BMC Medical Imaging

https://doi.org/10.1186/s12880-025-02114-0

Article in Press

Accuracy of junior doctor plain trauma X-ray interpretation: a systematic review and meta-analysis

Received: 30 September 2025

Accepted: 8 December 2025

Published online: 17 December 2025

Cite this article as: Acquah G., Anyitey-Kokor I.C., Donkor A. *et al.* Accuracy of junior doctor plain trauma X-ray interpretation: a systematic review and meta-analysis. *BMC Med Imaging* (2025). https://doi.org/10.1186/s12880-025-02114-0

Godwill Acquah, Ijeoma Chinedum Anyitey-Kokor, Andrew Donkor, Yaw Amo Wiafe, Benard Ohene-Botwe, Michael J. Neep & Patrick C. Brennan

Title

Accuracy of junior doctor plain trauma X-ray interpretation: A systematic review and meta-analysis

Author Names

Godwill Acquah [1, 2], Ijeoma Chinedum Anyitey-Kokor [1, 3], Andrew Donkor [1, 4], Yaw Amo Wiafe [1], Benard Ohene-Botwe [5], Michael J Neep [6, 7], Patrick C Brennan [8].

Affiliations

[1] = Department of Medical Imaging, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana.

[2]= Department of Radiography, University of Ghana, P.O. Box KB 143, Korle Bu, Accra, Ghana

[3] = Komfo Anokye Teaching Hospital, Kumasi, Ghana.

[4] = IMPACCT (Improving Palliative Aged and Chronic care through Clinical Research and Translation), Faculty of Health, University of Technology Sydney, Sydney, New South Wales, Australia.

[5] = Division of Radiography, Department of Allied Health, City St George's, University of London, UK.

[6] = Department of Medical Imaging, Logan Hospital, Meadowbrook, Queensland, Australia.

[7] = School of Clinical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia.

[8] = Discipline of Medical Imaging Science, University of Sydney, Camperdown, NSW, Australia.

Corresponding Author

Godwill Acquah

Email: goacquah@ug.edu.gh

Tel: +233543980925

Abstract

Background: Plain radiography remains a first-line assessment tool for emergency departments' trauma patients. Given the urgency of trauma care, emergency department doctors, including junior doctors, often perform initial trauma X-ray interpretations to support timely patient management when there is an unavailability of an immediate radiologist report. However, trauma X-ray interpretation is challenging, and inaccuracies can impact patient care. This study evaluates the diagnostic accuracy of emergency department junior doctors on the initial interpretation of trauma X-rays by systematically reviewing and meta-analysing existing research on the subject.

Method: Studies were identified from PubMed, Scopus, Embase, Cochrane Library, and by checking the reference lists of relevant studies. Quality assessment of included studies was evaluated using the QUADAS-2 tool. Meta-analysis was conducted using bivariate models, with summary estimates reported as sensitivity, specificity, and the SROC. Meta-regression and subgroup analysis was performed to evaluate the sources of heterogeneity. Publication bias was assessed using Deeks' funnel plot.

Results: Seven studies were included in this meta-analysis. Across the studies, pooled sensitivity, specificity and area under the receiver operating characteristic curve (AUC) were 0.65 (95% CI= 0.47 – 0.80), 0.89 (95% CI= 0.77 – 0.95), and 0.86 (95% CI= 0.83 – 0.89), respectively. Covariate analysis per anatomical region of trauma X-rays showed that for skeletal region, pooled sensitivity, specificity, and AUC were 0.72 (95% CI= 0.55 – 0.85), 0.86 (95% CI= 0.73 – 0.93), and 0.87 (95% CI= 0.83 – 0.89), respectively and for appendicular region, pooled sensitivity, specificity and AUC were 0.68 (95% CI= 0.49 – 0.82), 0.82 (95% CI= 0.62 – 0.93), and 0.81 (95% CI= 0.77 – 0.84), respectively. Substantial heterogeneity was identified but was not due to a threshold effect (Spearman rho= 0.29(p=0.49)). Meta-regression and subgroup analysis revealed that anatomical-region-specific trauma X-ray interpretation and accuracy assessment techniques influenced heterogeneity. No publication bias was identified (p=0.41).

Conclusion: Emergency department junior doctors' accuracy in the initial interpretation of trauma X-rays was moderate. The findings further suggest a high likelihood of missed abnormalities when they interpret trauma X-rays. This highlights the need for support strategies to enhance their diagnostic accuracy to strengthen clinical decision-making in trauma care.

