Do the 'Eyes' have it? How good are surgeons at identifying appendicitis macroscopically? A retrospective study in New Zealand

Rennie X Qin, Damien L Ah Yen

ABSTRACT

AIMS: The management of a macroscopically normal appendix during diagnostic laparoscopy depends on the accuracy of surgeons' intra-operative assessments. This study aims to determine the accuracy of this assessment and identify factors affecting it.

METHODS: We reviewed appendicectomies on adult patients at Waikato District Health Board in 2017. The primary outcome was the agreement between the operative assessment and the gold standard histopathologic assessment. Secondary outcomes were predictors of this agreement.

RESULTS: 420 patients were included. Among 74 appendixes assessed as normal by surgeons, 16 (21.6%) had appendicitis on histology. Surgeons assessed 346 appendixes as inflamed; however, 22 (6.3%) were revealed to be histologically normal. Only 2 of the 14 appendiceal neoplasms on histology were identified at the time of laparoscopy. Overall, there was disagreement in 9.1% of cases. This yielded a kappa of 0.69, indicating moderate inter-rater reliability. An inflamed appendix was significantly more likely to be falsely assessed as normal by non-trainee registrars, in female patients and in patients with a pre-operative ultrasound. A pre-operative computerised tomography scan (CT) decreased the odds of false negative operative diagnoses, but it increased the odds of false positives.

CONCLUSIONS: Macroscopic assessment of the appendix lacks accuracy and may be challenging in certain groups of operators and patients.

Appendicectomy is the most-commonly performed emergency abdominal surgery worldwide. In New Zealand, more than 5,000 hospital admissions for appendicitis occur annually.[1] The widespread adoption of a laparoscopic approach to appendicectomy in recent decades has allowed for the option of leaving a macroscopically normal appendix in situ. However, whether to do so is an ongoing debate , reflected by recent international surveys of experts.[2,3] Those who argue for leaving a macroscopically normal appendix in situ quote the high morbidity of negative appendicectomies compared to appendicectomies for uncomplicated appendicitis found by some studies.[4,5] Those who argue for removing it fear recurrent presentations and complications arising from missed appendicitis.[6] A decision, either way, depends on the ability of surgeons to make an accurate intra-operative diagnosis. Previous studies have demonstrated that operative assessment may lack accuracy. Between 27% to 31% of macroscopically normal appendixes encountered at the time of surgery were found to be in-

flamed on histologic examination.[6–9] Studies have found the accuracy of intra-operative assessment to be unaffected by seniority and to be more challenging in female patients.[7,10] With the increasing prevalence of and familiarity with a laparoscopic approach, the accuracy of macroscopic assessment may be different in recent times compared to earlier cohorts. Therefore, this study aims to determine the accuracy of intra-operative assessment of appendicitis compared to the gold standard of histopathological examination and to identify factors affecting this accuracy.

Methods

Patient eligibility

This study retrospectively reviewed all adults aged 15 years and over who underwent an appendicectomy for suspected appendicitis at Waikato District Health Board between 1 January 2017 and 31 December 2017. We excluded appendicectomies that were planned or a part of another procedure, such as a trauma laparotomy or a Hartmann's procedure. Approval to carry out the review was granted by the Clinical Audit and Support Unit.

Data collection

Data were extracted from the electronic medical records, operation notes and histology reports. Data on demographic characteristics, pre-operative investigations and operative details including the operator, the approach and the findings, as well as the histopathological diagnosis, were collected using a standardised proforma. Data were collected in a secured database and anonymised prior to analysis.

Outcomes

The primary outcome was the agreement between surgeons' macroscopic assessment obtained from the operation note with the histopathologic diagnosis. Operative and histological findings were both graded as normal and inflamed. A histological diagnosis of appendicitis was made at our institution on the basis of neutrophilic infiltrates in the wall of the appendix in the correct clinical context.[11] Other operative findings were also recorded, such as suspicion for neoplasia. A further category of macroscopically uncomplicated appendicitis was recorded in cases with the American Association for the Surgery of Trauma (AAST) grade 1 appendicitis, which includes appendixes that appeared acutely inflamed but not gangrenous or perforated.[11] False negatives occurred when a macroscopically normal appendix turned out to have appendicitis on histology, and vice versa for false positives.

