April 13, 2019

Dr. **Dennis A. Bloomfield**
Dr. Sandro Vento
Editors-in-Chief
World Journal of Clinical Cases

Dear Drs. Bloomfield and Vento:

Thank you for considering publishing our manuscript titled "**Clinical differentiation of acute appendicitis and right colonic diverticulitis: a case-control study**" in your esteemed journal, *World Journal of Clinical Cases*, subject to revision. We wish to re-submit the revised version of our manuscript for your review. The manuscript ID is 47070.

We appreciate the reviewers' invaluable comments, which have helped us improve our manuscript. We have provided point-by-point responses to each of the reviewer's comments and describe the corresponding revisions. **As per your instructions, changes have been indicated by underlined font in the revised manuscript.**

We look forward to any further comments regarding our revised manuscript.

Sincerely yours,
Yosuke Sasaki, MD, PhD
Department of General Medicine and Emergency Care
Toho University School of Medicine, Omori Hospital
6-11-1 Omori-Nishi, Ota-ku
Tokyo 143-8541, Japan
Phone: +81-3-3762-4151
Fax: +81-3-3765-6518
Email: pqrstbb@yahoo.co.jp, yousuke.sasaki@med.toho-u.ac.jp

Responses to the Reviewers' Comments

Responses to reviewer 00069988's comments

Abstract-Result

Comment 1:

"Median age was 35.5"

Does that mean that the children are excluded? Please define the study population by age.

Response:

Thank you for your comment. As you pointed out, pediatric patients were excluded in the present study. We did define and state the age limitation in the "Methods" of the Abstract as follows (page 4, lines 16-17):

> **"We performed an exploratory single-center retrospective case-control study that evaluated 369 Japanese patients (age ≥16 years),…"**

Comment 2:

"Prevalence of epigastric/periumbilical pain, nausea/vomiting, anorexia, and history of unresected appendicitis"

Please define this better. What is the point? Of course, that patients with acute appendicitis have a significantly lower rate of appendectomy

Response:

We apologize for the lack of clarity. Because a history of appendicitis that is treated without appendectomy is a well-known risk factor of recurrent appendicitis, we intended to evaluate whether a history of appendicitis that was treated without appendectomy (i.e., unresected appendicitis) was a significant risk factor of recurrent appendicitis in the present study.

Another reviewer fully understood our intention and requested the following:

> ***"'Patients with previous history of appendectomy were excluded because appendicitis was quite unlikely for those patients.' Please remove 'because appendicitis was quite unlikely for those patients' as this is understood implicitly and it is a fair inclusion criteria."***

However, as per your suggestion, we have added the following sentence in the "Methods" section of the main manuscript to facilitate readers' understanding (page 9, lines 3-4):

> **"We reviewed whether the patients had a history of acute appendicitis that was treated without appendectomy because previous appendicitis is a well-known risk factor of recurrent appendicitis if appendectomy was not performed[13]."**

In this sentence, we newly cited reference number 13: Salminen P, Tuominen R, Paajanen H, Rautio T, Nordström P, Aarnio M, Rantanen T, Hurme S, Mecklin J-P, Sand J, Virtanen J, Jartti A, Grönroos JM. Five-year follow-up of antibiotic therapy for uncomplicated acute appendicitis in the APPAC randomized clinical trial. *JAMA* 2018; 320: 1259-1265 [PMID: 30264120 DOI: 10.1001/jama.2018.13201]

The low prevalence of appendectomy in patients with previous appendicitis may be surprising. However, conservative treatment for appendicitis with antibiotics is very popular in Japan, and it is the *de facto* standard in our institute. Therefore, as mentioned in the manuscript, only 6/41 patients with a history of appendicitis were excluded from the study because they underwent appendectomy. To clearly explain this, we have added the following sentence in the "Results" (page 11, lines 19):

> **"In 41 patients with a history of appendicitis, only 6 patients with a history of appendectomy were excluded before the study because conservative treatment for AA was the *de facto* standard treatment in our hospital; 32 and 3 patients with a history of appendicitis were eventually included in the AA group and ARCD group, respectively (p<0.001, Table 1)."**

Comment 3:

"…RLQ pain and history of diverticulitis…"

How was the history of diverticulitis diagnosed/confirmed?

Response:

Thank you for your comment. Because this study had a retrospective design and used medical records, we diagnosed previous diverticulitis based on patients' reports and medical records. We mentioned this in the "Materials and methods" as follows (page 8, lines 1):

> **"In this exploratory single-center retrospective case-control study, we evaluated medical records from patients…"**

Unfortunately, we did not confirm the presence of diverticula by colonoscopy or other methods. We have mentioned the limitation of the study due to the retrospective design in the "Discussion" as follows (page 18, lines 14):

> **"Our study has some limitations. First, because the present study was a retrospective case-control study that used medical records, we could not collect some previously reported important information…"**

However, we did not mention the lack of confirmation of diverticula by colonoscopy or barium enema examination. Thus, we have added a new paragraph concerning this limitation in the "Discussion" as follows (page 19, lines 3-4):

> **"As mentioned earlier, we diagnosed AA and ARCD based on the findings of CT because it has been the most common diagnostic tool of AA and ARCD in Japanese clinical practice thanks to the fact that CT is the most available modality[6]. We believe that we could appropriately diagnose AA and ARCD by CT because its respective sensitivities and specificities are reportedly 90-100% and 91-99% for diagnosing AA[20], and 94% and 99% for diagnosing colonic diverticulitis[30]. Of note, we lacked the confirmation of diverticula by colonoscopy or barium enema examination in the present retrospective study."**

