Alsabri *et al. International Journal of Emergency Medicine* (2026) 19:41
https://doi.org/10.1186/s12245-026-01134-z

International Journal of Emergency Medicine

REVIEW **Open Access**

Accuracy of ultrasound for intussusception in pediatric emergency presentations: a systematic review and diagnostic meta-analysis

Mohammed Alsabri[1,2*], Shree Rath[3], Mohamed Amr Elkarargy[4], Amira A. Aboali[5], Khaled Abouelmagd[6], Abdelaziz Abdelaziz Abdelftah Ramadan[7], Luis L. Gamboa[8], Patrick Yoo[9] and Yisha Cheng[9]

Abstract

Introduction Intussusception is a common cause of acute abdominal emergencies in children. This systematic review and diagnostic meta-analysis aimed to determine the diagnostic accuracy of ultrasound for intussusception in pediatric emergency presentations, providing pooled estimates for sensitivity, specificity, predictive values, and diagnostic odds ratios to inform clinical practice.

Methods Adhering to PRISMA-DTA guidelines, a comprehensive search was conducted in PubMed, Scopus, Cochrane Library, and Web of Science up to July 2025. Bayesian bivariate random-effects meta-analyses were performed to estimate pooled sensitivity, specificity, and other measures, with subgroup and meta-regression analyses to explore heterogeneity.

Results A total of 44 studies comprising 4,142 pediatric patients were included in the quantitative synthesis. The pooled sensitivity of ultrasound for diagnosing intussusception was 96.3% (95% credible interval [CrI] 94.9–97.5%), and the pooled specificity was 95.7% (95% CrI 93.3–97.5%). The area under the hierarchical summary receiver operating characteristic curve (AUC) was 0.81–0.82, indicating good discriminative ability. Positive predictive value (PPV) ranged from 54.1% at 5% prevalence to 99.8% at 95% prevalence, while negative predictive value (NPV) decreased from 99.8% to 57.7% across the same prevalence range. The overall certainty of evidence for sensitivity and specificity was rated as high, with moderate certainty for prevalence due to substantial heterogeneity.

Conclusion Ultrasound demonstrates excellent diagnostic performance for pediatric intussusception in emergency settings, with high sensitivity and specificity maintained across patient subgroups and operator backgrounds. These findings support the continued use of ultrasound as the first-line diagnostic modality in both high- and low-resource environments and highlight the importance of structured training to optimize its accuracy. Future research should focus on multicenter prospective studies, standardization of ultrasound protocols, and the integration of artificial intelligence to further enhance diagnostic reliability.

*Correspondence:
Mohammed Alsabri
alsabri5000@gmail.com

Full list of author information is available at the end of the article

Keywords Intussusception, Emergency department, Pediatrics, Surgery, Point-of-care ultrasound

Introduction

Intussusception is a medical emergency characterized by the telescoping of a proximal segment of the intestine into an adjacent distal segment, resulting in luminal obstruction, venous congestion, and subsequent bowel wall edema [1,2]. Progressive vascular compromise leads to ischemia, mucosal necrosis, and, if untreated, bowel perforation and peritonitis [1,3]. It represents the most common cause of intestinal obstruction in infants and young children, and one of the leading causes of acute abdominal emergencies in this population [1–3].

The condition primarily affects infants aged 3 months to 3 years, with a peak incidence between 5 and 9 months of age [1–3]. Reported annual incidence rates vary substantially across regions, ranging from 3 to 40 cases per 10,000 live births in the United States, Europe, and Australia [4–6]. This variation partly reflects differences in case ascertainment, vaccination practices, and healthcare infrastructure. Notably, higher incidence rates have been documented in developing countries, where infectious enteritis, adenovirus, and rotavirus infections are more prevalent, and delayed access to healthcare contributes to underdiagnosed or late-presenting cases [7,8].

More than 90% of pediatric intussusception cases are idiopathic, though around 5–10% are attributed to identifiable pathological lead points. These include Peyer's patch hyperplasia, Meckel's diverticulum, duplication cysts, intestinal polyps, mesenteric lymphadenopathy, lymphoma, and postoperative adhesions [1,2,7]. The ileocolic type accounts for approximately 80–90% of cases, while small-bowel and colocolic variants are less frequent [1,2].

If unrecognized, acute intussusception can rapidly progress to bowel ischemia, necrosis, or perforation, resulting in peritonitis and shock [9]. Early recognition is therefore critical to prevent morbidity and mortality. Historically, contrast enema using air, barium, or water-soluble media served as both a diagnostic and therapeutic modality, achieving successful non-surgical reduction in 70–95% of cases [10]. However, this technique carries several limitations, including radiation exposure, risk of bowel perforation, operator dependency, and restricted availability in emergency or resource-limited settings [10,11].

In contrast, ultrasound (US) has emerged as the preferred first-line imaging modality due to its non-invasive nature, lack of ionizing radiation, bedside applicability, and high diagnostic accuracy [10]. Although ultrasound cannot directly assess bowel viability, it allows real-time evaluation of bowel wall perfusion, peristaltic motion, and presence of free fluid, which can indirectly suggest ischemia. Characteristic sonographic findings include the "target" or "doughnut" sign on transverse imaging, representing concentric intestinal layers and the "pseudokidney" sign on longitudinal scans reflecting invaginated bowel with mesenteric fat and vessels [11–13]. The sensitivity and specificity of ultrasound for diagnosing pediatric intussusception generally range from 85% to 100%, depending on operator experience, probe frequency, patient cooperation, presence of bowel gas, and timing of presentation [11–13].

Recent meta-analyses, each encompassing between 20 and 30 primary studies with pooled sample sizes exceeding 10,000 children, have reported summary sensitivity and specificity values of approximately 96% and 97%, respectively [14,15]. Despite these encouraging figures, significant methodological and population heterogeneity persists across studies. Variations in reference standards (contrast enema vs. surgical confirmation), operator training levels, and study designs (prospective vs. retrospective) continue to influence pooled estimates. Furthermore, prior reviews have primarily relied on bivariate frequentist models and lack Bayesian hierarchical modeling, subgroup analyses, and evaluation of small-study effects, which are necessary for robust diagnostic accuracy synthesis. Accordingly, further evidence synthesis is warranted to refine the pooled diagnostic performance of ultrasound in detecting pediatric intussusception, incorporating contemporary studies and advanced statistical modeling.

This systematic review and meta-analysis therefore aims to determine the diagnostic accuracy of ultrasound for pediatric intussusception, using contrast enema or surgery as reference standards. It will provide pooled estimates of sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratios, and predictive values. The findings are expected to guide diagnostic pathways in pediatric emergency and radiology settings, inform training and credentialing standards for point-of-care ultrasound, and support evidence-based resource allocation in both high- and low-resource healthcare environments.

Methods

Study design

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines [16]. The protocol was registered in PROSPERO prior to data extraction (CRD420251116473).

Eligibility criteria

The review included prospective and retrospective cohort studies, cross-sectional studies, and case–control studies evaluating the diagnostic accuracy of abdominal ultrasound for intussusception in children aged 0–18 years. Eligible studies used surgical findings, contrast (air or barium) enema, or combined clinical confirmation as the reference standard.

