Accuracy and Interrater Reliability of Point-of-Care Ultrasonography Image Interpretation for Intussusception

Published

2022

Journal Title

Pediatric Emergency Care

Version

Accepted Manuscript (AM)

DOI

10.1097/PEC.0000000000002786

Accuracy and interrater reliability of point-of-care ultrasonography image interpretation for intussusception

Kelly R. Bergmann[1], DO, MS, Marshal Khant[1], DO, MPH, Shea Lammers[1], MS, Alexander C. Arroyo[2], MD, Pablo Avendano[3], MD, MPH, Lindsey Chaudoin[4], MD, Stephanie G. Cohen[5], MD, J. Kate Deanehan[6], MD, Aaron E. Kornblith[7], MD, Samuel H. F. Lam[8], MD, MPH, Margaret Lin-Martore[7], MD, Laurie Malia[9], DO Kathryn H. Pade[10], MD, Daniel B. Park[11], MD, MBA, Adam Sivitz[12], MD, Keren Shahar-Nissan[13], MD, Peter J. Snelling[14], MBBS, Mark O. Tessaro[15], MD, Rosemary Thomas-Mohtat[16], MD, Valerie Whitcomb[1], MD, Adriana Yock-Corrales[17], MD, MSc, Paige Walsh[18] MS, Dave Watson[18], PhD, Manu Madhok[1], MD, in collaboration with the P2 Network Intussusception Study Group

Author Affiliations:

[1] Department of Emergency Medicine, Children's Minnesota, Minneapolis, MN, USA. kelly.bergmann@childrensmn.org; marshal.khant@childrensmn.org; shea.lammers@childrensmn.org; valerie.whitcomb@childrensmn.org; manu.madhok@childrensmn.org

[2] Department of Emergency Medicine, Maimonides Medical Center, Brooklyn, NY, USA. aarroyomd@gmail.com

[3] Division of Pediatric Emergency Medicine, University of Minnesota Masonic Children's Hospital, Minneapolis, MN, USA. aven0001@umn.edu

[4] Department of Emergency Medicine, Atrium Health's Carolinas Medical Center, Charlotte, NC, USA. lindsey.chaudoin@atriumhealth.org

[5] Department of Pediatrics and Emergency Medicine, Children's Healthcare Atlanta and Emory University, Atlanta, GA, USA. stephanie.cohen@emory.edu

[6] Department of Pediatric Emergency Medicine, Johns Hopkins Children's Center, Baltimore, MD, USA. jdeaneh1@jhmi.edu

[7] Departments of Emergency Medicine and Pediatrics, University of California, San Francisco, San Francisco, CA, USA. Aaron.Kornblith@ucsf.edu

[8] Department of Emergency Medicine, Sutter Medical Center Sacramento, Sacramento, CA, USA. docp2slam@gmail.com

[9] Department of Emergency Medicine, Division of Pediatric Emergency Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York Presbyterian-Morgan Stanley Children's Hospital, New York, NY, USA. lm3396@cumc.columbia.edu

[10] Department of Pediatrics, University of California, San Diego, CA, USA. kathrynpade@gmail.com

[11] Department of Pediatrics and Emergency Medicine, University of North Carolina, Chapel Hill, NC, USA. daniel.park@unc.edu

[12] Department of Emergency Medicine, Newark Beth Israel Medical Center, Newark, NJ, USA. asivitz@gmail.com

[13] Department of Emergency Medicine, Schneider Children's Medical Center of Israel, Petah Tikva, Israel. kerens411@gmail.com

[14] Department of Pediatric Emergency Medicine, Gold Coast University Hospital and Griffith University, Southport, Queensland, Australia. peter.j.snelling@gmail.com

[15] Division of Pediatric Emergency Medicine, Hospital for Sick Children, Toronto, Ontario, Canada. oliver.tessaro@gmail.com

[16] Department of Pediatric Emergency Medicine, Children's National Hospital, Washington, DC, USA. RThomasm@childrensnational.org
[17] Department of Emergency Medicine, Hospital Nacional de Niños: "Dr. Carlos Sáenz Herrera," CCSS, San José, Costa Rica. adriyock@gmail.com
[18] Department of Research and Sponsored Programs, Children's Minnesota, Minneapolis, MN, USA. paige.walsh@childrensmn.org; dave.watson@childrensmn.org

Address correspondence to: Kelly R. Bergmann, DO, Department of Emergency Medicine, Children's Minnesota, 2525 Chicago Ave S, Mail Stop 32-1488, Minneapolis, MN, 55404, [kelly.bergmann@childrensmn.org], Tel: (612) 813-6843, Fax: (612) 813-6114.

Short Title: Interrater reliability of ultrasonography for intussusception

Funding Source: This work was supported (in part) by a grant from the Research Committee of Children's Hospitals and Clinics of Minnesota

Conflict of Interest and Financial Disclosure: All authors have indicated they have no potential conflicts of interest or financial relationships relevant to this article to disclose.

Abbreviations: ED – Emergency Department; EM – Emergency Medicine; PEM – Pediatric Emergency Medicine: POCUS – Point-of-Care Ultrasound; RADUS – Radiology Department Ultrasound; RDMS – Registered Diagnostic Medical Sonographer

Word Count (excluding abstract, references, tables, and figures): 3,024

Study Group Authorship

P2Network Intussusception Study Group

Danielle Amona[1], MD, Taylor Anderson[2], MD, Carmen Avendano[3], MD, Amanda Balmages[4], DO, Deeksha Borkar[5], MD, Lea Cohen[6], MD, Maytal Firnberg[7], MD, Natalie Ford[8], MD, Marci Fornari[9], MD, Jessica Gomez-Vargas[10], MD, Danielle Hatt[11], MD, Ted Kouo[12], MD, Rachel Krieger[13], DO, Abby Lewis[13], MD, Eleanor Nicholson Thomas[14], MBBS, Sabrina Lue-Tam[15], MD