In this revision, we newly cited reference number 30: Laméris W, van Randen A, Bipat S, Bossuyt PMM, Boermeester MA, Stoker J. Graded compression ultrasonography and computed tomography in acute colonic diverticulitis: Meta-analysis of test accuracy. *Eur Radiol* 2008; 18:2498-2511 [DOI: 10.1007/s00330-008-1018-6]

Comment 4:

"Median CRP levels in AA and ARCD groups were 1.1 and 4.9 mg/dL"

Ranges are very important. Interestingly this group of AA has low mean CRP values. Can authors explain in the discussion section.

Response:

Thank you for the valuable comment. We agree that "ranges are very important." As per your suggestion, we have added the interquartile ranges as follows (page 5, lines 8-9):

> **"…median CRP levels were 1.1 (IQR: 0.2-4.1) and 4.9 (IQR: 2.9-8.5) mg/dL, respectively ($p<0.001$)."**

We had already discussed the lower CRP level in the AA group in the manuscript (page 17, lines 4-9).

Comment 5:

"ORs were significantly lower with a longer onset-to-visit interval"

How is longer interval defined? Is there a cut-off value?

Response:

Thank you for your comment. Although we already preliminary described the interval of both groups, we added the interquartile ranges in the manuscript to provide more precise information as follows according to your comment (page 5, lines 6-7):

"Median onset-to-visit intervals were 1 (Interquartile range [IQR]: 0-1) and 2 (IQR: 1-3) days in the AA and ARCD groups, respectively (p<0.001)."

We have addressed the definition of "prolonged interval" in a response to your comment number 7.

Comment 6:

"ORs were significantly lower with a longer onset-to-visit interval (OR: 0.84, CI: [0.72-0.97]), RLQ pain (OR: 0.28, CI: [0.11-0.71]), history of diverticulitis (OR: 0.034, CI: [0.005-0.20]), and CRP>3.0 mg/dL (OR: 0.25, CI: [0.14-0.43]), suggesting that ARCD was more likely"

Is this conclusion for all races or only Asians?

Response:

Thank you for this important comment. All participants were Japanese patients. According to your comment, we have changed "369 patients" to "369 Japanese patients" in the Abstract (page 4, line 17) and added "All patients were Japanese (immigrants or tourists were not included)." to the "Materials and methods" section (page 8, lines 6-7).

Introduction

Comment 7:

"Although prolonged…"

Define prolonged. There are definition for acute, subacute and chronic pain.

Response:

Thank you for your comment. We appreciate the point of your question. However, "prolonged" pain/symptoms were cited from a few previous studies with different results/definition (reference #9: Chen SC, Chang KJ, Wei TC, Yu SC, Wang SM. Can cecal diverticulitis be differentiated from acute appendicitis? *J Formos Med Assoc* 1994;**93**:263–265 [PMID: 7920071] **and reference #10: Shin JH**, Son BH, Kim H.

Clinically distinguishing between appendicitis and right-sided colonic diverticulitis at initial presentation. *Yonsei Med J* 2007; 48: 511-516 [PMID: 17594161 DOI: 10.3349/ymj.2007.48.3.511]**) (page 7, line 9). Thus, it is difficult to precisely define the cut-off interval of "prolonged" period.**

Comment 8:

"Although prolonged pain, initial right lower quadrant (RLQ) pain, lack of migration of pain, leukocytosis, nausea/vomiting, constipation, and systemic toxic signs have been proposed as differential findings of ARCD from AA"

I partly disagree. Appendicitis can start as RQL pain, leukocytosis is also present, nausea/vomiting are also frequently present. If authors want to have these as predictors of ARCD then comparison of incidence of these parameters should be made.

Response:

We apologize for the lack of clarity. However, we agree with your comment, and the sentence you pointed out corresponds with your opinion (page 7, lines 9-14). In the sentence, we initially introduced the results of previous studies (not our present study) as follows: "Although prolonged pain, initial right lower quadrant (RLQ) pain, lack of migration of pain, leukocytosis, nausea/vomiting, constipation, and systemic toxic signs have been proposed as differential findings of ARCD from AA." We subsequently stated that "these findings are based on several small studies using univariate comparisons from Korea and Taiwan; we could not find any published previous studies from Japan or confounder-adjusted studies." We just compared parts of our results to those of previous studies in order to evaluate the consistency with previous study results.

Material and Methods

Study variables

Comment 9:

Why did authors made CT to all patients? What about transabdominal ultrasound scan. According to Asian studies it is very accurate in diagnosing ARCD.

Response:

We agree with your comment on the accuracy of abdominal ultrasonography in diagnosing ARCD. However, we screened the patients diagnosed as having AA or ARCD by CT in this retrospective study because CT is the most common diagnostic tool in Japanese practice. We clearly mentioned this fact in the "Introduction" as follows (page 7, lines 4-7):

"Thanks to qualitive improvement and the high penetration rate of computed tomography (CT) scanning in Japan[6], differentiation of ARCD and AA mainly depends this modality."

We also stated that we enrolled the patients with AA and ARCD diagnosed by CT in the "Materials and methods" section as follows (page 8, lines 7-8):

"Diagnoses were confirmed by CT scans in all cases and for both groups."