Secondary outcomes were predictors of false positive and false negative operative assessment. This comprised of a set of pre-determined factors elicited from the literature, including training level, age, gender and pre-operative imaging. The training level of the operating surgeon was graded as (1) junior registrar, who has not yet entered the Surgical Education and Training Program (SET) of the Royal Australasian College of Surgeons, (2) senior registrar, who is training under the SET program and (3) consultant, who has completed SET training. When more than one operator was present, the training level of the more senior operator was counted.

Statistical analysis

Statistical analyses were performed using R Studio version 1.3.1056-1. Inter-observer reliability between the surgeon and the pathologist's assessment was tested using Cohen's kappa. A value of less than 0.39 indicated minimal or no agreement; 0.40–0.58 indicated weak agreement; 0.60–0.79 indicated moderate agreement; and above 0.80 indicated strong or near-perfect agreement.[12] Univariate logistic regression was used to assess for factors that predicted false positive and false negative operative assessment. A p-value of less than 0.05 was deemed to be significant in this study.

Results

Demographics

A total of 573 patients underwent appendicectomy at Waikato Hospital in 2017. The inclusion criteria were met by 425 patients, and five cases were removed due to incomplete documentation. In total, 420 cases were included in the final analysis, of which 54.5% were female and the median age was 28 (Table 1). The majority (95.0%, 399/420 cases) of appendicectomies were laparoscopic; the other 21 cases (5.0%) were open or converted to open.

Table 1: Demographic and clinical characteristics of patients undergoing appendicectomy.

	n = 420 [n (%)]
Age (years)	
15–34	260 (61.9%)
35–49	83 (19.8%)
≥50	77 (18.3%)
Gender	
Female	229 (54.5%)
Male	191 (45.5%)
Ethnicity	
European	282 (67.1%)
Māori and Pacific	97 (23.1%)
Other	28 (6.7%)
Missing	13 (3.1%)
Pre-operative imaging	
None	238 (56.7%)
USS	80 (19.0%)
CT	98 (23.3%)
Both	4 (1.0%)
Approach	
Laparoscopic	399 (95.0%)
Open and conversion to open	21 (5.0%)
Surgeon seniority	
Junior registrar	28 (6.7%)
Senior registrar	377 (89.8%)
Consultant	15 (3.6%)

Correlation between operative and histological diagnoses

Among the 74 appendixes that appeared macroscopically normal to surgeons, 16 (21.6%) were revealed to have appendicitis on histology. Twenty cases (27.0%) did not have appendicitis but had other histological findings. This included faecaliths in 10, fibrous obliteration in 5, lymphoid hyperplasia in 4 and polyp in 1. Of the 346 appendixes that appeared to be inflamed to surgeons, 22 (6.3%) turned out to be histologically normal (Table 2). An operative AAST grade 1 appendicitis was presented by 19 of the 22 cases of false positive operative assessment.

Overall, the negative appendicectomy rate was 19.0%. There was an agreement between operative and histological assessment in 90.9% of cases. This yielded a Cohen's kappa of 0.69, which indicates a moderate correlation.

Appendixes assessed to have uncomplicated or AAST grade 1 appendicitis operatively had a higher rate of false positive macroscopic assessment (8.6%) and a lower rate of agreement between operative and histological diagnoses (90.5%) compared to the overall group. However, these differences were not statistically significant.

Where neoplasm was found on histology, only 2 out of 14 cases (14.3%) were detected intra-operatively. The 14 cases included 6 adenomatous and sessile serrated polyps, 4 neuroendocrine tumours, 3 mucinous neoplasms and 1 appendiceal adenocarcinoma.

The effect of seniority on operative assessment

Table 3 describes the effect of seniority on the operative assessment of the appendix.

Junior registrars' operative assessment of the appendix had the lowest correlation with the histological assessment. They had higher rates of false positive and false negative operative diagnoses compared to the other groups. In particular, they were significantly more likely to falsely assess a histologically inflamed appendix as normal compared to senior registrars (Table 3). Consultants, however, did not significantly differ in the accuracy of their operative assessment compared to the other groups.