Affiliations:
[1] Department of Emergency Medicine, Newark Beth Israel Medical Center, Newark, NJ, USA. danielle.amona@gmail.com
[2] Department of Pediatrics, Atrium Health's Carolinas Medical Center, Charlotte, NC, USA. taylor.anderson@atriumhealth.org
[3] Division of Pediatric Emergency Medicine, University of Minnesota Masonic Children's Hospital, Minneapolis, MN, USA. aven0008@umn.edu
[4] Department of Emergency Medicine, Desert Regional Medical Center, Palm Springs, CA. amanda.balmages@gmail.com
[5] Department of Pediatrics, University of Minnesota Masonic Children's Hospital, Minneapolis, MN, USA. borka015@umn.edu
[6] Department of Emergency Medicine, Maimonides Medical Center, Brooklyn, NY, USA. lecohen@maimonidesmed.org
[7] Department of Emergency Medicine and Pediatrics, University of California, San Francisco, San Francisco, CA, USA. maytal.firnberg@gmail.com
[8] Department of Pediatrics, University of North Carolina, Chapel Hill, NC, USA. Natalie.Ford@unchealth.unc.edu
[9] Department of Pediatric Emergency Medicine, Children's National Hospital, Washington, DC, USA. mfornari@cnmc.org
[10] Department of Emergency Medicine, Hospital Nacional de Niños: "Dr. Carlos Sáenz Herrera," CCSS, San José, Costa Rica. jemagova@gmail.com
[11] Department of Pediatrics, University of California, San Diego, CA, USA. dhatt@health.UCSD.edu
[12] Department of Pediatric Emergency Medicine, Johns Hopkins Children's Center, Baltimore, MD, USA. tkouo@jhmi.edu
[13] Department of Pediatrics, Children's Healthcare Atlanta and Emory University, Atlanta, GA, USA. rachel.krieger@emory.edu; abby.lewis@emory.edu
[14] Department of Pediatric Emergency Medicine, Gold Coast University Hospital and Griffith University, Southport, Queensland, Australia. enthomas89@gmail.com
[15] Division of Pediatric Emergency Medicine, Hospital for Sick Children, Toronto, Ontario, Canada. sabrina.luetam@sickkids.ca

Author Contributions:

Study concept and design, acquisition of the data, analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, acquisition of funding,and study supervision: Dr Bergmann.

Study concept and design, analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, and study supervision: Drs. Madhok and Sivitz.

Study concept and design, analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content: Dr. Khant and Ms. Lammers.

Study concept and design, interpretation of the data and ultrasound studies, critical revision of the manuscript for important intellectual content: Drs. Arroyo, P. Avendano, Chaudoin, S. Cohen, Deanehan, Kornblith, Lam, Lin-Martore, Malia, Pade, Park, Shahar-Nissan, Snelling, Tessaro, Thomas-Mohtat, Whitcomb, and Yock-Corrales.

Interpretation of the data and ultrasound studies, critical revision of the manuscript for important intellectual content: Drs. Amona, Anderson, C. Avendano, Balmages, Borkar, L. Cohen, Firnberg, Ford, Fornari, Gomez-Vargas, Hatt, Kouo, Krieger, Lewis, Nicholson Thomas, Lue-Tam.

Study concept and design, analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, statistical expertise: Drs. Walsh and Watson.

Abstract

Objectives: To determine the accuracy and interrater reliability of 1) point-of-care ultrasound (POCUS) image interpretation for identification of intussusception and 2) reliability of secondary signs associated with intussusception among experts compared to novice POCUS reviewers.

Methods: We conducted a planned secondary analysis of a prospective, convenience sample of children aged 3 months to 6 years who were evaluated with POCUS for intussusception across 17 international pediatric emergency departments between October, 2018, and December, 2020. A random sample of 100 POCUS examinations was reviewed by novice and expert POCUS reviewers. The primary outcome was identification of the presence or absence of intussusception. Secondary outcomes included intussusception size and the presence of trapped free fluid or echogenic foci. Accuracy was summarized using sensitivity and specificity, which were estimated via generalized mixed effects logistic regression. Interrater reliability was summarized via Light's kappa statistics with bootstrapped standard errors (SEs). Accuracy and reliability of expert and novice POCUS reviewers were compared.

Results: Eighteen expert and 16 novice POCUS reviewers completed reviews. The average expert sensitivity was 94.5% (95% confidence interval [CI]: 88.6, 97.5) and the specificity was 94.3% (95% CI: 90.3, 96.7), significantly higher than the average novice sensitivity of 84.7% (95% CI: 74.3, 91.4) and specificity of 80.4% (95% CI: 72.4, 86.7). Kappa was significantly greater for expert (0.679, SE 0.039) compared to novice POCUS reviewers (0.424, SE 0.044; difference 0.256, SE 0.033). For our secondary outcome measure of intussusception size, kappa was significantly greater for experts (0.661, SE 0.038) compared to novices (0.397, SE 0.041; difference 0.264, SE 0.029). Interrater reliability was weak for expert and minimal for novice reviewers regarding the detection of trapped free fluid and echogenic foci.

Conclusion: Expert POCUS reviewers demonstrate high accuracy and moderate interrater reliability when identifying intussusception via image interpretation and perform better than novice reviewers.

Introduction

Point-of-care ultrasound (POCUS) has been shown to be a useful adjunct for the evaluation of pediatric abdominal complaints.[1-3] Recent investigations have shown that POCUS performs similarly to radiology-performed ultrasound (RADUS) for the identification of intussusception, with a diagnostic accuracy ranging from 92–98%.[4-9] However, routine use of POCUS for the evaluation of intussusception has been limited, in part due to varying levels of sonography experience among staff members and the relative nascency of POCUS within the field of pediatric emergency medicine (PEM).[10-13] Although nearly all PEM fellowship programs now include some form of POCUS training in their curriculum, both the educational content and competency requirements vary.[14] Given this variation in experience among clinicians, it is essential to understand the accuracy and reliability of POCUS findings among PEM physicians prior to the routine adoption of POCUS, particularly for more specific applications such as the identification of intussusception.

Although POCUS has been shown to be an accurate tool for the evaluation of intussusception among experienced sonologists, no studies have evaluated the accuracy and interrater reliability of POCUS image interpretation among PEM physicians with varying degrees of ultrasound experience. Therefore, our primary aim in the present study was to determine the accuracy and interrater reliability of POCUS image interpretation for the identification of intussusception among expert compared to novice PEM POCUS reviewers. Our secondary aim was to determine the interrater reliability of POCUS for the identification of secondary sonographic signs associated with intussusception.