Clinical trial number: Not applicable

Keywords: Traumatic injury, X-ray interpretation, Junior doctors, Diagnostic accuracy, meta-analysis, Emergency department.

Background

Traumatic injuries remain a major global health concern, with musculoskeletal trauma accounting for a substantial proportion of emergency department (ED) presentations.[1,2,3,4,5,6,7,8,9,10,11,12,13] Plain radiography is the first-line imaging modality for assessing these injuries due to its wide availability, cost-effectiveness, and efficiency in detecting fractures and other related abnormalities.[14,15,16,17] In most settings, radiologists traditionally interpret ED trauma X-rays (TXRs) to inform patient management. However, workforce shortages, increasing service demands, and the need for timely decision-making have shifted greater responsibility to ED clinicians, including junior doctors (JDs), as well as reporting radiographers.[15,18,19,20,21,22,23]

In many EDs across both high-income and low-resource settings, emergency department JDs routinely interpret TXRs to support their immediate clinical decisions, often before formal radiology reports are available.[15,18,19,23] Their initial interpretations directly affect patient triage, treatment, referral pathways, and decisions to discharge patients. [18,23]

Despite the advantage of JD's initial interpretation, the interpretation of TXRs is challenging, and errors can be of significant detriment to both patients and clinicians. This is intensified by most EDs being fraught with high patient turnover, time pressures, complex working environment and JDs' relative inexperience. [24,25,26,27,28,29]

For decades, existing research has shown variable levels of performance of ED JDs in the initial interpretation of TXRs, with inconsistent reporting estimates.[23,28,30,31,32] There have so far been limited attempts to produce a comprehensive synthesis of this evidence. The aim of this study was to conduct a systematic review and meta-analysis of the existing literature to establish the sensitivity and specificity of ED JDs' initial interpretation of TXRs.

Methods

Junior doctor definition for the review

The term "junior doctor" varies internationally and for over half a century, has encompassed a wide range of early-career medical practitioners including casualty officers, house officers, foundation year doctors, intern doctors, medical officers, senior house officers, residents (specialty trainee), registrars (specialty trainees), and in some jurisdictions resident doctors (both specialty trainees and non-specialty doctors).[22,24,33,34,35,36,37,38,39,40,41] Despite these variations, which sometimes result in terminological overlap, it is generally understood that they are fully qualified doctors within the first ten years after graduation. During this period, they may be undergoing postgraduate training (specialty training), employed in a non-training post or gaining clinical experience in various hospital departments under varying levels of supervision (postgraduate medical internship). The purpose of this stage is to build the knowledge and skills necessary to progress toward specialist or general practice roles.[23,39,40,41,42]

For this review, however, *JDs refer to fully qualified early-career doctors working in EDs who are not enrolled in formal residency (specialty) training programmes. This includes doctors employed in non-training posts, those undertaking postgraduate medical internships, and early-career medical officers rotating through emergency departments. They include house officers, foundation doctors, casualty officers, non-training senior house officers and medical officers.*

Protocol registration and reporting guidelines

The study was registered in the PROSPERO (CRD420251044917) and was guided by the step-by-step process outlined by Leeflang[43] and Deeks et al.[44] when conducting a diagnostic-test-accuracy systematic review and meta-analysis. Additionally, the Preferred Reporting Items for Systematic Reviews and Meta-Analysis of Diagnostic Test Accuracy (PRISMA-DTA) guidelines[45] were implemented to structure and summarise the findings from the literature search.

PICO question

To guide the focus of this review and ensure a structured approach to evaluating the accuracy of JDs, a clearly defined clinical question was formulated. This question specifies the patient group (population/patient), the diagnostic task being assessed (intervention), the comparison standard (comparison), and the outcomes of interest (outcome). It provided a framework for determining study eligibility (Table 1), directing data extraction, and shaping the synthesis of evidence. The question was:

In patients presenting with traumatic injury who undergo plain radiography (P), how accurately do emergency department junior doctors initially interpret their trauma X-rays (I) compared with radiologists or advanced practice radiographers (C) in identifying traumatic abnormalities (O)?