Predictors of false positive and false negative operative assessment

Table 4 describes the effects of other demographic and clinical characteristics on surgeons' macroscopic assessment. False negative operative diagnoses were significantly more likely to occur in female patients and patients who had undergone a pre-operative ultrasound (USS). A pre-operative computerised tomography scan

decreased the odds of false negative operative diagnoses, but it increased the odds of false positive operative diagnoses. False positive operative assessments were also significantly more likely to occur in Māori and Pacific patients.

Subgroup analysis

All of the 10 patients with a false negative operative assessment who had undergone a pre-operative ultrasound were female. The result of the ultrasound was inconclusive in eight patients, positive for appendicitis in one and negative for appendicitis in one.

Discussion

This study examined the agreement between operative assessment of the appendix and the gold standard of histopathologic assessment and identified factors affecting this agreement. When the appendix was assessed as normal by surgeons intra-operatively, 21.6% turned out to be inflamed on histology. When the appendix appeared inflamed to surgeons, 6.3% were revealed to be histologically normal; the majority of false positive operative diagnoses were in cases with operative AAST grade 1 appendicitis. Overall, operative and histological diagnoses disagreed in 9.1% of cases. Cohen's kappa was 0.69, which indicates a moderate inter-observer correlation.

Previous studies reported similar results. They demonstrated only moderate correlation between macroscopic and histological assessment of the appendix, with an agreement between the two in 83.5% to 87.5% of cases.[7,9,10] Previous studies found the negative predictive value of operative

Table 2: Surgeons' macroscopic assessment of the appendix compared to pathologists' histological assessment.

Operative assessment	Histology assessment		Total
	Normal	Inflamed	
Normal	58 (78.4%)	16 (21.6%)	74
Inflamed	22 (6.4%)	324 (93.6%)	346
Total	80	340	420

Table 3 : The effect of seniority on the macroscopic assessment of the appendix.

Operative assessment by seniority	Histological assessment		FP rate	p-value[a]	FN rate	p-value[a]	Cohen's kappa
	Normal	Inflamed					
Junior registrar							
Normal	5	3	28.6%		14.3%		0.55
Inflamed	2	18					
Senior registrar							
Normal	51	11	28.2%	0.96	3.5%	0.042*	0.71
Inflamed	20	293					
Consultant							
Normal	1	1	0.0%	0.99	7.1%	0.52	0.63
Inflamed	0	13					

[a] p-value when tested against the junior registrar group.
Abbreviations: FP, false positive; FN, false negative.
*p-value < 0.05

assessment to be low, and histological signs of appendicitis were present in between 27% and 31% of macroscopically normal appendixes.[6–9] A study of macroscopic diagnosis based on video recordings found not only weak inter-observer correlation but also weak intra-observer correlation when the same operator viewed different videos.[13] Previous studies also confirmed that operative diagnosis of appendicitis was less accurate in AAST grade 1 or uncomplicated appendicitis compared to complicated appendicitis with higher AAST grades.[8]

Our study found that there is potential for pathology to be missed on macroscopic assessment. Only 2 out of 14 appendixes with neoplastic changes on histology were detected during the operation. Previous studies echoed this. Roberts et al. reported only 3 out of 16 neoplastic lesions were being detected intra-operatively, and Sadot et al. reported 1 out of 22.[8,14] Twenty cases (27.0%) of macroscopically normal appendixes had other histological findings such as faecaliths or lymphoid hyperplasia. Previous studies reported similar results.[15,16] The exact implications of these findings are unclear; however, studies have found faecalith and lymphoid hyperplasia to be associated with right iliac fossa pain even in the absence of histological appendicitis.

Our study found the accuracy of operative assessment of the appendix to be the lowest in junior registrars. Junior registrars made significantly more false negative macroscopic assessment than senior registrars. Similarly, Al-Ghnaniem et al. also found that inexperienced surgeons were more likely than experienced surgeons to make an incorrect diagnosis during laparoscopic appendicectomy.[19] However, recent studies found that seniority did not affect the accuracy of operative assessment.[7,10,20] These differences in findings could be due to differences in the documentation of the presence of a more senior operator and different grading of seniority. For example, Pham et al. grouped early SET trainees and non-trainees, whereas this study compared non-trainees with all SET trainees.[10] All studies have not found the accuracy of consultants' operative assessment to be significantly different from the other groups. This is likely due to the consistent finding of a small number of consultants performing laparoscopic appendicectomies.