Studies were required to provide sufficient data to calculate sensitivity, specificity, predictive values, likelihood ratios, or diagnostic odds ratios. Case reports, reviews, letters, conference abstracts without data, animal studies, and studies involving adults were excluded.

Search strategy

A comprehensive literature search was performed in PubMed, Scopus, Cochrane Library, and Web of Science from inception to 27th July 2025. Search terms combined MeSH and free-text words related to intussusception, ultrasound, and pediatric. Reference lists of included studies and relevant reviews were also screened manually. A detailed search strategy is provided in Supplementary Table 1.

Study selection and data extraction

Two independent reviewers screened titles and abstracts, followed by full-text screening based on eligibility criteria. Data extraction included study characteristics (author, year, country, study design, patient number, mean age), ultrasound modality, operator type, reference standard, and diagnostic accuracy data (True Positive, False Positive, True Negative, False Negative). Disagreements were resolved by consensus or a third reviewer.

Quality assessment

Methodological quality and risk of bias were assessed independently by two reviewers using the QUADAS-2 tool [17], with any disagreements resolved through discussion or adjudication by a third reviewer. The QUADAS-2 framework evaluates studies across four key domains: (1) patient selection, assessing whether study participants were enrolled consecutively or randomly, and whether inappropriate exclusions or case-control designs introduced selection bias; (2) index test, evaluating whether the index test (in this case, abdominal ultrasonography for the diagnosis of intussusception) was interpreted without prior knowledge of the reference standard results and whether the test execution and interpretation were representative of clinical practice; (3)reference standard, determining whether the reference test used to the presence or absence of intussusception (such as contrast enema, surgical findings, or clinical follow-up) was likely to correctly classify the target condition and whether its results were interpreted independently of the index test; and (4) flow and timing, which assesses whether all patients received the same reference standard, the time interval between index and reference tests was appropriate, and whether all enrolled patients were included in the final analysis.

Each domain was rated for risk of bias as “low,” “high,” or “unclear,” and the first three domains were also evaluated for applicability concerns to ensure that study conditions reflected real-world clinical settings.

Statistical analysis

Meta-analysis was conducted in R using a combination of Bayesian and frequentist frameworks tailored to the data type and diagnostic outcomes. For the primary diagnostic accuracy outcomes, we employed a Bayesian bivariate random-effects model using the meta4diag package, which jointly modeled study-level sensitivity and specificity on the logit scale while estimating between-study variances (τ^2) and their correlation (ρ). Weakly informative half-Cauchy and Penalized Complexity (PC) priors were compared to assess model robustness under sparse data conditions.

For proportion-based outcomes (prevalence, responder rates, adverse events), we fitted generalized linear mixed models (GLMMs) using the metaprop function with logit transformation and random intercepts to estimate pooled proportions and heterogeneity (τ^2, I^2). Subgroup analyses and covariate effects (e.g., age, sex, symptom presence, operator specialty) were examined through Bayesian covariate extensions of the bivariate model, where α and β coefficients represented logit-scale effects on sensitivity and specificity, respectively.

Publication bias was evaluated using Deeks’ funnel plot asymmetry test, approximated by plotting the log diagnostic odds ratio against $1/\sqrt{\text{effective sample size}}$, with regression-based assessment of asymmetry. Model convergence, credible interval overlap, and between-prior agreement were used to ensure inferential stability and transparency across analyses.

Results

Study selection

A total of 3,580 records were obtained on an initial search, of which 1,335 records were screened following removal of duplicates. Of these, 64 were further assessed through a full-text screening. Finally, 44 records were included for qualitative and quantitative synthesis (Fig. 1).

Baseline characteristics

Across 44 studies, 4,142 patients were included. The sample size per study ranged from 47 to 127 children, with most studies enrolling children under 2 years of age. The mean age across studies varied between 1.7 and 6.5

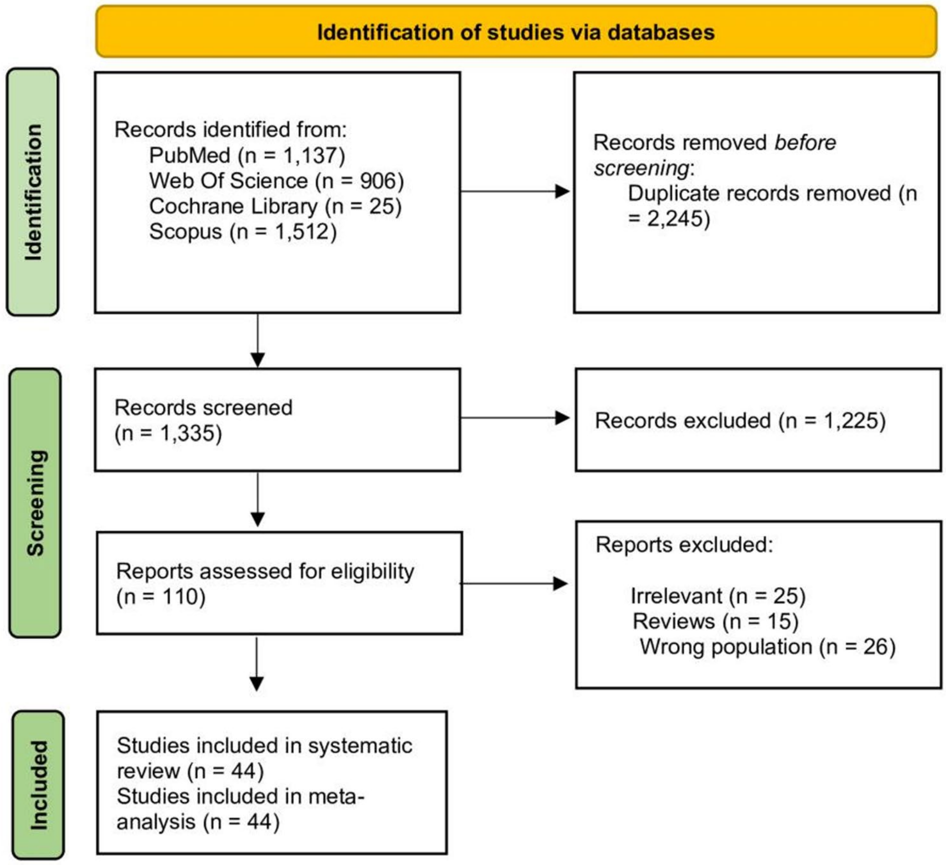

Fig. 1 PRISMA flowchart of screening and study selection

years, with a predominance of male participants (ranging from 58% to 70%). Duration of symptoms prior to presentation was inconsistently reported but generally occurred within the first 24–48 h of illness. Regarding ultrasound parameters, both radiologist-performed and point-of-care (POCUS) modalities were represented. In most studies, operators were emergency physicians or pediatric radiologists, with variable levels of ultrasound training and experience. High-frequency linear probes (7–10 MHz) were most commonly used, although some studies reported use of curvilinear or mixed-frequency probes. The most frequent identifiable pathological lead point was Meckel's diverticulum, with few cases attributable to polyps or volvulus. Reported causes for false-negative findings included technical limitations and dynamic reduction of intussusception prior to imaging. Detailed summary and baseline tables are provided in Supplementary Tables 1, 2.