Database and search strategy

A systematic search strategy was developed and refined to comprehensively identify eligible studies across four databases: PubMed, SCOPUS, Cochrane Library, and EMBASE, to retrieve relevant articles. This approach aligns with Leeflang's [43] and Deeks et al.'s [44] recommendations, which emphasise searching at least two databases to enhance the comprehensiveness and reliability of the review. Search terms included ("Junior doctor" OR "Foundation doctor" OR "Intern" OR "Medical officer" OR "Casualty officer" OR "House officer" OR "Doctor" OR "Houseman") AND ("radiographic interpretation" OR "image interpretation" OR "X-ray interpretation" OR "misinterpretation") AND ("trauma OR fracture"). Truncation and wildcards (*,?) were utilised to broaden retrieval and capture keyword variations. The literature search was conducted between 15th January 2025 and 15th April 2025 and updated on 15th August 2025 for any new publications. To prevent omissions, a hand search was conducted in the reference lists of relevant papers for eligible literature on the subject.

Inclusion and exclusion criteria

This review included studies published from 1990 to 15th August 2025, on TXR interpretation by JDs. There was no language restriction during the search. The year 1990 was chosen to include older, yet clinically relevant studies that reflect more contemporary trauma imaging practices and JD involvement. Eligible studies met the following criteria:

(1) reported on the accuracy of JDs' interpretation of TXRs

(2) provided complete data or sufficient information to construct 2 x 2 contingency tables

(3) used radiologist interpretation as the reference standard for accuracy. Studies that employed advanced practice radiographers as a reference standard were included based on jurisdictional certification and enough information to suggest their accuracy was equivalent to that of a radiologist.

Reports on TXR interpretation of other healthcare professionals (e.g., nurses, radiographers) or medical staff (e.g., residents (specialty trainees), registrars, senior doctors, and consultants) were excluded. Studies on JDs' plain X-ray interpretation of non-trauma cases, as well as interpretations of radiographs from other modalities like computed tomography (CT), and magnetic resonance imaging (MRI) were excluded. Studies that did not provide enough information to construct 2 x 2 contingency tables were excluded. Additionally, the population, intervention, comparator, and outcome (PICO) framework, which further framed the exclusion and inclusion criteria, has been summarised in Table 1.

Table 1. Eligibility criteria from PICO

PICO	Inclusion	Exclusion
Population/Patient	Patients who underwent plain radiography with a history of traumatic injury	Patients who underwent plain radiography for reasons other than traumatic injury or traumatic injury alone Patients who underwent examinations involving modalities like CT, MRI, and ultrasound, regardless of their clinical history
Intervention	Interpretation of TXRs by ED JDs	Interpretation of TXRs by other professionals (e.g., nurses, radiographers, residents (specialty trainees, senior doctors/registrars, and consultants).
Comparison	Interpretation of TXRs by radiologist/advanced practice radiographer	Interpretation of TXRs by other professionals (consultant emergency physicians and consultant orthopaedic physicians)
Outcome	2x2 table values (true positive, true negative, false positive, false negative), sensitivity, and specificity	Inadequate information to construct a 2x2 contingency table

NB: *MRI= Magnetic Resonance Imaging CT= Computed tomography JD= Junior doctor ED= Emergency department TXR= Trauma X-rays*

Selection strategy

Retrieved articles were imported into EndNote 21.0, where duplicates were identified and removed. Two reviewers (GA and ICA-K) independently read titles and abstracts to assess relevance. Full-text articles identified as potentially eligible were then reviewed in detail against predefined inclusion criteria by the same reviewers. Studies not meeting the criteria were excluded, with reasons documented. Any disagreements during the selection process were resolved through collaborative discussion.

Data extraction and quality assessment

Two authors (GA and ICA-K) independently extracted relevant data from selected articles. The extraction of data was in two parts.

(1) general characteristics of included studies in our study's context: study reference, country, study design, professional, clinical setting, assessment technique, and reference standard

(1) Characteristics of JDs' TXR interpretation: anatomical region of trauma X-rays, number of plain TXRs (total, normal and abnormal), number of interpretations per plain TXR, data (true positive (TP), false positive (FP), true negative (TN), false negative (FN)) and results (sensitivity and specificity).

The following considerations were applied during data extraction to ensure consistency and accuracy:

(1) In studies where the JDs interpreted the same TXRs using both film and digital workstations, the performance results from the variant yielding the highest diagnostic accuracy were selected for inclusion in the meta-analysis.