To explain this issue, we have added the following sentences in the "Discussion" (page 19, lines 3-10).

"As mentioned earlier, we diagnosed AA and ARCD based on the findings of CT because it has been the most common diagnostic tool of AA and ARCD in Japanese clinical practice thanks to the fact that CT is the most available modality[6]. We believe that we could appropriately diagnose AA and ARCD by CT because its respective sensitivities and specificities are reportedly 90-100% and 91-99% for diagnosing AA[20], and 94% and 99% for diagnosing colonic diverticulitis[30]. Of note, we lacked the confirmation of diverticula by colonoscopy or barium enema examination in the present retrospective study."

We apologize that we could not provide the results of ultrasonography in this retrospective study because ultrasonography is not a standard diagnostic modality in our institute.

Categorization of contentious variables

Comment 10:

"Fever was defined as a body temperature of ≥38.0°C"

Fever cannot be defined as 38. There are definitions for a fever that are universal. Also, AA has elevated temperature in many patients that is below 38, and it is also fever).

Response:

We agree with your comment that fever is not always a body temperature ≥38.0°C. However, the definitions of fever in previous studies were actually inconsistent (fever was defined as >37.2℃, >37.3℃, and ≥38.0°C, or no clear definition was

provided). We defined fever as a body temperature >38℃ according to a previous study (reference #9).

In response to your comment, we have recalculated the prevalence of fever in our participants with other definitions of fever according to previous studies, including the Alvarado score. The result did not affect the significance of body temperature compared to the result with the definition of fever as a body temperature ≥38.0°C.

If fever was defined as a body temperature >37.3℃, 80 (33.9%) patients and 47 (35.3%) patients with AA and ARCD had fever, respectively. The difference in the prevalence was not significant (p=0.78).

If fever was defined as a body temperature >37.2℃, 103 (43.6%) patients and 56 (42.1%) patients with AA and ARCD had fever, respectively. The difference in the prevalence was not significant (p=0.77).

To explain these results, we have added a new sentence in the "Results" as follows (page 12, lines 18 to page 13, line 2):

> **"Prevalence of fever (body temperature ≥38.0°C) was not significantly different between the AA and ARCD groups (AA 14.4% vs. ARCD 15.0%, p=0.87). Because the definition of fever differs among previous studies[9,11,17], we also evaluated the prevalence of fever with the definition of fever as a body temperature >37.2℃[11] or >37.3℃[17]. However, all results showed that the prevalence of fever was not significantly different between AA and ARCD (p=0.77-0.78)."**

We have also added reference #9 (Chen SC, Chang KJ, Wei TC, Yu SC, Wang SM. Can cecal diverticulitis be differentiated from acute appendicitis? *J Formos Med Assoc* 1994; 93: 263-265 [PMID: 7920071]**) as a citation for the definition of fever in "*Categorization of contentious variables*" (page 9, line 13).**

Comment 11:

Where was the point of temperature measurement (orally, analy, axillary, with which instrument.

Response:

Thank you for your comment. Because temperature was measured axillary in all patients of our study with an electric thermometer manufactured by Terumo, we have added the following sentence in the "Study variables" subsection (page 9, line 2-3):

"Body temperature was measured at the axilla with an electric thermometer (Terumo, Tokyo, Japan)."

We have also added "axillary measured" in front of "body temperature" in the "Categorization of contentious valuables" subsection (page 9, line 12).

Comment 12:

"Leukocytosis was defined as a leukocyte count >11,000 /mm^3"

Leukocytosis is not defined as 11.

Response:

Thank you for your comment. We agree with your opinion that leukocytosis is not always defined as 11,000/mm^3 in all studies. We recognize that several previous studies defined leukocytosis as 10,000/mm^3. However, we needed a generally accepted cut-off value of leukocytosis, and we cited the definition according to reference #15: Riley LK, Rupert J. Evaluation of patients with leukocytosis. *Am Fam Physician* 2015; 92: 1004-1111. [PMID: 26760415 DOI: d12293 [pii]]. **Because this article is a sophisticated review published in a reliable, popular journal, we believe the cut-off value is appropriate for our inclusion criteria.**

Comment 13:

Authors should be aware that AA and ARCD cannot be defined by a single laboratory parameter or single clinical symptom and sign. This can be misleading mostly for younger physicians. ARCD and AA are clinical presentations with symptoms and signs.

Response:

Thank you for this important comment. We agree with your opinion. Please understand that our study is not focused on the statistical difference of single parameters; our study was a multivariate comparison that used a logistic regression model. In the "Discussion," we just compared single parameters between our study and previous studies in order to evaluate the consistency of each parameter between our study and others. These comparisons were not the main point of the present study. To clarify, we have added the following phrase at the end of paragraph on study limitations in the "Discussion" (page 20, lines 11-12):

"Because previous studies focused on differences between each single parameter[9–12],…"

Comment 14:

Therefore for good comparison the authors should calculate Alvarado Score for these 2 groups to see the difference. Also, maybe some parameters would be absent in ARCD which would be very important for the clinical suspicion for the physician that had the first contact with the patient.

