopic surgery for suspected appendicitis. *British journal of surgery*. 2011;98(4):589-95.
23. Mock K, Lu Y, Friedlander S, Kim DY, Lee SL. Misdiagnosing adult appendicitis: clinical, cost, and socioeconomic implications of negative appendectomy. *The American Journal of Surgery*. 2016;212(6):1076-82.
24. Nepogodiev D, Matthews JH, Morley GL, Bhanderi S, Jain S, Mohamed W, et al. Right Iliac Fossa Pain Treatment (RIFT) Study: protocol for an international, multicentre, prospective observational study. 2018.
25. Davidson GH, Flum DR, Talan DA, Kessler LG, Lavalley DC, Bizzell BJ, et al. Comparison of Outcomes of antibiotic Drugs and Appendectomy (CODA) trial: a protocol for the pragmatic randomised study of appendicitis treatment. *BMJ open*. 2017;7(11):e016117.
26. Lu Y, Friedlander S, Lee SL. Negative appendectomy: clinical and economic implications. *The American Surgeon*. 2016;82(10):1018-22.
27. Mällinen J, Rautio T, Grönroos J, Rantanen T, Nordström P, Savolainen H, et al. Risk of appendiceal neoplasm in periappendicular abscess in patients treated with interval appendectomy vs follow-up

with magnetic resonance imaging: 1-year outcomes of the Peri-Appendicitis Acuta randomized clinical trial. JAMA surgery. 2019;154(3):200-7.

Figures

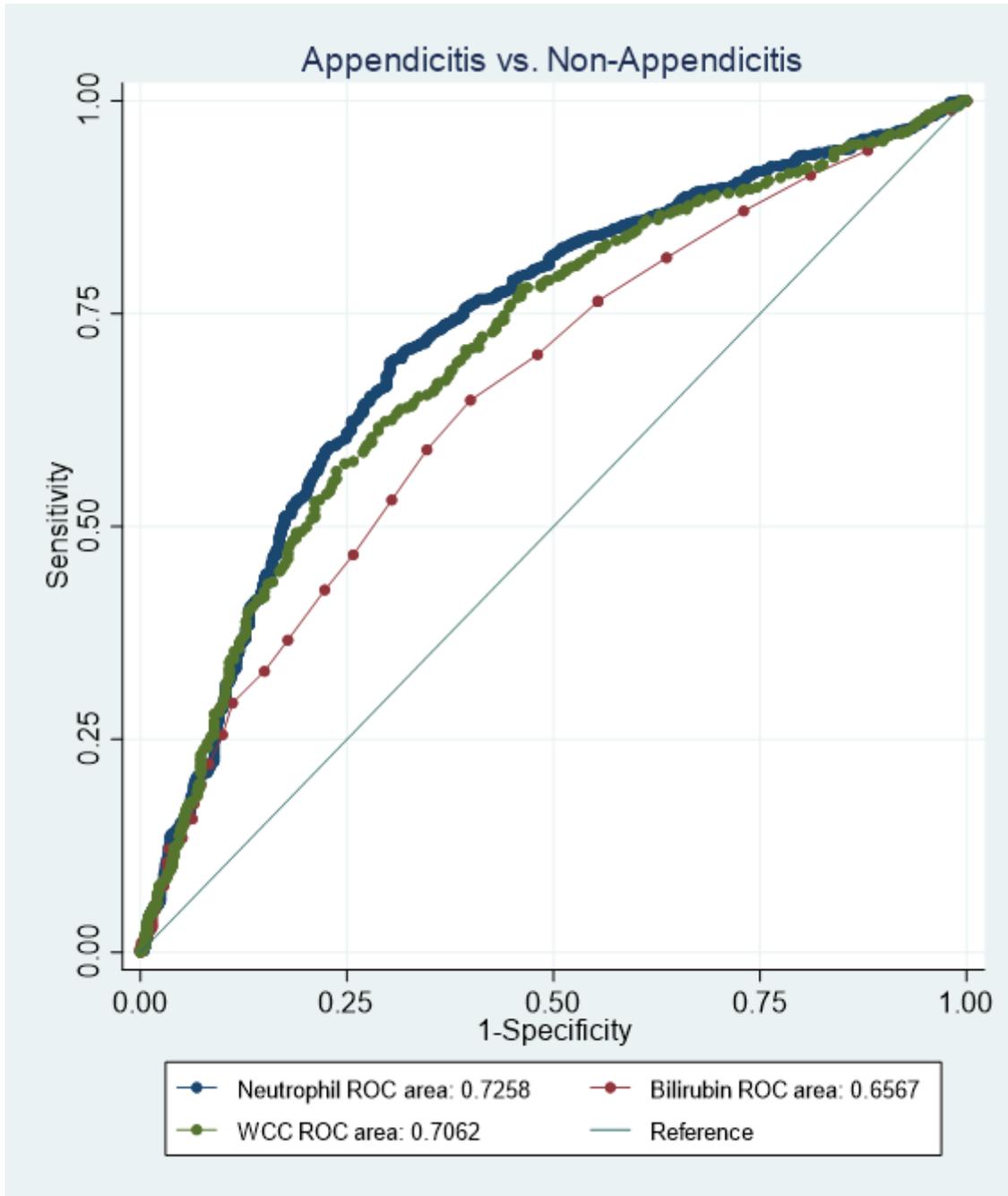


Figure 1

Receiver-Operator curves for Bilirubin, White cell count and Neutrophil count for the diagnosis of appendicitis versus non-appendicitis. Neutrophil count had a significantly higher area under the curve compared to white-cells and bilirubin. This was statistically significant ($p < 0.0001$), indicating neutrophils has the highest sensitivity/specificity for the detection of appendicitis.