Methods

Study Design and Setting

We conducted a planned secondary analysis of POCUS images obtained from a prospective, convenience sample of children aged 3 months to 6 years who were evaluated for intussusception between October, 2018, and December, 2020, in an international diagnostic study.[5] In the primary study, children were enrolled in 17 urban, pediatric emergency departments (EDs) across North and Central America, Europe, and Australia. Twelve sites were university-based academic medical centers with affiliated children's hospitals, four were large community children's hospitals, and one was a large urban academic center with a dedicated pediatric ED. For the present study, de-identified POCUS images/video clips from the primary study were reviewed.[5] The institutional review board or ethics committee at each participating institution approved this study.

Summary of Primary Study

Enrollment in the primary study occurred consecutively when an expert POCUS user was available.[5] All children received both POCUS and RADUS, and a total of 35 expert POCUS sonologists participated in enrollment. Novice POCUS reviewers did not participate in the primary study.

Our POCUS protocol recommended the use of a linear, high-frequency (10-5 MHz, 15-6 MHz) transducer. The manufacturers and models of ultrasound machines varied by site. Expert POCUS sonologists were instructed to place the patient in a supine position and move the transducer superiorly from the right lower quadrant to the hepatic flexure, laterally to the splenic flexure, inferiorly to the left lower quadrant, and across the mid-abdominal region at their discretion. We asked that at least two still images or video clips were obtained in transverse and

longitudinal planes in each quadrant, for a total of 8 images/clips. If fewer than 8 images/clips were obtained, all available images were included. If an intussusception was identified, POCUS sonologists were asked to capture at least one still image in the short axis plane of the intussusception and one still image in the longitudinal plane.

POCUS Image Database

Out of a total of 201 POCUS examinations available, a random sample of 100 examinations was selected for review. Size measurements and images with color Doppler flow were removed prior to review. Of the random sample of 100 POCUS examinations, the mean number of video clips per study was 8.0 (standard deviation [SD] 3.3), and the mean number of still images per study was 8.0 (SD 3.4). The design was fully crossed in that each POCUS reviewer evaluated all 100 randomly sampled examinations.[15] POCUS reviewers were blinded to the initial POCUS and RADUS interpretations from the primary study. De-identified POCUS examinations were accessed via a secure, HIPAA-compliant website (www.sonoclipshare.com). Reviews were collected and managed using a secure online database (REDCap).[16]

POCUS Reviewers

Expert POCUS reviewers were defined as those who completed an ultrasound fellowship, hold a Registered Diagnostic Medical Sonographer (RDMS) certification, or who have completed at least 20 POCUS examinations for intussusception with at least one positive result. Novice POCUS reviewers were defined as physicians not meeting the expert reviewer criteria and who were either a) pediatric or emergency medicine (EM) residents, b) PEM fellows, or c) pediatric or EM attending physicians with minimal POCUS experience. Additionally, novice reviewers

were required to have basic POCUS competency as determined by the expert POCUS user at their site. All POCUS reviewers watched a brief 20-minute video on POCUS findings for intussusception prior to performing the image review.

Outcomes

Our primary outcome was whether POCUS reviewers identified any intussusception (i.e., ileocolic or ileolileal) or no intussusception. This outcome was chosen over the identification of ileocolic intussusception alone because POCUS reviewers could not readily measure intussusception size on image review (e.g., measurement >2 cm). For the assessment of accuracy, the reference standard was the interpretation of both the original RADUS and POCUS; images with discordant RADUS and POCUS interpretations were excluded because the RADUS diagnosis likely could not be arrived at from the POCUS images. Our secondary outcomes were estimation of the size of intussusception as ≥2 cm or <2 cm based on the scale on POCUS images, and secondary sonographic signs associated with ileocolic intussusception, including the presence or absence of trapped free fluid or echogenic foci within the intussusception. Accuracy and reliability were assessed for the primary outcome, but only reliability was assessed for the secondary outcomes because a reference standard was not available.

Additional outcome measures included POCUS reviewer confidence in their evaluation and POCUS examination quality. POCUS reviewer confidence was dichotomized as low (i.e., not at all, somewhat, or moderately confident) or high (i.e., very or completely confident). POCUS examination quality was rated by an independent POCUS experts not involved in the examination reviews using a modified POCUS image quality (IQ) scale (**Supplement**).[17] Specifically, the modified POCUS IQ scale ranks seven items on a scale of 0, 1, or 2 for a

highest possible score of 10. Items scored in the modified POCUS IQ scale include depth, gain, probe control, visualization of relevant anatomy or landmarks, and completeness.

Statistical Analysis

The characteristics of both the expert and novice POCUS reviewers were summarized, including the level of medical training (EM vs pediatrics trained), completion of PEM fellowship, level of POCUS training, and number of years of POCUS experience (<2, 3-4, 5-6, 7-8, 9-10, >10).

Accuracy was assessed using sensitivity and specificity with respect to the reference standard. The sensitivity and specificity of the POCUS reviewers were summarized descriptively and modelled via generalized mixed effects logistic regression. The latter method models the dichotomous outcome of a correct interpretation as a function of the reference standard and the POCUS reviewers' experience and also accounts for rater and image random effects (estimated separately for experts and novices). The percentage of (latent) variation explained was used to summarize the degrees to which variation in the images and variation in the raters explained the (latent) variance.[18]

For the assessment of reliability, kappa statistics were used to measure agreement for our primary outcome. Kappa statistics were also applied to secondary outcomes. In particular, Light's kappa statistic was used to account for multiple raters and a fully crossed design.[15,19] The interpretation of Light's kappa statistic is the average Cohen's kappa statistic of all possible pairwise combinations of raters. Kappa statistics were compared between expert and novice POCUS reviewers, and standard errors (SEs) were estimated via bootstrapping (by resampling images). The Kappa level of agreement was interpreted as follows: 0-0.20 was interpreted as none, 0.21-0.39 as minimal, 0.40-0.59 as weak, 0.60-0.79 as moderate, 0.80-0.90 as strong, and

>0.9 as almost perfect.[20] All analyses were performed using R Statistical Software (version 4.1.1; R Foundation for Statistical Computing, Vienna, Austria).