Response:

We agree with your opinion and apologize that we could not evaluate the Alvarado score because of the retrospective study design, as we mentioned in the manuscript as follows (page 18, lines 14-19):

> **"Our study has some limitations. First, because the present study was a retrospective case-control study that used medical records, we could not collect some previously reported important information, such as parameters included in the Alvarado score[17], e.g., migration of abdominal pain or neutrophilia/left shift, despite previous studies clearly showing a high prevalence of pain migration and higher Alvarado score in AA compared to ARCD[9,11,12]."**

We performed logistic regression analysis as a substitute for the calculation of the Alvarado score. We believe that logistic regression is a common method of multivariable analysis and therefore acceptable.

Comment 15:

"We divided the patients into age groups, based on the median age of the patients, as follows: young, ≤40 years and old, >40 years; we did not use receiver operator characteristic (ROC) analysis of age for prediction of ARCD because it was poorly accurate (area under the curve (AUC) was 0.41)."

If ROC for age is poorly accurate, why did the authors make a division into patients young, ≤40 years and old, >40 years? Is this done based on previous studies that found this age as some cut-off?

Response:

We apologize for the lack of clarity. However, we mentioned how we divided the age in the "Materials and methods" as follows in the manuscript before revision:

> **"We divided the patients into age groups, based on the median age of the patients, as follows: young, ≤40 years and old, >40 years; we did not use receiver operator characteristic (ROC) analysis of age for prediction of ARCD because it was poorly accurate (area under the curve [AUC] was 0.41)."**

We defined age groups based on the mean age of the study participants because no previous studies defined the cut-off age, and ROC analysis for age was poorly accurate for determining the specific cut-off value.

According to your comment, we have added the following phrases to clarify the reason why we defined age groups as young, ≤40 years and old, >40 years (page 9, line 20 to page 10, line 2):

> **"Because we could not find previous studies that defined a specific cut-off of age groups, we divided the patients into age groups, based on the median age of the patients, as follows: young, ≤40 years and old, >40 years; we did not use receiver operator characteristic (ROC) analysis of age for prediction of ARCD because it was poorly accurate (area under the curve (AUC) was 0.41)."**

Results

Comment 16:

"The median age was 38 years and 212 patients (57.5%) were male."

Is this for both groups? What is the point of this data?

Response:

We apologize for the lack of clarity. As you pointed out, we stated the basic information of both groups because this information was not presented in Table 1 (only comparative data between the two groups are listed in Table 1).

Comment 17:

Again, the authors concentrated on the statistical difference of single by a single parameter. This is not very helpful for the clinician. Authors should make Alvarado score comparison and to instruct and propose how to differentiate clinically and based on laboratory parameters between these 2 diseases. Maybe an additional parameter to Alvarado score should be added for the duration of pain (1 vs. 2 days) if the authors' opinion is that this is important for making a correct diagnosis.

Response:

Thank you for this important comment. Please refer to our responses to your previous comment numbers 10 and 14.

Discussion

Comment 18:

"... longer onset-to-visit interval..."

again what is longer interval, 1, 2, 3, 4 or 7 or 14 days.

Response:

Unfortunately, it is difficult to define "prolonged" of "longer interval" because the definitions of prolonged/longer period were different as mentioned above in our response to your comment number 7. In the present study, the difference of the interval was described as follows (page 12, lines 5-7):

> **"The onset-to-visit interval was 1 day longer in the ARCD group; the median interval was 1 day (IR, 0-1) in the AA group and 2 days (IR, 1-3) in the ARCD group ($p<0.001$)."**

Comment 19:

"previous history of diverticulitis"

How is that proven?

Response:

Thank you for this important comment. Please see our responses to your previous comment number 3.

Comment 20:

"previous history of diverticulitis"

Do authors mean right-sided diverticulitis or any colonic diverticulitis?

Response:

We apologize for the lack of clarity. "Previous diverticultis" in this study indicated diverticulitis of any parts of the colon. We have added "(including any parts of the colon)" in the "Materials and methods" (page 8, line 18).

Comment 21:

What about neutrophilia/left shift for the diagnosis? Please include that because this is a part of Alvarado Score.

Response:

Thank you for this important comment. Unfortunately, we could not evaluate neutrophilia/left shift in this retrospective study. We have revised the sentence in the paragraph about study limitations as follows (page 8, lines 18-19):

> **"…we could not collect some previously reported important information, such as parameters included in the Alvarado score[17], e.g., migration of abdominal pain or neutrophilia/left shift, despite previous studies clearly showing a high**

prevalence of pain migration and higher Alvarado score in AA compared to ARCD[9,11,12].”

Comment 22:

“Higher age in the ARCD group and higher leukocyte count in the AA group were observed in the univariate comparison although they were insignificant in logistic regression.”

This is a confirmation of my objections. There should be a scoring system used for the differentiation of these 2 entities. Please include Alvarado score and if possible male improved or modified scoring system. Maybe it should include the race also.

Response:

Thank you for your comment. Please refer to our previous response to your comment number 21. As mentioned above, we agree with your opinion, but we could not evaluate the Alvarado score because of the study design. We performed logistic regression instead of using the scoring system for diagnosis in the present study.

Comment 23:

“While contradicting evidence has been proposed, AA is traditionally thought to be initiated by elevation of intraluminal pressure of the appendix, which can be caused by luminal obstruction associated with fecalith, enlarged lymphoid tissue, barium, worms, or tumors, or appendiceal ulcer due to unknown etiology”

This is only one presentation of AA. Please do not simplify it. There are two pathophysiologic mechanisms – one is obstructive the other is inflammatory.