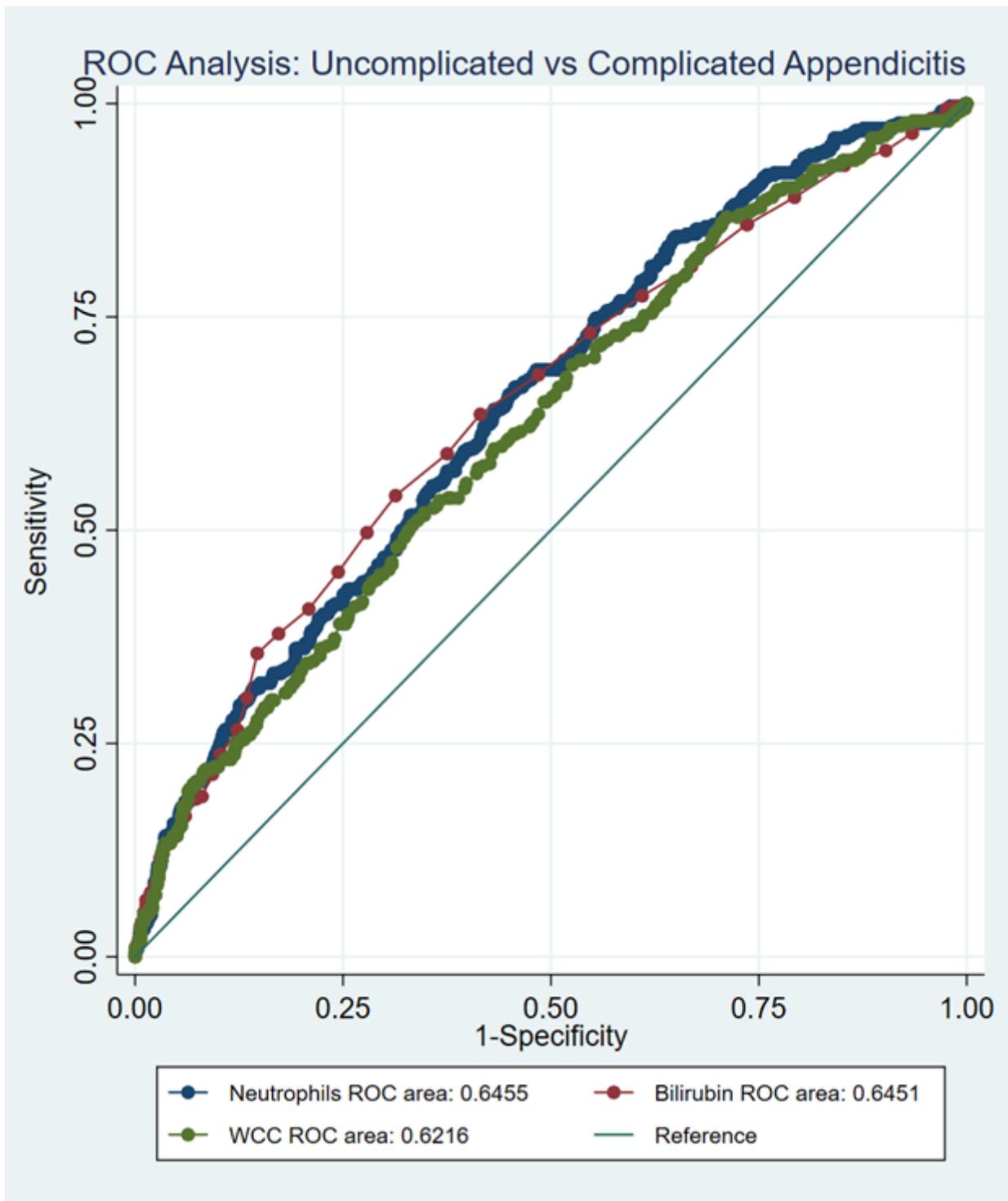


Figure 2

Receiver-Operator curves for bilirubin, white cell count (WCC) and neutrophil count for the diagnosis of uncomplicated versus complicated appendicitis. Bilirubin and neutrophils had a significantly higher area under the curve compared to WCC (both $p < 0.001$). The areas did not statistically significantly differ between neutrophils and bilirubin ($p = 0.968$). Thus bilirubin is no better than neutrophils in the detection of complicated appendicitis.