We found significantly higher rates of false negative operative assessment in female patients. Other studies have also found female patients to be at particular risk of inaccurate operative assessment with higher rates of both false positive and false negative

Table 4: Demographic and clinical predictors of false positive and negative operative assessment.

		False positive				False negative			
		Odds ratio	**95% CI**		**p-value**	**Odds ratio**	**95% CI**		**p-value**
Age	<35								
	>35	1.80	0.66	4.72	0.32	0.32	0.10	0.86	0.08
Gender	Male								
	Female	0.85	0.36	2.06	0.76	7.00	2.31	31.40	0.011*
Ethnicity	European and other								
	Māori and Pacific	2.96	1.20	7.34	0.040*	0.22	0.02	0.88	0.143
Pre-operative imaging	None								
	USS	0.46	0.18	1.10	0.157	11.19	4.67	28.42	<0.001***
	CT	12.67	2.34	137.40	0.027*	0.16	0.02	0.64	0.050*

Abbreviations: CI, confidence interval; CT, computerised tomography; USS, ultrasound.
*p-value<0.05, ** p-value<0.01, ***p-value<0.001.

operative diagnoses of appendicitis.[10,20] The assessment of female patients with suspected appendicitis is generally challenging with female gender being associated with a high negative appendicectomy rate.[5,21]

With regards to imaging, we found the use of pre-operative ultrasound to increase the odds of a false negative operative assessment. This was also reported by Thong et al.[20] A recent meta-analysis found ultrasound to have a low specificity at 81%, which reflects its lack of ability at excluding appendicitis.[22] However, ultrasound could reduce the number of CTs performed, and it remains the recommended first-line investigation in certain patient groups due to its non-radiative nature.[23] Female patients tended to be more frequently selected for an ultrasound, which again reflects the challenges of the operative assessment of appendicitis in female patients. A non-contributory ultrasound with the appendix not being visualised could bias subsequent operative assessment.

On the other hand, this study found the use of CT imaging significantly reduced the odds of a false negative operative assessment but significantly increase the odds of a false positive macroscopic diagnosis. Indeed, studies have generally found CT to have good sensitivity and specificity in diagnosing appendicitis.[24] Strong et al. also found CT to be associated with more false positive operative diagnosis.[7] A possible explanation could be that patients with CT proven appendicitis may have recovered with antimicrobial therapy while they waited for an operation. However, the assessment of the duration between imaging and operative intervention was outside of the scope of this study.

More false positive operative assessment of appendicitis was also found in Māori and Pacific patients in our study. The reason for this is not entirely clear. The association between ethnicity and macroscopic assessment of the appendix has not been well studied previously and warrants further investigation.

This study is limited by its single-centre setting, which reduces its generalisability. Although histologic assessment of the appendix was used as the gold standard, potential inter-rater variability may exist between pathologists, and the diagnostic criteria used by different pathologists may differ. The retrospective nature of the study limits it in several ways. Firstly, data on operative assessment and operators present rely on accurate documentation in the operation note. Secondly, junior operators may have been selected for appendicectomies that were perceived to be easier to perform, predisposing them to more challenging macroscopic diagnosis in patients with mild and uncomplicated appendicitis. Thirdly, due to a lack of blinding, surgeons' macroscopic assessment may not have been entirely based on objective operative findings and may have been influenced by the patients' clinical presentations or pre-operative imaging results. Lastly, in the analysis of the accuracy of operative assessment, this study did not include a number of macroscopically normal appendixes that were left in situ by surgeons and therefore did not receive a histological assessment. Additionally, we cannot comment on the long-term implications of leaving behind appendixes that appear normal to surgeons. There are suggestions that mild appendicitis or endoluminal appendicitis without obvious macroscopic abnormality may resolve spontaneously without intervention.[25] However, studies following-up patients with macroscopically normal appendixes left in situ have only included a relatively small number of patients thus far.[26,27] A recent large retrospective cohort study supported the possibility that endoluminal appendicitis may be a pathological entity that tends to occur in young females with non-diagnostic imaging who do benefit symptomatically from appendicectomies.[28]