(2) For studies that reported both pre- and post-intervention assessments of JDs' diagnostic accuracy, only the pre-intervention (baseline) results were included in the analysis.

(3) In studies where the initial interpretation accuracy of JDs on TXRs was categorised by anatomical region, and each category met the inclusion criteria, data were extracted and analysed separately for each region as individual study parts.

(4) Authors of relevant articles were contacted for clarity on information where necessary.

(5) In cases where values for a 2 x 2 table (TP, FP, TN, and FN) were not directly reported, and authors who were contacted could not provide the raw data, these values were mathematically calculated from available data, i.e. total sample size, sensitivity, specificity, and prevalence.[16,4647]

The methodological quality of included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool,[48] as recommended by Deeks et al.[44]. GA and ICA-K independently conducted the assessment. For consistency, details of how the plain TXRs were sampled by each relevant study were used to judge whether the patient spectrum in each study was selected consecutively or randomly.[49] All disagreements during extraction and quality assessment were resolved through collaborative discussion.

Statistical analysis and data synthesis

A bivariate random-effects meta-analysis was conducted using STATA v14.0 (StataCorp), specifically the *Midas* command package.[50] A coupled forest plot was used to display the pooled estimate of sensitivity and specificity. A Summary

Receiver Operating Characteristic (SROC) curve was generated to illustrate the diagnostic performance of JDs. The interpretation of resultant area under the Receiver Operating Characteristic curve (AUC) was low ($0.5 \geq AUC \leq 0.7$), moderate ($0.7 \geq AUC \leq 0.9$), or high ($0.9 \geq AUC \leq 1$) accuracy. [50,51]

To assess heterogeneity, the Higgins I^2 statistic was calculated.[52] An $I^2 \leq 50\%$ indicated low heterogeneity, while $I^2 > 50\%$ suggested high heterogeneity.[52] Where high heterogeneity was observed, Spearman's correlation was used to evaluate heterogeneity due to threshold effects (values > 0.6 indicating significance).[52] If a threshold effect was not present, univariable meta-regression and subgroup analysis was conducted to explore alternative sources of heterogeneity.[50] Deeks' funnel plot was used to assess publication bias, with $p < 0.05$ indicating statistical significance.

Results

Literature search

Figure 1 presents the flow of the literature screening process. Overall, 1541 articles were retrieved from the comprehensive search, of which 63 duplicates were excluded. One thousand four hundred and fifty-four articles were then excluded after reading the titles and abstracts. Following this, the full text of 44 articles was sought, of which one was not retrieved. The full texts of the 43 articles remaining were read thoroughly, and 7 studies were ultimately included in this meta-analysis.[15,18,53,54,55,56,57]

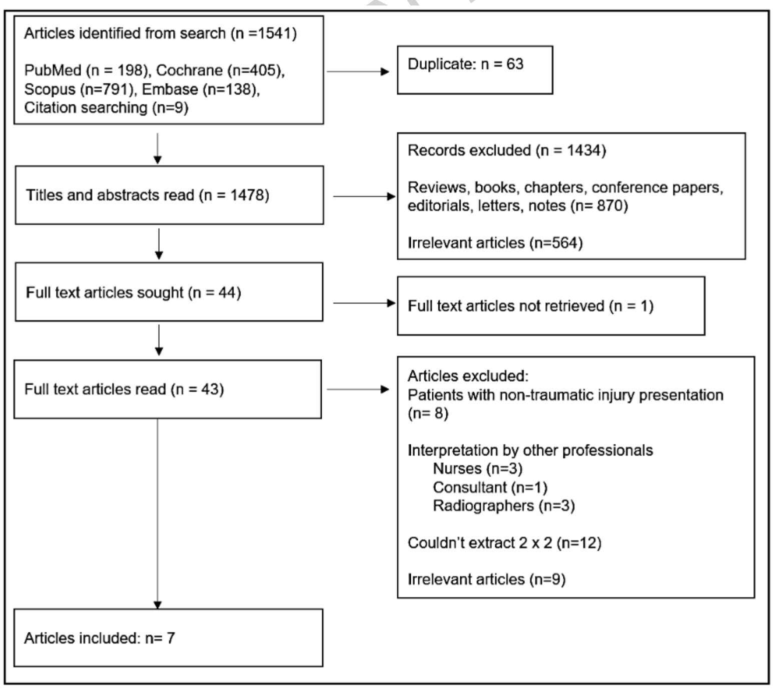

Figure 1. Flow chart of study selection process. *NB: Full details of reasons for full-text exclusion have been presented in the supplementary file*.