Quality assessment

The methodological quality of the included studies was assessed using the QUADAS-2 tool. Most studies were judged to have low risk of bias across the domains of patient selection, index test, reference standard, and flow and timing. Three studies (Al-Ani 2017; Lam 2014; Gata 2016) were judged to be at high risk of bias due to concerns in patient selection. In addition, 12 studies were rated as unclear in one or more domains, most often relating to patient selection or interpretation of the index test. No study was judged to be at high risk in the domains of index test, reference standard, or flow and timing. Applicability concerns were generally low, with the majority of studies considered directly relevant to the review question of ultrasound diagnosis of intussusception in pediatric emergency settings, although a small subset had unclear concerns regarding patient selection. (Supplementary Fig. 1).

Bayesian modelling

We estimated the pooled diagnostic accuracy of ultrasound for pediatric intussusception using two Bayesian bivariate models in meta4diag. The primary model used H-Cauchy priors for the between-study variance (weakly informative), while the sensitivity analysis used penalized complexity (PC) priors, which shrink extreme variance estimates more aggressively. This approach allows us to check the robustness of pooled estimates without relying on leave-one-out or influence diagnostics, which are not available in meta4diag.

Both models produced very high and consistent estimates of diagnostic performance, demonstrating the clinical reliability of ultrasound in this context. The pooled sensitivity was 96.3% in both models (95% credible interval [CrI] 94.9–97.5% for H-Cauchy; 95.0–97.5% for PC), and the pooled specificity was 95.7% for H-Cauchy (95% CrI 93.3–97.5%) and 95.6% for PC (95% CrI 93.4–97.3%) (Figs. 2 and 3, Supplementary Figs. 2, 3). The credible intervals were notably tight, reflecting low uncertainty and reinforcing the precision of these pooled estimates (Table 1).

Leave-one-out diagnostics were not feasible in meta4diag; instead, penalized complexity priors were applied as a robustness check, which yielded consistent estimates. The PC prior yielded slightly lower between-study variance for specificity (var_psi 2.54 vs. 3.12) than the H-Cauchy prior, indicating its stronger shrinkage effect, but this did not materially alter the pooled Se/Sp. The correlation between sensitivity and specificity remained high in both models (0.78–0.79), consistent with the natural trade-off between false positives and false negatives; as well as, The hierarchical summary ROC curves demonstrated strong diagnostic performance, with areas under the curve (AUCs) of 0.82 (H-Cauchy model) and 0.81 (PC model) (Fig. 4). In this context, the AUC quantifies the overall ability of ultrasound to correctly distinguish children with intussusception from those without across all possible thresholds. Values above 0.80 are generally interpreted as good, supporting the robustness of our pooled accuracy estimates.

Clinically, these results highlight that ultrasound is a highly accurate, rapid, and non-invasive tool for diagnosing pediatric intussusception, capable of guiding prompt management decisions with minimal risk of misclassification.

Posterior variance distributions indicated that sensitivity estimates were consistently stable across studies, suggesting that high sensitivity of ultrasound is reproducible in most clinical contexts. Specificity showed greater between-study variability, particularly under the Cauchy prior, implying that false-positive rates may be more influenced by study setting or operator factors. Nevertheless, the similarity of the posterior curves under both priors supports the robustness of the pooled accuracy estimates.

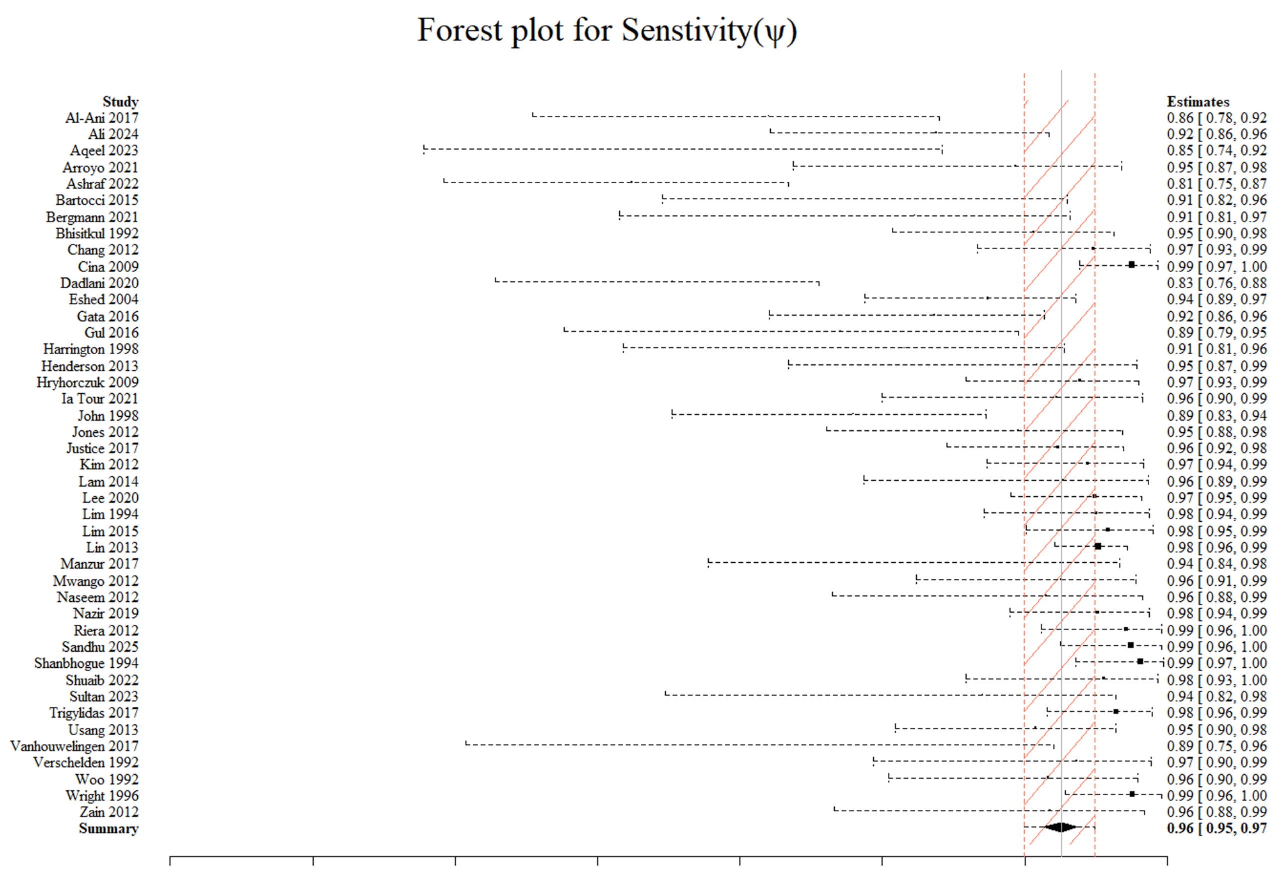

Fig. 2 Forest plot of study-level sensitivity, with 95% credible intervals. The pooled estimates are shown as diamonds