Response:

We apologize for oversimplifying the pathophysiology of appendicitis, and we agree with your opinion. That is, we do not think that all cases with appendicitis are caused by appendiceal obstruction, although most articles, authorized textbooks, and online reviews such as “Up to date (R)” describe appendiceal obstruction as a primary cause of appendicitis. Therefore, we did mention the contradicting study in the manuscript with the corresponding citation in the manuscript before revision: “While contradicting evidence has been proposed[18]…”

As per your suggestion, we have also revised the manuscript as follows to clearly indicate that appendicitis is caused by both appendiceal obstruction and

inflammation to avoid the readers from misunderstanding that all cases of appendicitis are caused by obstruction alone (page 14, lines 10-):

> **"While contradicting evidence has been proposed[18], most cases of AA are traditionally thought to be initiated by elevation of the intraluminal pressure of the appendix (with concurrent inflammation),…"**

Comment 24:

Also the reference is from the book (Internal medicine book ???). The reference should be from the experimental work that has the proof of your concept.

Response:

Thank you for the comment. We have reviewed several original experimental articles and review articles, and decided to cite the chapter of *Harrison's Principles of Internal Medicine* written by Dr. William Silen. We think that this citation is appropriate to support discussion about the pathophysiology of appendicitis even it is a textbook because of following reasons:

1) **The content was the most informative and comprehensive in the articles and books we reviewed probably because the chapter cited and integrated many citations.**
2) **The author of the chapter, Dr. William Silen, is one of the famous authorities in diagnosing acute abdomen.**
3) **The chapter does cite several review articles published from reliable peer-review journals.**
4) ***Harrison's Principles of Internal Medicine* is a textbook with an established reputation.**

Thus, we have left the citation for *Harrison's Principles of Internal Medicine*.

As per your suggestion, we have added reference #18 (Arnbjörnsson E, Bengmark S. Role of obstruction in the pathogenesis of acute appendicitis. *Am J Surg* 1984; 147: 390-392 [PMID: 6322604 DOI: 10.1016/0002-9610(84)90174-0]**) as new citation for explaining the pathophysiology of appendicitis.**

Comment 25:

What is the percentage of this classic migration of pain? This is important because significant number of patients does not have this migration. Please include references that are original articles for the percentages.

Response:

We apologize for the lack of information on migration. At first, we thought that it was not appropriate to discuss the migration of pain because we could not collect the patients' information on migration in this retrospective study that used medical records. However, we reconsidered this and have noted the previously reported prevalence and significance of the migration of pain according to your comment. As per your suggestion, we have added the following sentence (page 14, lines 22 to page 15 line 8):

"Previous studies comparing AA and ARCD reported that the prevalence of the migration of pain in AA was 39.2-82.0%, which is a significantly higher prevalence than that in ARCD (15.4-54.0%)[9–11]. Another review on AA reported that the sensitivity and specificity of the migration for the diagnosis of AA were 64% and 82%, respectively[21]. Although we lacked the prevalence of the migration of pain in the present study, our result indicating a significantly high OR of nausea/vomiting and anorexia in the logistic regression model for differentiating AA from ARCD is compatible with the generally accepted pathophysiology described above and that discussed in previous studies."

Comment 26:

"In four previous studies, two studies showed a significantly high proportion of nausea/vomiting in AA cases; the proportions were 8-16% in ARCD groups and 32-72% in AA groups [10,12]. The proportions were insignificant in the other two studies"

This is also a confirmation of my objection that scoring system should be used for the diagnosis, not the comparison of a single parameter.

Response:

Thank you for your comment. As previously mentioned, we agree with your opinion that we should not depend on single parameters; therefore, we performed logistic regression analysis instead of using a scoring system in consideration of the limitation of data collection due to retrospective study design. The aim of the current study was not comparisons of single parameters. Please refer to our previous response to your comment number 13.

Comment 27:

"Similar to a previous study, our univariate comparison showed that a higher median age of patients in the ARCD group was comparable to the AA group. But it was insignificant in logistic regression. Increase of diverticulosis in proportion to patients' age and higher prevalence of AA in younger patients may explain the difference of age."

What is the differentiation important for disctionction of these 2 disease. Can we say that patients several years younger can be strictly diagnosed with AA not ARCD – again scoring system is needed.

Response:

We agree with your opinion. Please refer to our previous responses to your comment numbers 13 and 26. According to your comment, we have added the following sentence at the end of the paragraph about study limitations (page 20, lines 12-14):

> **"Further multivariable analysis or scoring system studies that address the limitations of the present study are warranted."**

Comment 28:

"On the other hand, the age of the patients with ARCD is reportedly younger than the age of patients with left colonic diverticulitis"

Why is this info important for your study and clinical differentiation between AA and ARCD.

Response:

We apologize for the lack of clarity. In this paragraph, we just compared the difference in the result of age between our study and a previous study, and we intended to explain a possible reason why our study did not show a significant difference in age between the two groups. Again, this is just a comparison between our study result and results of previous studies.

Comment 29:

"We believe our findings will provide new evidence on utility of CRP for diagnosis of acute abdominal conditions."

How is there an algorithm based of these results? Can you again discuss Alvarado Score?

Response:

Thank you for this comment. This result is based on the result of logistic regression analysis as mentioned in our responses to your comment numbers 13 and 26. Please also refer to our previous responses to your comment numbers 13 and 26 about the Alvarado score.