Study characteristics

From Table 2, the included studies were conducted between 1995 and 2023. Geographically, four studies were conducted in the UK,[18,53,54,55] and three were conducted in Africa.[15,56,57] Generally, the studies assessed the plain TXR interpretation accuracy among JDs. Regarding study design, four studies were comparative,[54,55,56,57] one study was retrospective and comparative,[53] one study was retrospective[15] and one study was prospective.[18] The comparative studies compared accuracy between JDs and other professionals, including nurses,[53,55] radiographers,[55,56,57] and radiologist specialist registrars[54] against a reference standard. All studies were conducted at the ED. In the assessment of JDs TXR interpretation accuracy, four studies employed a test bank approach,[54,55,56,57] two studies employed a retrospective analysis approach,[15,53] and one study employed a prospective analysis approach.[18] Across the studies, and in each case, the X-rays interpreted by JDs were obtained from multiple anatomical sites of patients' injury. While radiologist interpretation was the primary reference standard in most included studies, one study in the UK relied solely on a reporting radiographer as the reference standard.[18] This study was included based on jurisdictional evidence supporting radiographer reporting roles and sufficient detail in the study to suggest diagnostic performance equivalent to that of a radiologist.[19] Common to all included studies was the fact that each reference standard chosen per study was applied to all TXRs included in each study.

Table 2. General characteristics of studies included per the review's context.

Study reference	Country	Study design	Professional	Clinical setting	Assessment technique	Sites of patients' injury	Reference standard
Freij et al. 1996 [53]	UK	Retrospective and comparative	Senior House Officers (SHOs)	A&E	Retrospective analysis	Hand, wrist, forearm, foot, ankle, lower leg	Interpretation by a consultant radiologist
Gillard et al. 1998 [54]	UK	Comparative	3 Casualty officers	A&E	Test bank approach	Unspecified	Interpretation by a consultant radiologist
Coleman and Piper, 2009 [55]	UK	Comparative	7 Casualty officers	A&E	Test bank approach	Forearm, hand, foot, elbow, ankle, knee, shoulder	A consensus interpretation from a consultant radiologist, a senior radiology registrar and an advanced practitioner radiographer on each case.
Snaith and Hardy, 2014 [18]	UK	Prospective	Foundation years 1 and 2 doctors	A&E	Prospective analysis	Upper limb, lower limb, pelvis/hip, spine, thoracic cage, skull/face	Interpretation by a reporting radiographer
du Plessis and Pitcher, 2015 [56]	South Africa	Comparative	8 medical officers < 3 years of experience	A&E	Test bank approach	Appendicular skeleton	Consensus interpretation of three consultant radiologists on each case
Ofori-Manteaw & Dzidzornu 2019 [57]	Ghana	Comparative	12 JDs (6 months to 2 years of experience)	A single health facility	Test bank approach	Appendicular skeleton	Interpretation by a consultant radiologist
Liu et al. 2022 [15]	South Africa	Retrospective	JDs	A&E	Retrospective analysis	Skull, facial bones, spine, upper/lower limbs, pelvis, chest, abdomen	A consensus interpretation by a radiology consultant with 11 years' experience and a senior radiology registrar with 5 years' experience on each case

NB: *UK= United Kingdom, A&E= Accident and Emergency, JD= Junior doctor, Unspecified= the study wasn't specific about the actual sites of patients' injury, but was skeletally oriented, thus from either the appendicular or axial region. "Study design" describes the methodological approach used, such as retrospective (analysing previously recorded JDs' interpretations), prospective (collecting data of JDs' interpretations in real time), or comparative (comparing the interpretation of JDs to other healthcare professionals). "Professional" specifies the type and level of junior doctors or medical staff assessed, reflecting their* *training and experience. "Clinical setting" denotes the environment in which the assessment took place, typically the Accident & Emergency (A&E) department. "Assessment technique" explains how the doctors' interpretation skills were evaluated, either using a test bank approach (pre-selected set of trauma X-ray cases) or retrospective/prospective analysis of real cases. "Sites of patients' injury" refer to the anatomical sites where traumatic injuries occurred and from which the corresponding TXRs interpreted by junior doctors were taken. "Reference standard" indicates the gold standard against which the junior doctors' interpretations were compared, usually the assessment of a consultant radiologist or a consensus interpretation by multiple experienced radiology professionals."*