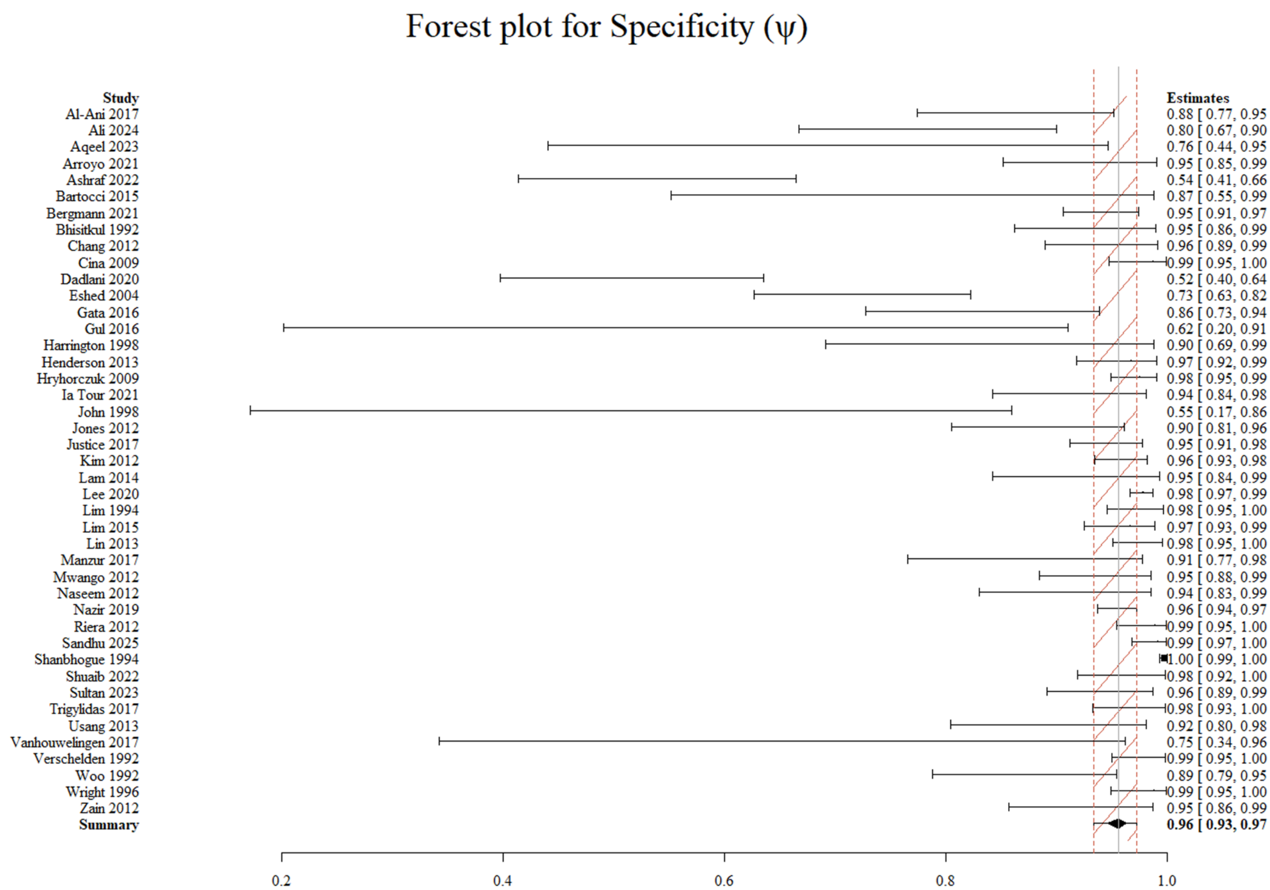

Fig. 3 Forest plot of study-level specificity, with 95% credible intervals. The pooled estimates are shown as diamonds

Table 1 Bayesian model estimates

Model	Metric	Estimate (%)	95% CrI (%)	Variance	Correlation
H-Cauchy	Sensitivity	96.3	94.9–97.5	1.12	0.79
H-Cauchy	Specificity	95.7	93.3–97.5	3.12	0.79
H-Cauchy	AUC	0.81	-	-	-
PC	Sensitivity	96.3	95.0–97.5	1.04	0.78
PC	Specificity	95.6	93.4–97.3	2.54	0.78
PC	AUC	0.82	-	-	-

Prevalence

A separate random-effects GLMM was used to estimate the pooled prevalence of intussusception. The model yielded a logit-scale estimate of 0.0228 (SE 0.2473; 95% CI −0.4619 to 0.5074), corresponding to a pooled prevalence of 50.6% (95% CI 38.8–62.5%) on the probability scale. Between-study heterogeneity was substantial (τ^2 = 2.325, I^2 = 98.2%), and the 95% prediction interval indicated that prevalence in a new study could plausibly range from 5.0% to 95.4% (Supplementary Table 4). This very high heterogeneity confirms that pooled prevalence is not directly generalized across settings but remains valuable for contextualizing predictive values.

Using the pooled HSROC sensitivity (96.3%) and specificity (95.6%), we derived positive and negative predictive values (PPV and NPV) across observed prevalence bands. Of the 40 studies reporting prevalence, 3 (7.3%) were in the low-prevalence band (<10%), 9 (22.0%) in the medium band (10–30%), and 29 (70.7%) in the high band (>30%). PPV increased with rising prevalence, from 54.1% at 5% prevalence to 99.8% at 95% prevalence, while NPV decreased from 99.8% to 57.7% across the same range (Fig. 5). These findings highlight the strong dependence of predictive values on baseline risk and provide a framework for interpreting ultrasound results in low-, medium-, and high-prevalence emergency settings.

Subgroup analysis

We explored potential sources of variability in diagnostic performance through subgroup analyses using Bayesian bivariate models for key symptoms and demographic factors. In these models, α (alpha) reflects the effect on sensitivity and β (beta) on specificity, with back-transformed outputs expressed as percentages for clarity.

For individual symptoms, the effects were minimal. For abdominal pain, α = −0.9% (95% CrI −4.3% to 2.3%) and β = 0.6% (95% CrI −4.1% to 5.1%), indicating negligible