Acknowledgement for the reviewer 00069988:

Thank you for your focused, clear, and well-structured comments. They were very valuable and helped us improve our manuscript. Please understand that we agree with your opinion that we should not depend on single parameters alone. Although scoring systems such as the Alvarado score are good methods, we performed logistic regression instead of using a scoring system because of the lack of data in the present retrospective study.

Response to reviewer 03479745's comments

1. Title.

Comment 1:

The authors could probably add their methodology into the study title (STROBE statement). This would probably help portray a more accurate reflection of the hypothesis/methodology of this manuscript.

Response:

Thank you for your comment on the title. We have revised the title as follows (page 1, lines 5): "Clinical differentiation of acute appendicitis and right colonic diverticulitis: a case-control study." The revised title has 12 words, which is the maximum word limit according to the guidelines for authors.

2 Abstract.

Comment 2:

The abstract is well summarized and reflects the work, the authors could perhaps provide a better conclusion rather than to merely just report the findings that have been reflected already in their results section of the abstract. It defeats the purpose of the conclusion section if you are merely reporting the results which you have already represented earlier.

Response:

Thank you for your comment. We agree that the conclusion should include what we would like to convey to the readers. As per your suggestion, we have revised the conclusion of the "Abstract" as follows (page 5, lines 18-20): "Clinical findings can differentiate AA and ARCD before imaging studies; nausea/vomiting and anorexia suggest AA, and longer onset-to-visit interval, RLQ pain, previous diverticulitis, and CRP level >3.0 mg/dL suggest ARCD." This revised sentence has 30 words, the maximum of the conclusions according to the guidelines for authors.

3 Key words.

Comment 3:

The keywords reflect the manuscript purpose.

Authors may want to consider removing "nonetheless" as the opening word in their core tip section.

Response:

Thank you for your comment. As per your suggestion, we have removed "Nonetheless" from the "Core tip" section (page 6, line 4).

4 Background.

Comment 4:

The authors have set the stage well for this uniquely Asian problem in their background. I applaud them for identifying this gap in literature and attempting to contribute to literature with this methodology.

Response:

Thank you for your positive feedback. We would like to applaud the researchers who published the previous studies we cited.

5 Methods.

Comment 5:

The authors have described their methodology well and appears to have done due diligence accordingly as well. However, may I suggest the following:

1. Under the Design and patients section

A. "Based on the findings of the CT scan and an ultrasound, patients with AA were diagnosed with complicated appendicitis if they had gangrenous appendicitis, perforated appendicitis, or appendicitis complicated with an intra-abdominal abscess." This statement is redundant and may appear to be confusing when the earlier sentence says they have already confirmed all diagnosis by CT. It can simply be said the both simple and complicated appendicitis were included in this study. As there is no subgroup analysis in this group of complicated appendicitis, there is no need to specifically address this point now. Furthermore, the authors discuss extensively later in the manuscript their rationale for including this group which addresses their inclusion.

Response:

We apologize for the redundant sentence about the inclusion of complicated appendicitis. As per your suggestion, we have removed the sentence "Based on

the findings of the CT scan and an ultrasound, patients with AA were diagnosed with complicated appendicitis if they had gangrenous appendicitis, perforated appendicitis, or appendicitis complicated with an intra-abdominal abscess.” and revised it as follows to simplify the sentence (page 8, lines 8-9): “We included both simple and complicated appendicitis in the AA group.”

B. "All CT findings were reviewed by several different radiologists and surgeons. " Why is this line used again after stating that it was diagnosis was achieved with CT at clinical presentation? Does it imply that retrospectively the CT images and diagnosis was reviewed again after the IRB was approved and this study conducted or was it just obtaining data from the reports? Appears confusing again, especially if 2 lines before this it was mentioned that diagnosis was confirmed by CT.

Response:

We apologize for the confusion. We intended to state that several different surgeons and radiologists reviewed the CT scans of all patients when they were admitted. As per your suggestion, we have removed the sentence to prevent confusion among readers.

C. "Patients with previous history of appendectomy were excluded because appendicitis was quite unlikely for those patients." Please remove "because appendicitis was quite unlikely for those patients" as this is understood implicitly and it is a fair inclusion criteria.

Response:

We agree with your comment and apologize for the redundancy. However, another reviewer did not understand the reason why we evaluated history of unresected appendicitis and requested a detailed explanation. Therefore, we have added the following sentences:

> **“We reviewed whether the patients had a history of acute appendicitis that was treated without appendectomy because previous appendicitis is a well-known risk factor of recurrent appendicitis if appendectomy was not performed[13].”**

D. State exclusion criteria for Diverticulitis

Does your cohort of patient include complicated diverticulitis? If not would I suggest you state explicitly at the start of your article that this study involve comparing acute appendicitis (both simple and complicated) with acute simple diverticulitis. If it

included complicated diverticulitis, can I suggest you include that explicitly as well as well as reflect the numbers and proportion somewhere in your article. Also with regards to this, is the protocol of the centre thereafter to perform a colonoscopy or any other form of repeat imaging to confirm the diagnosis of diverticula in the right colon? If there was, you could mention it.

Response:

Thank you for this valuable comment. We did include cases with complicated diverticulitis in the ARCD group. Therefore, we have added the following sentences at the end of the "Design and patients" subsection to explain that we included cases of complicated diverticulitis (page 8, lines 10-11):

"We included both simple and complicated right colonic diverticulitis in the ARCD group."