Results

A total of 18 expert and 16 novice POCUS reviewers participated. The expert reviewers were more likely to have received PEM fellowship training and had more years of POCUS experience compared to the novices. Across all reviewers, most had received training in pediatrics (**Table 1**).

Of the 100 randomly sampled POCUS images from the database, the original POCUS and RADUS operators both classified 28 as having intussusception (ileocolic, n=26; ileoileal, n=2) and 71 as having no intussusception (**Figure 1**). One image had discordant diagnoses, and this image was excluded from the accuracy assessment. Median modified POCUS IQ score was 8 (IQR 7, 9).

The expert reviewers were more accurate at interpreting POCUS examinations (**Figure 2, Table 2**). Among the expert reviewers, the average sensitivity was 88.9% (SD 12.4) and the average specificity was 89.3% (SD 9.6), whereas among novice reviewers the average sensitivity was 77.0% (SD 14.3) and the specificity was 75.7% (SD 14.8). Mixed effects logistic regression models estimated the average expert sensitivity at 94.5% (95% confidence interval [CI]: 88.6, 97.5) and specificity at 94.3% (95% CI: 90.3, 96.7), which were significantly higher than the average novice sensitivity of 84.7% (95% CI: 74.3, 91.4) and specificity of 80.4% (95% CI: 72.4, 86.7). The percent of (latent) variation explained by POCUS examination images (37% and 34% for experts and novices, respectively) was higher than the variation explained by reviewers (10% and 6%).

Interrater reliability was moderate among expert and weak among novice POCUS reviewers for identification of any intussusception and for differentiating ileocolic from ileoileal intussusception (**Table 2**). For the binary interpretation of identification of any intussusception, kappa was significantly greater for expert (0.679, SE 0.039) compared to novice POCUS reviewers (0.424, SE 0.044; difference 0.256, SE 0.033). Similar findings were seen for the ordered 3-level interpretation of no intussusception versus ileoileal intussusception versus ileocolic intussusception.

The proportion of studies for which the reviewer had high confidence in their interpretation was greater for expert compared to novice reviewers, with average proportions of 77.6% (SE 3.9) for expert and 22.1% (SE 1.3) for novice POCUS reviewers. Notably, 10 novice users had high confidence in less than 10% of their evaluations.

For the secondary outcome measure of intussusception size, interrater reliability was moderate among expert and weak among novice POCUS reviewers (**Table 2**); kappa was significantly greater for expert (0.661, SE 0.038) compared to novice POCUS reviewers (0.397, SE 0.041; difference 0.264, SE 0.029). Interrater reliability was weak for expert and minimal for novice reviewers regarding detection of trapped free fluid and echogenic foci.

Discussion

In this study of pediatric and general EM physicians with varying levels of proficiency using POCUS, expert POCUS reviewers demonstrated high accuracy and moderate interrater reliability for the identification of intussusception via image interpretation. In contrast, novice POCUS reviewers demonstrated modest accuracy and weak interrater reliability, suggesting that curricula emphasizing image interpretation, in addition to image acquisition and clinical

integration, may be needed for residents and PEM fellows in training. Our results have implications for pediatric hospital systems seeking to expand POCUS use and training, as well as stakeholders in medical education.

As interest in utilizing POCUS continues to grow within pediatric EDs, it is important to develop evidence-based implementation strategies that improve patient outcomes without compromising safety. Recent research has focused on the use of POCUS for intussusception and has shown that POCUS has high sensitivity, specificity, and diagnostic accuracy for this application when compared to RADUS.[4,5,9,21-27] The use of POCUS for the evaluation of intussusception has several potential benefits to patients, including improvement of patient throughput in the ED, focusing the need for further diagnostic testing, and informing decisions regarding transfer from general to pediatric centers. In one recent multicenter study, 38% of children transferred from an outside hospital had a negative intussusception evaluation at the receiving hospital, suggesting that these transfers may have been avoidable assuming there was no other indications for transfer.[5] Recent studies have also shown that clinical pathways for intussusception using a POCUS-first strategy lead to reduced times to diagnosis and treatment, and shorter ED length of stay without significant increases in negative outcomes.[28,29] Despite these benefits, the identification of intussusception using POCUS is not routinely taught across pediatric residency or PEM fellowship programs. In a survey of PEM fellowship programs across the U.S., only 64% of fellows reported specific POCUS training for intussusception.[14] However, the need for greater POCUS training for intussusception is also becoming increasingly recognized. Two recent modified Delphi studies outlined core content for PEM POCUS training, and both included identification of intussusception.[30,31]

Several studies have examined whether novice POCUS reviewers perform similarly to experts in diagnosing intussusception. Tonson la Tour et al. recently showed that the diagnostic accuracy for the identification of POCUS was 92% among novice POCUS reviewers compared to 98% among experts, where novice reviewers were defined as having no previous experience with bowel sonography other than a one-day ultrasound course.[9] However, few expert POCUS reviewers were included in this study, leading to wide CIs when evaluating the difference between diagnostic accuracy among experienced compared to inexperienced POCUS reviewers.[9] Another study among PEM physicians who participated in a one-hour course followed by hands-on scanning showed that novice POCUS reviewers were able to identify intussusception with a sensitivity of 85% (95% CI 54-97) and specificity of 97% (95% CI 89-99) compared to RADUS.[4] Similar to previous studies, expert POCUS reviewers in our study demonstrated greater accuracy evaluating images compared to novices, who also tended to be less confident than experts. Our findings suggest that those with more POCUS experience may be best suited to perform this application in clinical practice. This has been suggested by one recent survey of PEM fellow and program leaders, which showed that PEM program leaders endorsed significantly greater competency for identifying intussusception on POCUS compared to fellows.[14] Similarly, it has been estimated that the sensitivity of POCUS for the identification of intussusception is 90% among PEM physicians with focused training in bowel sonography compared to 79% among those without formal training.[8] More recently, Hsiao et al. showed that misdiagnosis of intussusception by POCUS decreased from 44% to 13% after the implementation of formal POCUS training; however, misdiagnosis rates improved more among senior compared to junior physicians with less clinical experience.[32] Given the modest sensitivity

and specificity among novice reviewers in our study, it is likely that more focused training curricula, including image interpretation, is needed for residents and PEM fellows.