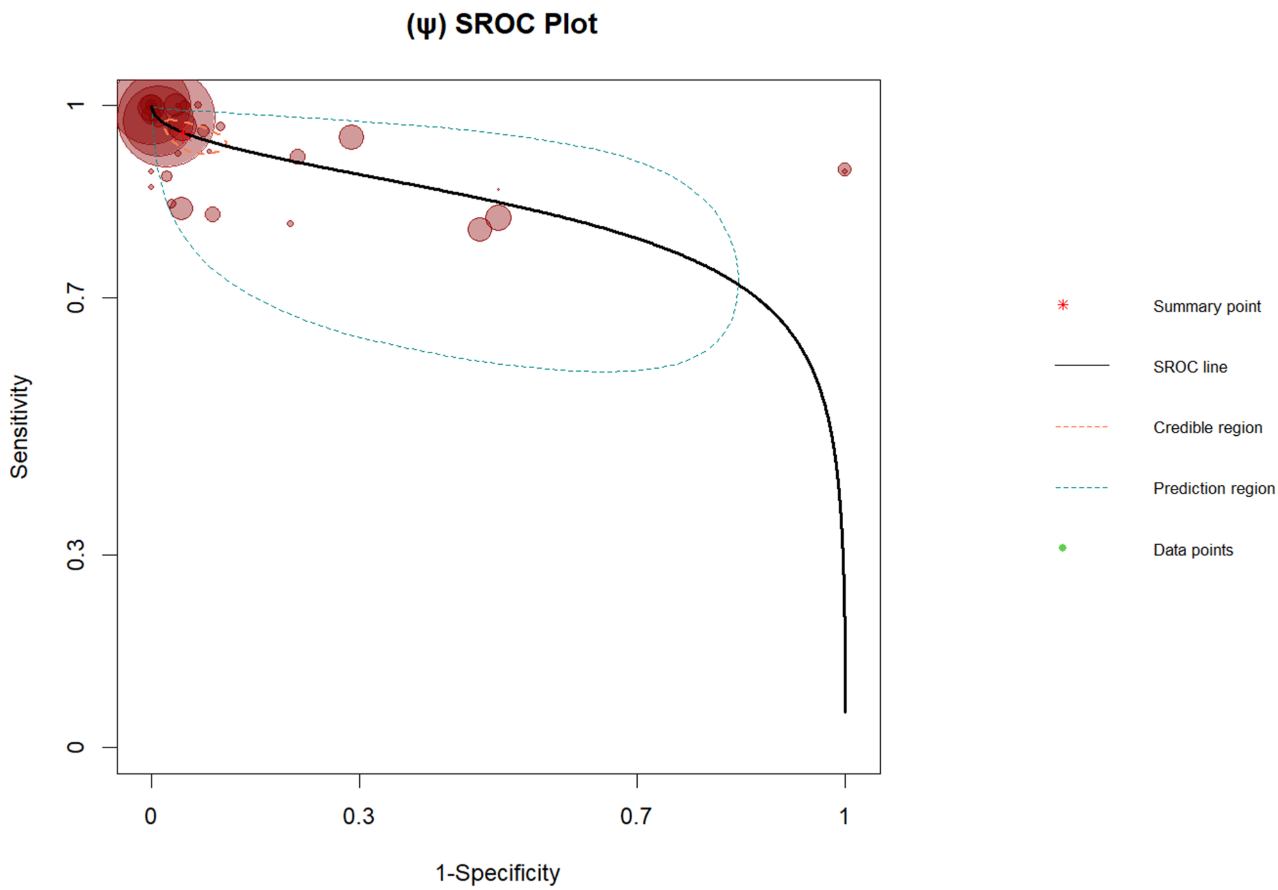

Fig. 4 Hierarchical summary ROC (HSROC) curve with 95% confidence (inner ellipse) and 95% prediction regions (outer ellipse), summarizing the trade-off between sensitivity and specificity across studies

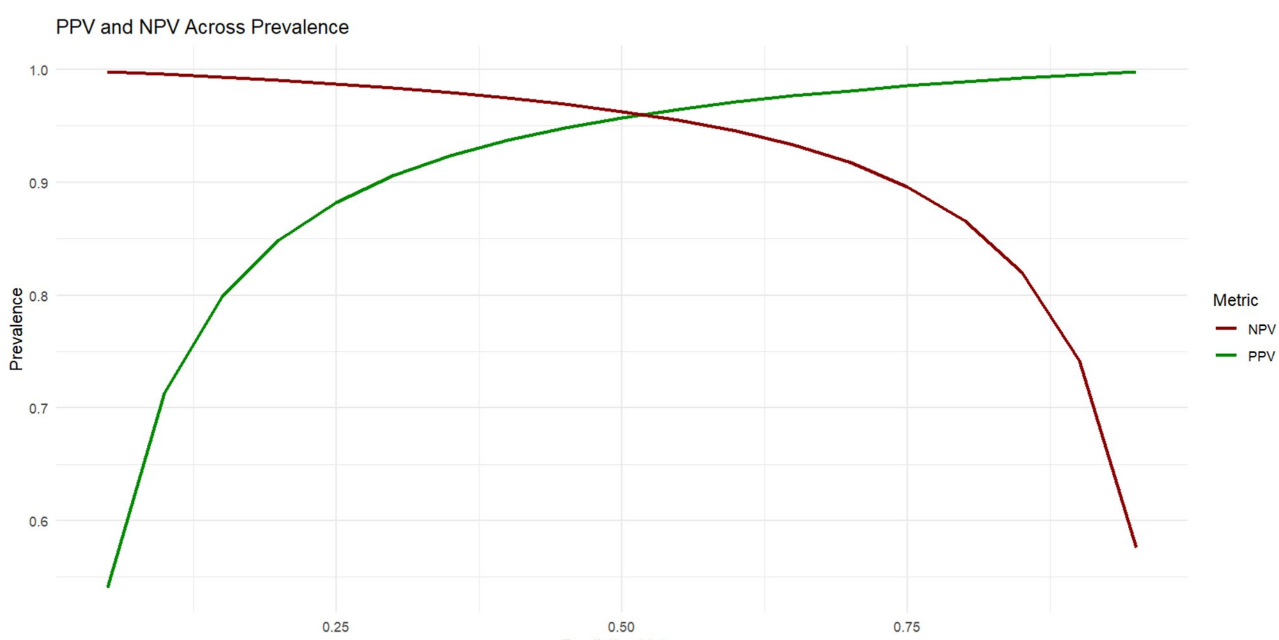

Fig. 5 Fagan nomogram illustrating post-test probability of intussusception at pre-test probabilities of 10%, 30%, and 50%, based on pooled sensitivity and specificity

impact on sensitivity or specificity. Fever showed α = −7.5% (−21.4% to 5.5%) and β = −3.3% (−14.3% to 7.3%), while vomiting had α = 4.2% (−2.8% to 3.7%) and β = 13.6% (−37.0% to 75.2%). Other symptoms, including tender abdomen, bloody stools, constipation, abdominal mass, and lethargy, similarly demonstrated small, statistically non-significant changes in sensitivity and specificity.

For demographic covariates, age produced α = 1.3% (−3.3% to 5.7%) and β = 5.9% (−2.3% to 13.7%), while male proportion showed α = 2.6% (−2.9% to 8.3%) and β = 0.6% (−10.1% to 11.6%). These results suggest that neither age nor sex composition substantially altered test performance across studies.

A separate subgroup analysis examined diagnostic accuracy by operator specialty, comparing radiologist-performed studies to the overall pooled dataset. The overall model yielded a pooled sensitivity of approximately 97.5% (95% CrI: 94–99%) and specificity of 98.5% (95% CrI: 95–99.6%), with moderate heterogeneity (τ^2_{se} = 1.15; τ^2_{sp} = 2.93) and strong positive correlation (ρ = 0.76). In contrast, radiologist-only studies demonstrated a sensitivity of about 65% (95% CrI: <1–99%) and specificity of 65% (95% CrI: <1–99%), accompanied by wider credible intervals and greater between-study variability (τ^2_{se} = 1.14; τ^2_{sp} = 4.48). The broad uncertainty intervals and small number of radiologist-only studies suggest that these apparent differences may reflect sampling variability rather than a true disparity in performance. Taken together, the results indicate no conclusive evidence that operator background substantially influences diagnostic accuracy, although precision was lower in the radiologist-only subgroup (Table 2).

Between-study heterogeneity remained considerable, with variance components for sensitivity (var_phi) ranging from 0.37 to 2.88 and for specificity (var_psi) from 1.74 to 5.93, and correlations between sensitivity and specificity ranging from −0.20 to 0.79. Collectively, these subgroup analyses indicate that the observed variation in diagnostic accuracy is not strongly driven by individual symptoms or demographic factors, supporting the robustness of our pooled sensitivity, specificity, and prevalence-stratified predictive value estimates presented in earlier sections.