This study found that the operative assessment could lack accuracy and may miss pathology. The accuracy of macroscopic assessment was lower in the most junior operators. The implementation of standardised and objective grading criteria for the macroscopic assessment of the appendix during laparoscopy has been found to improve the accuracy of the operative diagnosis of appendicitis.[13,29] This may be of particular value in the training of junior operators. Female patients and patients who have undergone a pre-operative ultrasound are prone to false negative operative

assessment. The use of a pre-operative CT could reduce the odds of a false negative assessment in these patients; however, it needs to be balanced with the potential harm of ionising radiation. Low-dose CT has been found to be non-inferior in diagnostic utility compared to standard-dose CT and could be considered for this patient group.[30] The implications of leaving behind a macroscopically normal appendix warrant further exploration through large, randomised and blinded trials.

This study found that macroscopic diagnosis of appendicitis could be challenging with 21.6% of macroscopically normal appendixes having histological signs of inflammation.

Competing interest:
Nil.
Acknowledgements:
The authors acknowledge the Clinical Audit and Support Unit at Waikato District Health Board.
Author information:
Rennie X Qin MBChB, BMedSc (Hons): Registrar, Department of General Surgery, Waikato District Health Board, Hamilton, New Zealand.
Damien L Ah Yen MBChB, FRACS: General and Trauma Surgeon, Department of General Surgery, Waikato District Health Board, Hamilton, New Zealand.
Corresponding author:
Mr Damien Ah Yen, Department of General Surgery, Waikato District Health Board, Pembroke Street, Hamilton 3240, New Zealand, +64211004937
DamienAhYen@waikatodhb.health.nz
URL:
www.nzma.org.nz/journal-articles/do-the-eyes-have-it-how-good-are-surgeons-at-identifying-appendicitis-macroscopically-a-retrospective-study-in-new-zealand

REFERENCES

1. Hider P, Wilson L, Rose J, Weiser TG, et al. The role of facility-based surgical services in addressing the national burden of disease in New Zealand: An index of surgical incidence based on country-specific disease prevalence. Surgery. 2015;158(1):44–54.
2. Jaunoo SS, Hale AL, Masters JPM, Jaunoo SR. An international survey of opinion regarding investigation of possible appendicitis and laparoscopic management of a macroscopically normal appendix. Ann R Coll Surg Engl. 2012;94(7):476–80.
3. Turner EJH, Lightwood R. Management of the Normal Appendix during Laparoscopy for Right Iliac Fossa Pain. Mishra R, editor. World J Laparoscopic Surg. 2009;2(2):15–7.
4. Dubrovsky G, Rouch J, Huynh N, et al. Clinical and socioeconomic factors associated with negative pediatric appendicitis. J Surg Res. 2017;218:322–8.
5. Mock K, Lu Y, Friedlander S, Kim DY, Lee SL. Misdiagnosing adult appendicitis: clinical, cost, and socioeconomic implications of negative appendectomy. Am J Surg. 2016;212(6):1076–82.
6. Bhangu A, Begaj I, Ray D. Population level analysis of diagnostic laparoscopy versus normal appendicectomy for acute lower abdominal pain. International Journal of Surgery. 2014;12(12):1374–9.
7. Strong S, Blencowe N, Bhangu A. How good are surgeons at identifying appendicitis? Results from a multi-centre cohort study. International Journal of Surgery. 2015;15:107–12.
8. Roberts JK, Behravesh M, Dmitrewski J. Macroscopic findings at appendicectomy are unreliable: implications for laparoscopy and malignant conditions of the appendix. Int J Surg Pathol. 2008;16(4):386–90.
9. Phillips AW, Jones AE, Sargen K. Should the macroscopically normal appendix be removed during laparoscopy for acute right iliac fossa pain when no other explanatory pathology is found? Surg Laparosc Endosc Percutan Tech. 2009;19(5):392–4.
10. Pham H, Devadas M, Howle J. Effect of surgical experience on the macroscopic diagnosis of appendicitis: A retrospective cohort study. International Journal of Surgery. 2015;16:78–82.