Although recommendations for PEM POCUS training include the identification of intussusception,[30,31,33] they do not differentiate between ileocolic versus ileoileal intussusception. This distinction is clinically important, as these patients are managed differently. Earlier studies have demonstrated moderate to strong interrater reliability for differentiating ileocolic from ileoileal intussusception. Park et al. showed that the reliability (kappa) of determining ileocolic versus ileoileal versus an equivocal interpretation was 0.85 (95% CI 0.68-1.0) when comparing the initial sonologist's interpretation to that of an expert reviewing the POCUS images.[34] The authors also noted that all discrepancies involved determination of ileoileal intussusception versus an equivocal interpretation, and that no discrepancies involved ileocolic intussusception.[34] In a more recent study, the interrater reliability (kappa) between two expert reviewers was 0.747 (95% CI 0.518-0.976) for identifying ileocolic versus ileoileal versus no intussusception.[5] Our results show a marginally lower ordered 3-level kappa among expert POCUS reviewers,[5] which may be a result of including many more reviewers in the present study or of the low number of examinations showing ileoileal intussusception (n=2). Further research may be needed to determine whether PEM physicians can accurately differentiate ileocolic from ileoileal intussusception, and this distinction should be incorporated into PEM POCUS training.

Limitations

Several factors may have influenced the reviewer's interpretation of POCUS examinations. These include the potential for reviewers to have reviewed images from their own institution (e.g., expert reviewers may have recognized images previously collected), variation in image

quality for each POCUS examination (i.e., poor image acquisition affecting image interpretation), and the number of images/clips obtained per examination. However, variation in IQ also reflects what is obtained in clinical practice when performing a POCUS examination on a young child who may be fussy or inconsolable. Further, the median modified POCUS IQ score was 8 out of 10, suggesting that most images were of high quality. Second, we removed measurements and color Doppler from all images to reduce bias in the reviewer's interpretations, but omitting such information may have influenced reviewer ratings as well. We also did not include information regarding the location of the intussusception, which may have influenced the determination of ileocolic versus ileoileal intussusceptions.[35] Third, one criterion that we utilized to differentiate expert from novice POCUS reviewers was the completion of at least 20 POCUS exams for intussusception with at least one positive result, which is marginally less than what has been suggested for reaching competence.[36] However, we also defined expert reviewers as having completed an ultrasound fellowship or an RDMS certification, so it is unlikely that our expert reviewers lacked sufficient POCUS experience. Further, approaches that emphasize individualized training and direct observation,[37] particularly with regard to the assessment of technique and image interpretation,[33] may be more appropriate for determining POCUS competency. Lastly, image interpretation of POCUS images in isolation is not the same as obtaining images in real time at the patient's bedside, where the physician can integrate the history and physical examination findings into clinical decision-making. Reviewers in our study did not physically perform the POCUS examination or have access to clinical findings, which may have led to lower than expected interrater reliability among both novice and expert reviewers. This possibility is supported by one recent study showing that the sensitivity of POCUS improved from 87% to 97% after excluding physicians who had an inconclusive

physical examination,[9] suggesting that the incorporation of physical examination findings may impact the likelihood of identifying an intussusception on POCUS.

Conclusions

Using a large database of POCUS examinations for intussusception, we showed that expert POCUS reviewers demonstrate high accuracy and moderate interrater reliability for the identification of intussusception. In contrast, novice POCUS reviewers demonstrated modest accuracy and weak interrater reliability. Our results suggest that further training emphasizing image interpretation may be needed, in addition to hands-on training, prior to implementation in clinical practice.

References

1. Sivitz AB, Cohen SG, Tejani C. Evaluation of acute appendicitis by pediatric emergency physician sonography. *Ann Emerg Med.* 2014;64(4):358-364 e354.
2. Garcia AM, Asad I, Tessaro MO, et al. A Multi-institutional Case Series With Review of Point-of-Care Ultrasound to Diagnose Malrotation and Midgut Volvulus in the Pediatric Emergency Department. *Pediatr Emerg Care.* 2019;35(6):443-447.
3. Malcom GE, 3rd, Raio CC, Del Rios M, Blaivas M, Tsung JW. Feasibility of emergency physician diagnosis of hypertrophic pyloric stenosis using point-of-care ultrasound: a multi-center case series. *J Emerg Med.* 2009;37(3):283-286.
4. Riera A, Hsiao AL, Langhan ML, Goodman TR, Chen L. Diagnosis of intussusception by physician novice sonographers in the emergency department. *Ann Emerg Med.* 2012;60(3):264-268.
5. Bergmann KR, Arroyo AC, Tessaro MO, et al. Diagnostic Accuracy of Point-of-Care Ultrasound for Intussusception: A Multicenter, Noninferiority Study of Paired Diagnostic Tests. *Ann Emerg Med.* 2021;78(5):606-615.
6. Arroyo AC, Zerzan J, Vazquez H, et al. Diagnostic Accuracy of Point-Of-Care Ultrasound for Intussusception Performed by Pediatric Emergency Medicine Physicians. *J Emerg Med.* 2021.
7. Bhisitkul DM, Listernick R, Shkolnik A, et al. Clinical application of ultrasonography in the diagnosis of intussusception. *J Pediatr.* 1992;121(2):182-186.
8. Chang YJ, Hsia SH, Chao HC. Emergency medicine physicians performed ultrasound for pediatric intussusceptions. *Biomed J.* 2013;36(4):175-178.
9. Tonson la Tour A, Desjardins MP, Gravel J. Evaluation of bedside sonography performed by emergency physicians to detect intussusception in children in the emergency department. *Acad Emerg Med.* 2021;28(8):866-872.
10. Chamberlain MC, Reid SR, Madhok M. Utilization of emergency ultrasound in pediatric emergency departments. *Pediatr Emerg Care.* 2011;27(7):628-632.
11. Cohen JS, Teach SJ, Chapman JI. Bedside ultrasound education in pediatric emergency medicine fellowship programs in the United States. *Pediatr Emerg Care.* 2012;28(9):845-850.
12. Marin JR, Zuckerbraun NS, Kahn JM. Use of emergency ultrasound in United States pediatric emergency medicine fellowship programs in 2011. *J Ultrasound Med.* 2012;31(9):1357-1363.
13. Ramirez-Schrempp D, Dorfman DH, Tien I, Liteplo AS. Bedside ultrasound in pediatric emergency medicine fellowship programs in the United States: little formal training. *Pediatr Emerg Care.* 2008;24(10):664-667.
14. Cramer N, Cantwell L, Ong H, et al. Pediatric emergency medicine fellowship point-of-care ultrasound training in 2020. *AEM Educ Train.* 2021;5(4):e10643.
15. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol.* 2012;8(1):23-34.
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381.