Publication bias

Potential publication bias was evaluated using funnel plots for sensitivity and specificity. Visual inspection suggested some asymmetry, with a subset of smaller studies lying outside the expected distribution, indicating a moderate likelihood of publication bias. However, the effect on the pooled estimates appears minimal. Begg's rank correlation test indicated no statistically significant small-study effects (z = 1.40, $p = 0.162$; bias estimate = 53.0, SE = 37.86), suggesting that any publication bias is likely minimal. Both Bayesian models (H-Cauchy and PC priors) produced robust sensitivity and specificity estimates, and the narrow credible intervals reinforce the precision of the pooled results despite the minor asymmetry observed (Supplementary Fig. 4).

Certainty of evidence (GRADE Assessment)

The certainty of evidence was appraised using the GRADE framework across pooled accuracy, predictive values, prevalence, and subgroup analyses (Table 3). Sensitivity and specificity were rated as high certainty, supported by consistent results, tight credible intervals, and robustness across priors and subgroups. Predictive values were also rated as high certainty; estimates were precise and clinically meaningful across low-, medium-, and high-prevalence scenarios.

In contrast, the certainty for prevalence was moderate due to substantial between-study heterogeneity (I^2 ≈ 98%), which limits generalizability of the pooled estimate. Subgroup analyses by symptoms, age, and sex were consistently rated as high certainty, as none materially altered diagnostic performance (Table 3).

Table 2 Summary table of exploratory covariates outcomes

Covariate	α (logit)	β (logit)	Sensitivity % (Δ)	Specificity % (Δ)	var_φ	var_ψ	ρ
Abdominal pain	-0.009 (-0.043, 0.023)	0.006 (-0.041, 0.051)	49.8 → 49.7%	50.1 → 50.1%	1.467	3.208	0.693
Fever	-7.493 (-21.407, 5.476)	-3.340 (-14.304, 7.271)	0.06 → 0.09%	3.3 → 4.8%	3.352	1.735	0.465
Vomiting	0.420 (-2.768, 3.694)	1.356 (-3.703, 7.522)	60.3 → 61.3%	79.8 → 79.9%	1.495	3.956	0.626
Tender abdomen	-4.536 (-9.343, -0.128)	-4.836 (-13.401, 4.232)	1.1 → 1.0%	1.3 → 1.2%	0.373	3.834	-0.198
Bloody stools	-0.989 (-3.188, 1.091)	-0.977 (-4.563, 3.442)	27.5 → 26.9%	27.3 → 27.1%	0.779	2.732	0.516
Constipation	-11.750 (-28.328, 8.159)	-21.288 (-56.005, 11.637)	~0 → 0%	~0 → 0%	1.823	6.593	0.093
Abdominal mass	-1.589 (-4.366, 0.763)	3.546 (-4.643, 14.697)	16.0 → 16.5%	47.0 → 48.0%	0.574	5.931	0.615
Lethargy	-6.529 (-11.367, -0.893)	-5.598 (-15.724, 1.952)	0.15 → 0.14%	0.35 → 0.33%	2.881	1.912	-0.006
Age (continuous)	0.013 (-0.033, 0.057)	0.059 (-0.023, 0.137)	~50 → 50.3%	~50 → 51.5%	1.326	3.916	0.767
Male proportion	2.591 (-2.927, 8.334)	0.577 (-10.053, 11.625)	55 → 92%	50 → 64%	1.295	4.528	0.787
Operator specialty (Radiologist)	1.206 (−26.53, 28.94)	1.218 (−26.52, 28.96)	65.0 → 65.5%	65.0 → 65.5%	1.139	4.477	0.449
Operator specialty (Overall)	−0.147 (−0.460, 0.178)	−0.393 (−0.852, 0.062)	97.5 → 97.1%	98.5 → 98.0%	1.148	2.927	0.764

Table 3 Grade assessment table

Outcome	**No. of Studies (*n*)**	**Pooled Estimate (95% CrI/CI)**	**Certainty of Evidence (GRADE)**	**Key Considerations / Comments**
Sensitivity	44 studies	96.3% (94.9–97.5)	●●●● High	Consistent across studies and subgroups; tight CrIs; robust across priors
Specificity	44 studies	95.6% (93.3–97.5)	●●●● High	Minimal variation between priors; estimates stable across models
Prevalence	40 studies	0.51% (logit scale −0.46–0.51)	●●●○ Moderate	Substantial heterogeneity (I^2 = 98.2%) limits generalizability
Positive Predictive Value (PPV)	44 studies	54.1–99.8% (at 5–95% prevalence)	●●●● High	PPV increases with prevalence; narrow CrIs from pooled Se/Sp
Negative Predictive Value (NPV)	44 studies	57.7–99.8% (at 5–95% prevalence)	●●●● High	NPV decreases with prevalence; precise estimates across scenarios
Subgroup: Symptoms (all)	3–14 (per symptom)	α: −7.5–4.2%; β: −21.3–13.6%	●●●● High	Individual symptoms did not materially affect Se/Sp
Subgroup: Age	35 studies	α = 1.3%; β = 5.9%	●●●● High	No clinically important effect of age on test performance
Subgroup: Male proportion	37 studies	α = 2.6%; β = 0.6%	●●●● High	Sex composition did not materially affect test performance
Outcome/Subgroup	No. of Studies (n)	Pooled Estimate (95% CrI/CI)	Certainty of Evidence (GRADE)	Key Considerations/Comments
Subgroup: Operator Specialty (Radiologist)	12 studies	Se = 65% (CrI < 1–99%), Sp = 65% (CrI < 1–99%)	●●●○ Moderate	Wide CrIs and higher between-study heterogeneity; performance consistent with overall estimates despite uncertainty
Subgroup: Operator Specialty (All Operators)	43 studies	Se = 97.5% (94–99%), Sp = 98.5% (95–99.6%)	●●●● High	Pooled diagnostic accuracy remained high across operator backgrounds; minimal influence of operator variation

Discussion

This meta-analysis provides one of the most comprehensive Bayesian syntheses to date on the diagnostic performance of ultrasound for detecting pediatric intussusception in emergency settings. Drawing on data from 44 studies encompassing 4,142 pediatric patients, the findings demonstrate that ultrasound possesses excellent diagnostic accuracy, with pooled sensitivity and specificity of 96.3% and 95.7%. The results remained remarkably stable across different prior distributions, subgroups, and operator backgrounds.

The pooled sensitivity of 96.3% and specificity of 95.7% found in this meta-analysis indicate that ultrasound is highly effective in distinguishing true cases of intussusception from non-cases. These findings reaffirm the established role of ultrasound as the diagnostic modality of choice for pediatric intussusception, aligning closely with previous meta-analyses conducted, which reported pooled sensitivities and specificities ranging between 93 and 98% and 88–96%, respectively [18, 19]. The high sensitivity ensures that few cases of intussusception are missed [20], while the high specificity minimizes unnecessary interventions, such as contrast enemas or surgical exploration [21]. Given that ultrasound is non-invasive, radiation-free, and widely accessible, these results strongly support its use as a first-line diagnostic test in both high- and low-resource emergency settings. Furthermore, the HSROC-AUC value of 0.82 indicate excellent overall discriminative ability, reflecting ultrasound's capacity to maintain diagnostic accuracy across variable thresholds and operator contexts.