To clarify the number and proportion of cases of complicated diverticulitis in the ARCD group, we have also added the following sentence in the "Results" section (page 11, lines 17-18):

> **"On the other hand, 10/133 patients (7.5%) with ARCD were diagnosed with complicated right colonic diverticulitis."**

We have also added the following sentence in the paragraph about limitations in the "Discussion" (page 20, lines 6-11):

> **"We also included 10 (7.5%) patients with complicated right colonic diverticulitis in the ARCD group. Although we took priority in examining appendicitis or diverticulitis in the present study, further studies on the clinical difference of simple and complicated ARCD are required. Because this study included potentially different cases in the same groups, we have to be cautious in applying our results to individual cases."**

Unfortunately, we do not routinely perform colonoscopy or any other repeat imaging to confirm the diagnosis of right colonic diverticula. Because this study was retrospective, we could not obtain retrospective confirmation of right colonic diverticula in all cases. To address this issue, we have added the following sentence in the paragraph about study limitations (page 19, lines 3-10):

"As mentioned earlier, we diagnosed AA and ARCD based on the findings of CT because it has been the most common diagnostic tool of AA and ARCD in

Japanese clinical practice thanks to the fact that CT is the most available modality[6]. We believe that we could appropriately diagnose AA and ARCD by CT because its respective sensitivities and specificities are reportedly 90-100% and 91-99% for diagnosing AA[20], and 94% and 99% for diagnosing colonic diverticulitis[30]. Of note, we lacked the confirmation of diverticula by colonoscopy or barium enema examination in the present retrospective study."

2. Study variables

Comment 6: Most appear appropriate - why did the authors choose to include ALT as a variable?

Response:

Thank you for your comment. We recognize that ALT has not been reported as a potential confounder in any previous studies, and therefore, ALT may not be necessarily evaluated as a variable in this study. However, we collected and evaluated ALT, which is commonly measured in daily practice just to ensure that liver function abnormality was not a confounder of the study. In order to explain this, we have added the following sentence in the "Study variables" subsection (page 9, lines 6-9):

"Although ALT has not been reported as a potential confounder in any previous studies, we collected and evaluated the ALT level to ensure that liver function abnormality was not a confounder in this study."

3. Categorization of continuous variable

Comment 7:

From a methodology point of view, I can understand what the authors are trying to do with regards to using AUROC for CRP but is there a reason they chose to do this rather than using literature defined cut-offs of CRP from previous studies looking at this issue? Issue being using your own cohort to establish this cut off would introduce an element of reproducibility, especially since this was a study done retrospectively in nature. Otherwise, the authors have done well to explain their methodology from a univariate to a multivariate perspective.

Response:

Thank you for your comment. We agree that the cut-off value of the CRP level in our study can be regarded as arbitrarily defined because it was not previously defined. However, we defined the cut-off value of the CRP level based on the

result of ROC analysis because the cut-off values defined in previous studies vary widely. Furthermore, as mentioned in the manuscript, no previous study has compared CRP levels between ARCD and AA (page 17, lines 2).

In order to clarify why we defined the cut-off value of the CRP level based on the result of ROC analysis instead of using previously defined cut-off values, we have added the following sentence with relevant citations in the paragraph about study limitations in the "Discussion" (page 19, lines 11-16):

"Second, the cut-off value of the CRP level in our study can be regarded as arbitrarily defined because it was not previously defined. However, we determined 3.0 mg/dL as the cut-off value based on the result of ROC analysis (Figure 1) because of the lack of commonly used cut-off values of the CRP level; the cut-off values for diagnosing AA or ARCD vary widely between 3.0 and 20 mg/dL depending on the study[8,25–27]."

6 Results.

Comment 8: The results are appropriately represented.

Response: Thank you for the positive feedback.

7 Discussion.

Comment 9:

Does the manuscript interpret the findings adequately and appropriately, highlighting the key points concisely, clearly and logically? Are the findings and their applicability/relevance to the literature stated in a clear and definite manner? Is the discussion accurate and does it discuss the paper's scientific significance and/or relevance to clinical practice sufficiently? The authors once again summarize the results in the first paragraph again, this is unnecessary. They should consider using the last 2 lines of their first paragraph only "Previous studies have reported prolonged pain and higher age as predictors of ARCD and nausea/vomiting and leukocytosis as predictors of AA; most of our results were consistent with the results of previous studies [9–12]. On the other hand, history of diverticulitis, RLQ pain, and high serum CRP levels have not been previously established as predictors of ARCD."

Response:

Thank you for your comment. We tried to summarize the result again to make it easier for readers to understand. However, as you suggest, these sentences may be redundant. As per your recommendation, we have removed these sentences from

the "Discussion" and left only the following sentences (page 13, lines 20 to page 14, line 2):

"Previous studies have reported prolonged pain and higher age as predictors of ARCD and nausea/vomiting and leukocytosis as predictors of AA; most of our results were consistent with the results of previous studies[9–12]. On the other hand, history of diverticulitis, RLQ pain, and high serum CRP levels have not been previously established as predictors of ARCD."