17. Dessie AS, Calhou AW, Gilbert GE, et al. 112 Development and validation of a point-of-care ultrasound image quality assessment tool: the POCUS IQ scale. *Annals of emergency medicine.* 2019;74(4):S45-S46.
18. Rodriguez G, Elo I. Intra-class correlation in random-effects models for binary data. *The Stata Journal.* 2003;3(1):32-46.
19. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychological Bulletin.* 1971;76(5):365-377.
20. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276-282.
21. Lin-Martore M, Kornblith AE, Kohn MA, Gottlieb M. Diagnostic Accuracy of Point-of-Care Ultrasound for Intussusception in Children Presenting to the Emergency Department: A Systematic Review and Meta-analysis. *West J Emerg Med.* 2020;21(4):1008-1016.
22. Trigylidas TE, Hegenbarth MA, Patel L, Kennedy C, O'Rourke K, Kelly JC. Pediatric Emergency Medicine Point-of-Care Ultrasound for the Diagnosis of Intussusception. *J Emerg Med.* 2019;57(3):367-374.
23. Tsou PY, Wang YH, Ma YK, et al. Accuracy of point-of-care ultrasound and radiology-performed ultrasound for intussusception: A systematic review and meta-analysis. *Am J Emerg Med.* 2019;37(9):1760-1769.
24. Hom J, Kaplan C, Fowler S, Messina C, Chandran L, Kunkov S. Evidence-based diagnostic test accuracy of history, physical examination, and imaging for intussusception: a systematic review and meta-analysis *Pediatr Emerg Care.* 2020;Sep 15 [E-pub ahead of print].
25. Lee JY, Kim JH, Choi SJ, Lee JS, Ryu JM. Point-of-care ultrasound may be useful for detecting pediatric intussusception at an early stage. *BMC Pediatr.* 2020;20(1):155.
26. Arroyo AC, Zerzan J, Vazquez H, et al. Diagnostic Accuracy of Point-Of-Care Ultrasound for Intussusception Performed by Pediatric Emergency Medicine Physicians. *J Emerg Med.* 2021;60(5):626-632.
27. Lam SH, Wise A, Yenter C. Emergency bedside ultrasound for the diagnosis of pediatric intussusception: a retrospective review. *World J Emerg Med.* 2014;5(4):255-258.
28. Kim JH, Lee JY, Kwon JH, Cho HR, Lee JS, Ryu JM. Point-of-Care Ultrasound Could Streamline the Emergency Department Workflow of Clinically Nonspecific Intussusception. *Pediatr Emerg Care.* 2020;36(2):e90-e95.
29. Paek SH, Kim DK, Kwak YH, Jung JY, Lee S, Park JW. Effectiveness of the implementation of pediatric intussusception clinical pathway: A pre- and postintervention trial. *Medicine (Baltimore).* 2021;100(48):e27971.
30. Constantine E, Levine M, Abo A, et al. Core Content for Pediatric Emergency Medicine Ultrasound Fellowship Training: A Modified Delphi Consensus Study. *AEM Educ Train.* 2020;4(2):130-138.
31. Shefrin AE, Warkentine F, Constantine E, et al. Consensus Core Point-of-care Ultrasound Applications for Pediatric Emergency Medicine Training. *AEM Educ Train.* 2019;3(3):251-258.
32. Hsiao HJ, Wang CJ, Lee CC, et al. Point-of-Care Ultrasound May Reduce Misdiagnosis of Pediatric Intussusception. *Front Pediatr.* 2021;9:601492.

33. Vieira RL, Hsu D, Nagler J, et al. Pediatric emergency medicine fellow training in ultrasound: consensus educational guidelines. *Acad Emerg Med.* 2013;20(3):300-306.
34. Park BL, Rabiner JE, Tsung JW. Point-of-care ultrasound diagnosis of small bowel-small bowel vs ileocolic intussusception. *Am J Emerg Med.* 2019;37(9):1746-1750.
35. Zhang M, Zhou X, Hu Q, Jin L. Accurately distinguishing pediatric ileocolic intussusception from small-bowel intussusception using ultrasonography. *J Pediatr Surg.* 2021;56(4):721-726.
36. Abo AM, Alade KH, Rempell RG, et al. Credentialing Pediatric Emergency Medicine Faculty in Point-of-Care Ultrasound: Expert Guidelines. *Pediatr Emerg Care.* 2021;37(12):e1687-e1694.
37. Bell C, Hall AK, Wagner N, Rang L, Newbigging J, McKaigney C. The Ultrasound Competency Assessment Tool (UCAT): Development and Evaluation of a Novel Competency-based Assessment Tool for Point-of-care Ultrasound. *AEM Educ Train.* 2021;5(3):e10520.

Table 1. Characteristics of expert and novice POCUS reviewers

	Expert N=18 (%)	Novice N=16 (%)
Pediatrics residency training**	15 (83.3)	13 (81.3)
Emergency medicine residency training**	5 (27.8)	4 (25.0)
Fellowship training		
Pediatric emergency medicine	18 (100.0)	6 (37.5)
None	0 (0.0)	10 (62.5)
Current level of practice		
Attending, pediatric emergency medicine	18 (100.0)	2 (12.5)
Fellow, pediatric emergency medicine or POCUS	0 (0.0)	7 (43.8)
Resident, pediatrics or emergency medicine	0 (0.0)	7 (43.8)
POCUS experience, years		
0 – 3	1 (5.6)	15 (93.8)
4 – 6	4 (22.2)	0 (0.0)
7 – ≥10	13 (72.2)	1 (6.2)

*Fisher's exact test.

**Some reviewers reported dual training in both pediatrics and emergency medicine for residency.

POCUS, point-of-care ultrasound.