The pooled prevalence of intussusception across studies 50.6%, though with very high between-study heterogeneity (I^2 = 98.2%). This variation likely reflects differences in study design, patient selection criteria, and clinical thresholds for imaging [22, 23]. Despite this variability, the derived predictive values provide practical insights into the real-world performance of ultrasound. The PPV rose from 54.1% at a 5% prevalence to 99.8% at a 95% prevalence, while the NPV decreased from 99.8% to 57.7% across low- and moderate-prevalence settings. This suggests that in high-prevalence settings, a positive ultrasound nearly confirms the diagnosis, while in low-prevalence contexts, a negative ultrasound effectively rules it out. Thus, in tertiary pediatric emergency departments, where clinical suspicion for intussusception is typically high, the combination of suggestive symptoms and a positive ultrasound result can confidently guide enema reduction or surgical management [24]. Conversely, in general emergency departments or primary centers with lower prevalence, a negative ultrasound may suffice to exclude intussusception without the need for further imaging [25].

Subgroup analyses revealed that neither demographic factors nor clinical symptoms significantly influenced the diagnostic accuracy of ultrasound. Variations by age, sex distribution, or presenting features such as abdominal pain, vomiting, or bloody stools produced only minor and statistically non-significant changes in sensitivity and specificity. This suggests that ultrasound maintains high diagnostic reliability regardless of symptom profile or patient demographics, which is especially valuable in pediatric populations where symptom presentation can be variable and nonspecific [26].

Of particular note is the operator specialty analysis. While radiologist-performed studies showed wide credible intervals due to small sample sizes and higher between-study variance, the overall diagnostic performance remained excellent across operator backgrounds. Studies including both radiologists and trained emergency physicians demonstrated pooled sensitivities and specificities exceeding 97%, suggesting that operator background alone does not substantially influence diagnostic performance when adequate training and supervision are ensured [27, 28]. These findings align with recent evidence showing that point-of-care ultrasound (POCUS) conducted by trained pediatric emergency physicians achieves diagnostic accuracies comparable to formal radiology-performed examinations [29, 30]. This supports the expanding role of POCUS in resource-limited settings and highlights the value of structured ultrasound training in emergency medicine curricula.

A key strength of this meta-analysis lies in its comprehensive scope and methodological rigor. The inclusion of 44 studies encompassing diverse clinical settings enhances generalizability, while the use of a Bayesian hierarchical framework provides a robust approach to modeling diagnostic accuracy under uncertainty. However, several limitations should be acknowledged. First, despite the low risk of bias across most studies, heterogeneity in prevalence and reference standards could influence pooled estimates. Although all studies used confirmatory imaging or surgical findings as reference standards, differences in diagnostic thresholds or operator experience may have contributed to variability in specificity. Second, the subgroup analyses, particularly for operator specialty and rare symptoms, were limited by smaller sample sizes and wide credible intervals, precluding definitive conclusions in these strata. Third, leave-one-out diagnostics were unavailable in the Bayesian framework employed, restricting the ability to identify highly influential studies.

Future research should focus on several key areas. First, well-designed multicenter prospective studies are needed to evaluate ultrasound performance across diverse healthcare systems, especially in low- and middle-income countries where operator training and access to imaging resources vary widely. Second, future meta-analyses could incorporate individual patient data (IPD) to enable patient-level covariate modeling, allowing exploration of interaction effects between age, duration of symptoms, and diagnostic accuracy. Third, standardization of ultrasound training protocols and diagnostic criteria is crucial to minimize variability across operators and settings. Finally, integration of artificial intelligence (AI) and machine learning–based image recognition algorithms into ultrasound workflows offers a promising avenue to augment diagnostic accuracy and reproducibility. Early studies have shown that AI-assisted ultrasound interpretation can achieve diagnostic accuracies comparable to expert radiologists, suggesting significant potential for enhancing diagnostic capacity in resource-limited environments [31, 32].

Conclusion

This meta-analysis of 44 studies involving 4,142 children found that ultrasound has a pooled sensitivity of 96.3% (95% CrI 94.9–97.5) and specificity of 95.6% (95% CrI 93.3–97.5) for diagnosing pediatric intussusception, with an AUC of 0.81–0.82, indicating excellent diagnostic accuracy. Predictive values ranged from PPV 54.1–99.8% and NPV 99.8–57.7% across prevalence bands. These results confirm ultrasound as a highly accurate, rapid, and non-invasive diagnostic tool for pediatric intussusception. Future studies should focus on multicenter validation and AI-assisted interpretation to enhance diagnostic consistency across settings.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12245-026-01134-z.

Supplementary Material 1

Acknowledgements

The authors would like to thank all contributors and institutions that supported this project.

Author contributions

Mohammed Alsabri (MA): First and corresponding author. Conceived and designed the study, developed the protocol, coordinated all stages of manuscript development, drafted substantial portions of the manuscript, supervised the research team, and managed all communications with co-authors.Shree Rath (SR): Joint first author. Led literature screening, PROSPERO registration, creation of the data-extraction sheet, and finalization of included studies. Contributed to study concept development, data extraction, drafting of the discussion, and full manuscript review. Responsible for validation, search strategy, extraction, quality assessment, discussion synthesis, and final compilation.Mohamed Amr Elkarargy (MAE): Conducted the critical analysis and contributed significantly to the Results section.Amira A. Aboali (AAA): Participated in title and abstract screening with Dr. Ramadan, completed full-text screening, and selected eligible studies. Prepared the data-extraction sheet, reviewed extracted data and quality assessment, and contributed to the Introduction and Methods sections.Khaled Abouelmagd (KA): Contributed to data extraction and verification. Abdelaziz Abdelaziz Abdelftah Ramadan (AAR): Participated in screening and selection of included studies.Luis L. Gamboa (LLG): Provided expert supervision, conceptual feedback, and critical revision of the manuscript for important intellectual content.Patrick Yoo (PY): Reviewed, edited, and approved the final manuscript. Yisha Cheng (YC): Contributed to manuscript writing, critical review, and approval of the final version.All authors read and approved the final manuscript.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Pediatrics, Al-thawara Modern General Hospital, Sana'a, Yemen
[2]St. Christopher's Hospital for Children, Philadelphia, PA, USA
[3]Department of Medicine, All India Institute of Medical Sciences, Bhubaneswar, India
[4]University College Cork, Cork, Ireland
[5]Department of Diagnostic and Interventional Radiology, Damanhour Teaching Hospital, General Organization for Teaching Hospitals and Institutes, Damanhour, Egypt
[6]Cardiology Department, Faculty of Medicine, Al-Azhar University, New Damietta, Egypt
[7]Faculty of Medicine, Tanta University, Tanta, Egypt
[8]Drexel University College of Medicine, Pediatric Emergency Medicine Fellowship, St Christopher's Hospital for Children, Philadelphia, PA, USA
[9]Pediatric Emergency Department, Department of Pediatrics, St. Christopher's Hospital for Children, Philadelphia, PA, USA