Comment 10:

The authors then go on to describe the pathophysiology and relationship with clinical symptoms which is relevant. However, I have issue with this 2 lines at the end of a paragraph - "Interestingly, our study showed that 93.9% of the patients without RLQ pain had RLQ tenderness during palpation. This discrepancy of symptoms and physical findings may underscore the importance of careful palpation in patients with acute abdominal pain regardless of the position of abdominal pain." On many counts, I do not feel this statement should be included. Not only is it confusing as there is no discrepancy, I do not think the authors should be making this point based on their study hypothesis, simply for the fact that RLQ tenderness also exists in appendicitis, also in view of the results they represent. From my understanding based on their hypothesis background and results based on the methodology performed, patients who present with RLQ pain as symptoms are more likely to have diverticulitis, the role of RLQ tenderness, is not useful in differentiating AA from ARCD.

Response:

We apologize for the confusing text. We intended to mention that 93.9% of AA patients without RLQ pain had RLQ tenderness. Our statement is actually limited to the AA group. However, we agree with your comment, "I do not think the authors should be making this point based on their study hypothesis." Now we think that the discrepancy between RLQ pain and RLQ tenderness in patients with AA should be discussed separately from the present study. Thus, as per your suggestion, we have removed the following sentences from the "Discussion:" "Interestingly, our study showed that 93.9% of the patients without RLQ pain had RLQ tenderness during palpation. This discrepancy of symptoms and physical findings may underscore the importance of careful palpation in patients with acute abdominal pain regardless of the position of abdominal pain."

Comment 11:

I think the authors must be cautious in their discussion of CRP and Leukocytosis in this setting of clinical differentiation. It just simply falls into the concept of "it may be statistically significant but is it clinical significant?" I would strongly urge the authors to re-word this portion of the discussion to highlight this issue. So are you therefore proposing based on your study a CRP value of this is the clinical differentiator? Also not withstanding the fact that the appendicitis group has a proportion of complicated appendicitis?

Response:

Thank you very much for this important comment. Honestly, based on our personal experience as clinicians, both the CRP level and leukocytosis are useful for differentiating between AA and ARCD, especially in patients with simple appendicitis. However, our study showed a positive result for the CRP level and a negative result for leukocytosis. Furthermore, there are discrepancies between our study and previous studies on the significance of leukocytosis. Considering these facts, we agree with your comment. We have reconsidered whether to highlight the utility of CRP in our study. As per your suggestion, we have made the following major revision in the "Discussion."

1) **We have moved the following sentence to just after the suggestions from previous studies on CRP (page 17, lines 22 to page 18, line 4): "Although previous studies suggested that lack of systemic signs of toxicity and fever make ARCD more likely compared to AA, they did not mention CRP. Therefore, to the best of our knowledge, this is the first study to discuss differences of CRP between appendicitis and diverticulitis."**
2) **We have started a new paragraph after the paragraph on leukocytosis as follows (page 17, lines 19 to page 18, line 4):**

   "We believe that our findings will provide new evidence on the utility of CRP and leukocytosis for diagnosing acute abdomen. However, considering the inconsistent clinical significance of CRP for diagnosing acute abdomen in previous studies and the discrepancy of significance of leukocytosis between previous studies and the present study, a cautious attitude is required when applying our results to individual patients. We will separately discuss the heterogeneity of cases (simple and complicated) in the present study in the following paragraphs about our study limitations."

Comment 12:

also there is not date if the ARCD group has any patients with complicated diverticulitis (which hopefully the authors will address)

Response:

Thank you for this comment. Please refer to our response to your comment number 5.

8 Illustrations and tables.

Comment 13: Adequate.

Response: Thank you for the positive feedback.

9 Biostatistics.

Comment 14: Adequate

Response: Thank you for the positive feedback.

10 Units.

Comment 15: Meets standard unit references.

Response: Thank you for the positive feedback.

11 References.

Comment 16: Appears appropriate.

Response: Thank you for the positive feedback.

12 Quality of manuscript organization and presentation.

Comment 17:

The manuscript is appropriately organised and presented. There is no glaring issues with language and grammar. This is commendable as the authors are not native speakers of English.

Response: Thank you for the positive feedback.

13 Research methods and reporting.

Comment 18:

The authors have done well to come up with this hypothesis and their aim has been clearly spelt out. Their methodology was extensively described as well. They have probably complete most of the STROBE Statement. However, they did not mention the use of STROBE in their methodology in the manuscript but only in their cover letter

which I would encourage them to do if they have fulfilled the checklist. This would certainly make their manuscript more robust.

Response:

Thank you for your comment. Because we completed the STROBE checklist, we have added the following sentence at the end of the "Materials and Methods" (page 11, lines 9-10):

"The manuscript was written according to the STROBE Statement—checklist of items."

14 Ethics statements

Comment 19:

Authors have met the ethical standards. Submitted the necessary documents for proof.

Response: Thank you for the positive feedback.

Comment 20:

In summary, I appreciate this article and feel it may contribute to literature in this unique situation of an Asian disease. The authors did indeed identify a gap in terms of this question with a lack of robust methodology but however in view of the retrospective nature of this study with its inherent limitations. This is the short-fall of this study methodology. I urge the authors to correct and revise based on the suggestions to enhance the strengths.

Response:

We really appreciate your great efforts of reviewing our manuscript and providing valuable comments. We are confident that our manuscript has improved thanks to your suggestions.

Acknowledgement for reviewer 03479745:

Thank you for your focused, clear, and well-structured comments. They were very valuable and helped us improve our manuscript.