Table 2. Accuracy and interrater reliability between expert and novice POCUS reviewers

Accuracy	Expert (95% CI)	Novice (95% CI)	Odds Ratio (95% CI)
Sensitivity	94.5% (88.6, 97.5)	84.7% (74.3, 91.4)	3.1 (1.1, 8.7)
Specificity	94.3% (90.3, 96.7)	80.4% (72.4, 86.7)	4.1 (1.9, 8.4)

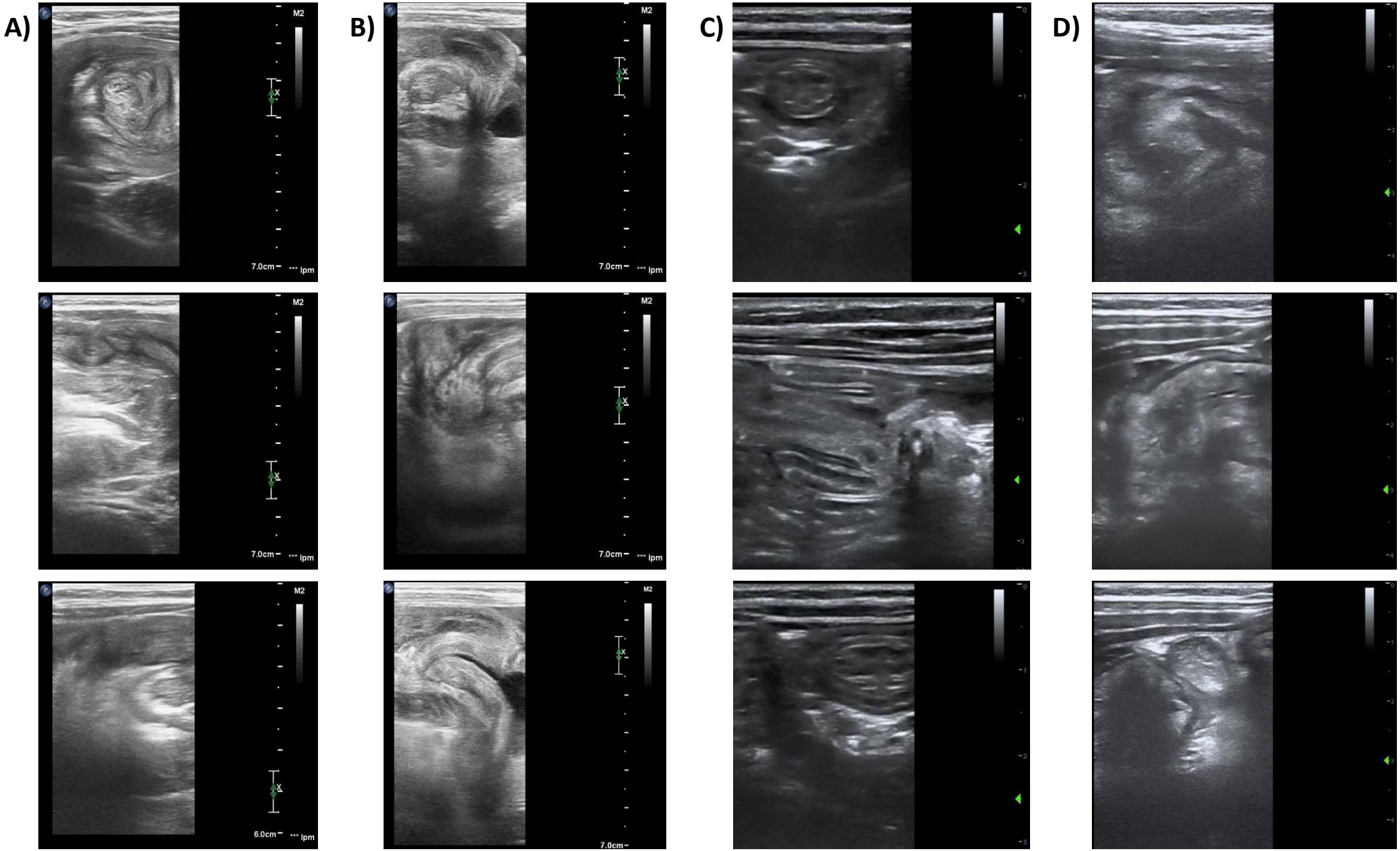
Interrater Reliability (Kappa)	Expert (Standard Error)	Novice (Standard Error)	Difference* (Standard Error)
No vs any intussusception	0.679 (0.039)	0.424 (0.044)	0.256 (0.033)
No vs ileoileal vs ileocolic intussusception	0.653 (0.036)	0.353 (0.036)	0.299 (0.028)
Size <2 or ≥2 cm	0.661 (0.038)	0.397 (0.041)	0.264 (0.029)
Trapped free fluid	0.517 (0.028)	0.307 (0.030)	0.210 (0.028)
Echogenic foci	0.517 (0.027)	0.291 (0.027)	0.226 (0.026)