11. Hernandez MC, Kong VY, Aho JM, et al. Increased anatomic severity in appendicitis is associated with outcomes in a South African population. J Trauma Acute Care Surg. 2017;83(1):175–81.
12. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.
13. van den Boom AL, de Wijkerslooth EML, Mauff K a. L, et al. Interobserver variability in the classification of appendicitis during laparoscopy. Br J Surg. 2018;105(8):1014–9.
14. Sadot E, Keidar A, Shapiro R, Wasserberg N. Laparoscopic accuracy in prediction of appendiceal pathology: oncologic and inflammatory aspects. The American Journal of Surgery. 2013;206(5):805–9.
15. Singhal V, Jadhav V. Acute Appendicitis: Are we Over Diagnosing it? Ann R Coll Surg Engl. 2007 Nov;89(8):766–9.
16. Chiarugi M, Buccianti P, Decanini L, et al. 'What you see is not what you get'. A plea to remove a "normal" appendix during diagnostic laparoscopy. Acta Chir Belg. 2001;101(5):243–5.
17. Grimes C, Chin D, Bailey C, et al. Appendiceal faecaliths are associated with right iliac fossa pain. Ann R Coll Surg Engl. 2010;92(1):61–4.
18. Giuliano V, Giuliano C, Pinto F, Scaglione M. Chronic appendicitis 'syndrome' manifested by an appendicolith and thickened appendix presenting as chronic right lower abdominal pain in adults. Emerg Radiol. 2006;12(3):96–8.
19. Al-Ghnaniem R, Kocher HM, Patel AG. Prediction of inflammation of the appendix at open and laparoscopic appendicectomy: findings and consequences. European Journal of Surgery. 2002;168(1):4–7.
20. Thong DW, Crouch S, Morgan S, et al. Can Surgeons Identify Appendicitis Macroscopically? Results From a Multicentre Prospective Study. Surg Laparosc Endosc Percutan Tech. 2019;29(5):344–8.
21. Lee M, Paavana T, Mazari F, Wilson TR. The morbidity of negative appendicectomy. Ann R Coll Surg Engl. 2014;96(7):517–20.
22. Giljaca V, Nadarevic T, Poropat G, et al. Diagnostic Accuracy of Abdominal Ultrasound for Diagnosis of Acute Appendicitis: Systematic Review and Meta-analysis. World J Surg. 2017;41(3):693–700.
23. Schuh S, Chan K, Langer JC, et al. Properties of serial ultrasound clinical diagnostic pathway in suspected appendicitis and related computed tomography use. Acad Emerg Med. 2015;22(4):406–14.
24. Di Saverio S, Birindelli A, Kelly MD, et al. WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. World Journal of Emergency Surgery. 2016;11(1):34.
25. Pieper R, Kager L, Näsman P. Clinical significance of mucosal inflammation of the vermiform appendix. Ann Surg. 1983;197(3):368–74.
26. van den Broek WT, Bijnen AB, de Ruiter P, Gouma DJ. A normal appendix found during diagnostic laparoscopy should not be removed: Diagnostic laparoscopy for suspected appendicitis. British Journal of Surgery. 2001;88(2):251–4.
27. Teh SH, O'Ceallaigh S, Mckeon JG, et al. Should an appendix that looks 'normal' be removed at diagnostic laparoscopy for acute right iliac fossa pain? Eur J Surg. 2000;166(5):388–9.
28. Mizumoto R, Cristaudo AT, Lai NK, et al. Dilemma of mucosal appendicitis: a clinico-pathological entity? A retrospective cohort study. ANZ J Surg. 2018 Apr;88(4):E284–8.
29. Gomes CA, Sartelli M, Di Saverio S, et al. Acute appendicitis: proposal of a new comprehensive grading system based on clinical, imaging and laparoscopic findings. World J Emerg Surg. 2015;10:60.
30. Sippola S, Virtanen J, Tammilehto V, et al. The Accuracy of Low-dose Computed Tomography Protocol in Patients With Suspected Acute Appendicitis: The OPTICAP Study. Ann Surg. 2020;271(2):332–8.