Received: 20 December 2025 / Accepted: 4 February 2026
Published online: 10 February 2026

References
1. Loukas M, Pellerin M, Kimball Z, de la Garza-Jordan J, Tubbs RS, Jordan R. Intussusception: an anatomical perspective with review of the literature. Clin Anat. 2011;24:552–61.
2. Edwards EA, Pigg N, Courtier J, Zapala MA, MacKenzie JD, Phelps AS. Intussusception: past, present and future. Pediatr Radiol. 2017;47:1101–8.
3. Applegate KE. Intussusception in children: evidence-based diagnosis and treatment. Pediatr Radiol. 2009;39(Suppl 2):140–3.
4. Ko HS, Schenk JP, Troger J, Rohrschneider WK. Current radiological management of intussusception in children. Eur Radiol. 2007;17:2411–21.
5. Tate JE, Simonsen L, Viboud C, Steiner C, Patel MM, Curns AT, et al. Trends in intussusception hospitalizations among US infants, 1993–2004: implications for monitoring the safety of the new rotavirus vaccination program. Pediatrics. 2008;121:e1125–32.
6. Buettcher M, Baer G, Bonhoeffer J, Schaad UB, Heininger U. Three-year surveillance of intussusception in children in Switzerland. Pediatrics. 2007;120:473–80.
7. Yang G, Wang X, Jiang W, Ma J, Zhao J, Liu W. Postoperative intussusceptions in children and infants: a systematic review. Pediatr Surg Int. 2013;29:1273–9.
8. Jiang J, Jiang B, Parashar U, Nguyen T, Bines J, Patel MM. Child intussusception epidemiology: A systematic review. BMC Gastroenterol. 2021;21(1):1–12.
9. Johnson B, Gargiullo P, Murphy TV, Parashar UD, Patel MM. Factors associated with bowel resection among infants with intussusception in the united States. Pediatr Emerg Care. 2012;28(6):529–32. https://doi.org/10.1097/PEC.0b013e3182587d12.
10. Daneman A, Navarro O. Intussusception. Part 1: A review of diagnostic approaches. Pediatr Radiol. 2004;34(2):95–108.
11. Hryhorczuk AL, Strouse PJ. Validation of US as a first-line diagnostic test for assessment of pediatric Ileocolic intussusception. Pediatr Radiol. 2009;39(10):1075–9.
12. 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):255–8.
13. Pari Gul, Mansoor MA, et al. The diagnostic accuracy of sonography in children suspecting intussusception keeping surgical findings as gold standard. Pak J Radiol. 2020;30(1):31–34. Available from: [https://www.pakjr.com/ojs/index.php/PJR/article/view/285] (https://www.pakjr.com/ojs/index.php/PJR/article/view/285).
14. Li XZ, Wang H, Song J, Liu Y, Lin YQ, Sun ZX. Ultrasonographic diagnosis of intussusception in children: A systematic review and Meta-Analysis. J Ultrasound Med. 2021;40(6):1077–84.
15. Rahmani E, Amani-Beni R, Hekmatnia Y, et al. Diagnostic accuracy of ultrasonography for detection of intussusception in Children; a systematic review and Meta-Analysis. Arch Acad Emerg Med. 2023;11(1):e24.
16. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and Meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA. 2018;319(4):388–96. https://doi.org/10.1001/jama.2017.19163.
17. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
18. Rahmani E, Amani-Beni R, Hekmatnia Y, et al. Diagnostic accuracy of ultrasonography for detection of intussusception in Children; A systematic review and Meta-Analysis. PubMed. 2023;11(1):e24. https://doi.org/10.22037/aaem.v11i1.1914.
19. Lin-Martore M, Firnberg MT, Kohn MA, Kornblith AE, Gottlieb M. Diagnostic accuracy of point-of-care ultrasonography for intussusception in children: A systematic review and meta-analysis. Am J Emerg Med. 2022;58:255–64. https://doi.org/10.1016/j.ajem.2022.06.025.
20. UpToDate. UpToDate. https://www.uptodate.com/contents/intussusception-in-children
21. Li X, Wang H, Song J, Liu Y, Lin Y, Sun Z. Ultrasonographic diagnosis of intussusception in children. J Ultrasound Med. 2020;40(6):1077–84. https://doi.org/10.1002/jum.15504.
22. Li Y, Zhou Q, Liu C, et al. Epidemiology, clinical characteristics, and treatment of children with acute intussusception: a case series. BMC Pediatr. 2023;23(1). https://doi.org/10.1186/s12887-023-03961-y.
23. Esteghamati A, Joulani M, Sayyahfar S, et al. Incidence of intussusception in children less than five years of age: a pre-rotavirus vaccine survey from Iran, 2010–2015. PubMed. 2020;34:40. https://doi.org/10.34171/mjiri.34.40.
24. Hwang J, Yoon HM, Kim PH, Jung AY, Lee JS, Cho YA. Current diagnosis and image-guided reduction for intussusception in children. Clin Experimental Pediatr. 2022;66(1):12–21. https://doi.org/10.3345/cep.2021.01816.
25. Lin-Martore M, Kornblith A, Kohn M, Gottlieb M. Diagnostic accuracy of Point-of-Care ultrasound for intussusception in children presenting to the emergency department: a systematic review and meta-analysis. Western J Emerg Med. 2020;21(4). https://doi.org/10.5811/westjem.2020.4.46241.
26. Ghani R, O'Connor A, Sajid I, Johnson G, Ullah S. Diagnostic accuracy of ultrasound in the paediatric population with acute right Iliac fossa pain, our district general hospital experience. Published Febr 11, 2022. https://pmc.ncbi.nlm.nih.gov/articles/PMC8835414/.
27. Vasey B, Ursprung S, Beddoe B, et al. Association of clinician diagnostic performance with machine Learning–Based decision support systems. JAMA Netw Open. 2021;4(3):e211276. https://doi.org/10.1001/jamanetworkopen.2021.1276.
28. Ceriani E, Schiavon R, La Cava L, Ruscitti C, Cogliati C. Point of care ultrasound: focus on evidence for a critical appraisal. European J Intern Medicine Published Online June. 2025;1. https://doi.org/10.1016/j.ejim.2025.06.005.
29. Katz-Dana H, Stackievicz R, Dana E, et al. Diagnostic accuracy of point-of-care ultrasound (PoCUS) for the diagnosis of hip effusion in the pediatric emergency department. Canadian J Emerg Medicine Published Online Oct. 2024;1. https://doi.org/10.1007/s43678-024-00788-z.
30. Gold DL, Marin JR, Haritos D, et al. Pediatric emergency medicine physicians' use of point-of-care ultrasound and barriers to implementation: a regional pilot study. AEM Educ Train. 2017;1(4):325–33. https://doi.org/10.1002/aet2.10049.

31. Khalifa M, Albadawy M. AI in diagnostic imaging: revolutionising accuracy and efficiency. Comput Methods Programs Biomed Update. 2024;5:100146. https://doi.org/10.1016/j.cmpbup.2024.100146.
32. Xu D, Sui L, Zhang C, et al. The clinical value of artificial intelligence in assisting junior radiologists in thyroid ultrasound: a multicenter prospective study from real clinical practice. BMC Med. 2024;22(1). https://doi.org/10.1186/s12916-024-03510-z.

Publisher's note