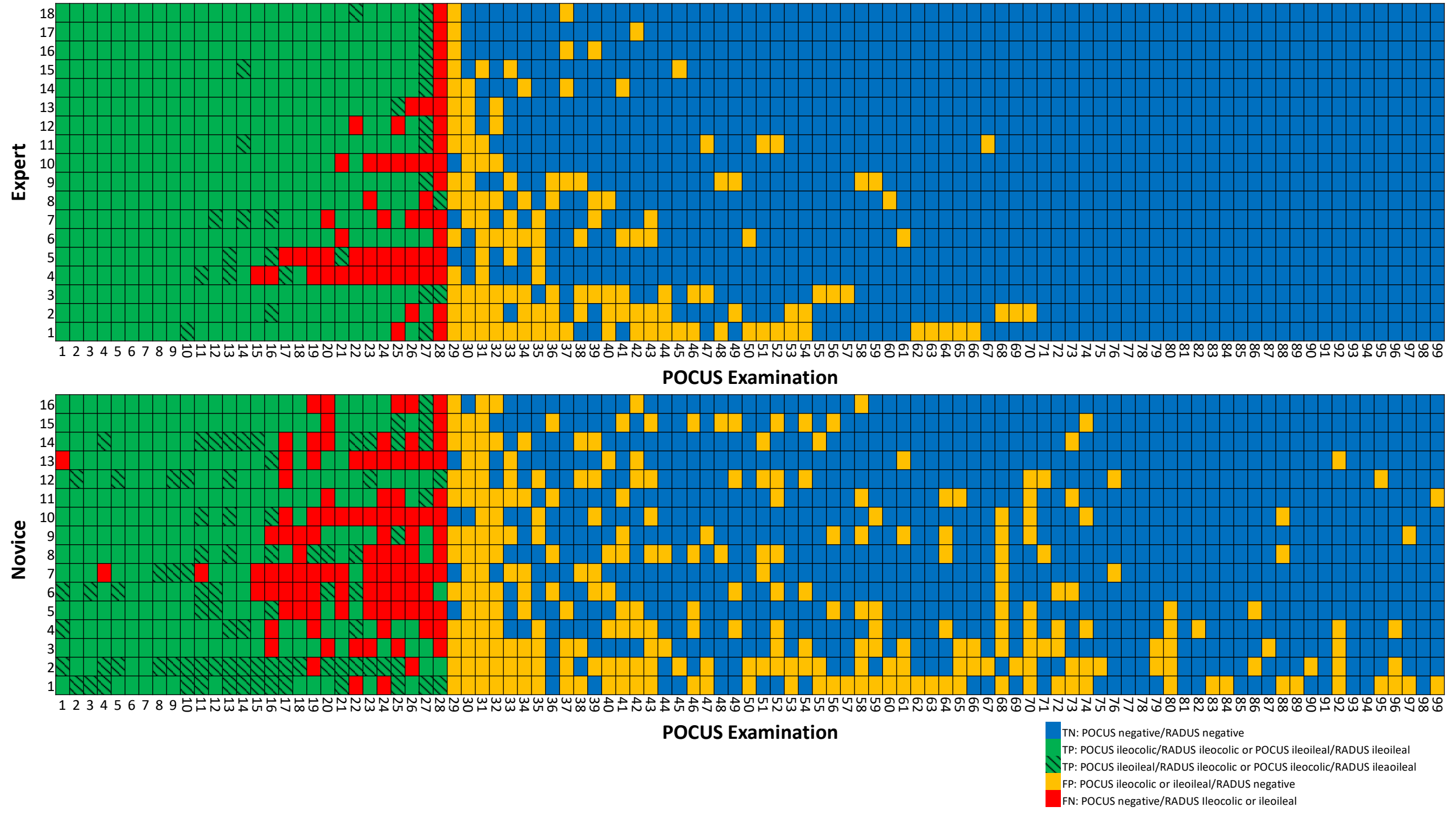
*For comparison of expert and novice POCUS reviewers, all p-values <0.0001.
Sensitivity and specificity were modelled via generalized mixed effects logistic regression, accounting for rater and image random effects (estimated separately for experts and novices).
POCUS, point-of-care ultrasound.

Figure Legends

Figure 1. Examples of POCUS examinations positive for ileocolic (A and B) and ileoileal intussusception (C), and negative for intussusception (D).

Figure 2. Accuracy of expert and novice POCUS reviewers for each POCUS examination.


A)
B)
C)
D)
M2
7.0cm
6.0cm
lpm


Expert
Novice
POCUS Examination
POCUS Examination
TN: POCUS negative/RADUS negative
TP: POCUS ileocolic/RADUS ileocolic or POCUS ileoileal/RADUS ileoileal
TP: POCUS ileoileal/RADUS ileocolic or POCUS ileocolic/RADUS ileaoileal
FP: POCUS ileocolic or ileoileal/RADUS negative
FN: POCUS negative/RADUS Ileocolic or ileoileal

POCUS IQ Scale User Guide

The POCUS IQ Scale is meant to evaluate a study in its entirety; all clips for a single application. For this study, it will be applied to all of the clips in a single day.

The scale includes 5 domains, each scored 0-2 on an ordinal scale. Provided are some descriptions of each domain and examples to orient you on suggested scoring.

Depth – Too shallow images, where important anatomy is missing, should score a 0 (inadequate). Well-centered images with area of interest in the focal zone score a 2 (ideal). If adjusting the depth would improve the image but is not entirely necessary, that is a 1 (adequate).

Gain/Presets – Images that are significantly over/undergained would be a 0 (inadequate). If the gain is at a level you would not change for the application that would be a 2 (ideal). Slightly over- or under- gained images would be a 1 (adequate).

Probe Control - This refers to overall smoothness in moving clips. Novice sonographers tend to have more jerky probe movements or lose contact with the skin (0), while experienced sonographers tend to have more fluid, smooth movements (2).

Anatomy/Landmarks – If every clip demonstrates both the superior and inferior aspect of the intussusception, that would correspond with a score of 2. If most clips demonstrate superior and inferior aspect of intussusception or the superior or inferior aspect are missing that would be a 1. If no clip demonstrates the correct landmarks, or if the anatomy is difficult to visualize due to angle of insonation, depth/zoom, or other artifacts- that would be a score of 0.

Completeness – Are all the appropriate views fully obtained? For example, ideal views include both upper and lower aspect of the intussusception for every clip/zone and views of the intussusception in both short and long axis would be scored a 2. If only 1 view of the intussusception was obtained or the intussusception is not visualized that would be scored a 1. If there was only partial visualization on the intussusception (e.g., bury image or quick view on cine clip) that would be 0. For studies with no identified intussusception, there should be multiple view at the appropriate depth (4-5 cm) with good visualization of bowel (2). If there is partial visualization of the bowel with only artifact from bowel gas that would be 1, and all views with predominantly all bowel gas would be 0.

	0	1	2
TECHNICAL			
Depth	Inadequate for visualization of area of interest	Adequate	Ideal
Gain/Presets	Inadequate for visualization of area of interest	Adequate	Ideal
SCANNING SKILL			
Probe Control	Poor probe control, frequent erratic movements	Fair probe control	Excellent probe control, consistently smooth movements
Anatomy/ Landmarks	Poor demonstration of anatomy/landmarks	Fair demonstration of anatomy/landmarks	Excellent demonstration of anatomy/landmarks
INTERPRETABILITY			
Completeness e.g. appropriate views, modes, measurements	Inadequate views for image interpretation	Adequate views	Ideal views
TOTAL SCORE ___ / 10			