From Table 3, there were 443 TXRs across studies, with 253 being normal and 190 being abnormal. One study[54] reported the performance of the JDs in interpreting the same TXRs on film and a digital workstation under controlled conditions. Their performance on film was recorded as it was slightly higher than on the digital workstation. Additionally, one study[18] also reported the interpretation accuracy of JDs in two different arms, i.e. immediate and delayed reporting arms. While reports were readily available from advanced practice radiographers in the immediate reporting arm, JDs had to provide their own interpretation in the delayed reporting arm. Therefore, JDs' performance in the delayed reporting arm was included in our analysis. Further, one study[57] also reported pre-training and post-training TXR interpretation accuracy among JDs. The pre-training results were selected for our analysis. In one study,[15] the overall initial interpretation accuracy of JDs was categorised by anatomical region of TXRs, i.e. appendicular, axial, chest, and abdomen, and each category met the inclusion criteria; therefore, data were extracted and analysed separately for each region, increasing the number of study parts from seven (7) from all seven studies to ten (10).

Five studies (five study parts)[15,53,55,56,57] reported the interpretive accuracy of JDs on TXRs from the appendicular region. One study each reported the interpretive accuracy of JDs on TXRs from both appendicular and axial regions[18], axial region only[15], chest[15] and abdomen[15]. One study[54] was not specific with the exact anatomical region of TXRs JDs interpreted, but was skeletally oriented. In four of the studies, sufficient information was provided for the mathematical calculation of a 2x2 table with matrix values (TP, FP, TN, FN). [54,55,56,57] Four studies[54,55,56,57] employed a specific number of JDs to each interpret the same set of TXRs; therefore, the number of interpretations is also presented.

Table 3. Characteristics of junior doctors' trauma X-ray interpretation.

Study	Anatomical region of trauma X-rays	Number trauma X-rays			Number of interpretations			Data				Results	
		Total	normal	abnormal	Total	normal	abnormal	TP	FP	TN	FN	Sensitivity	Specificity
Freij et al. 1996 [53]	Appendicular	124	80	44	124	80	44	41	6	74	3	93%	92%
Gillard et al. 1998 [54]	Skeletal	32	4	28	96	12	84	63*	2*	10*	21*	75%	83.3%
Coleman and Piper, 2009 [55]	Appendicular	20	6	14	140	42	98	50*	18*	24*	48*	51%	57%
Snaith and Hardy, 2014 [18]	Appendicular + Axial	57	45	12	57	45	12	12	2	43	0	100%	95.6%
du Plessis and Pitcher, 2015 [56]	Appendicular	40	10	30	320	80	240	165*	28*	52*	75*	68.7%	65.5%
Ofori-Manteaw & Dzidzornu 2019 [57]	Appendicular	30	15	15	360	180	180	122*	44*	136*	58*	67.8%	75.6%
Liu et al. 2022 (app.) [15]	Appendicular	70	44	26	70	44	26	11	1	43	15	42%	98%
Liu et al. 2022 (ax.) [15]	Axial	14	11	3	14	11	3	1	1	10	2	33%	91%
Liu et al. 2022 (cht.) [15]	Chest	52	36	16	52	36	16	6	1	35	10	38%	97%
Liu et al. 2022 (abd.) [15]	Abdomen	4	2	2	4	2	2	0	0	2	2	0%	100%

NB: ** = mathematically calculated value, TP= true positive, FP= false positive, FN= false negative, TN= true negative. Sensitivity= TP/(TP+FN). Specificity= TN/(FP+TN). app. = appendicular ax.= axial cht.= chest abd.= abdomen.* ***NB:*** *In one* ***Liu et al., 2022*** *the overall initial interpretation accuracy of JDs was categorised by anatomical region of TXRs, i.e. appendicular, axial, chest, and abdomen, and each category met the inclusion criteria; therefore, data were extracted and analysed separately for each region, increasing the number of study parts from seven (7) from all seven studies to ten (10). Appendicular includes the lower or upper extremities or the girdles. Axial includes the spine, skull/face, or thoracic cage. Skeletal means either from the appendicular or axial region. Chest means the soft tissues of the chest. Abdomen means the abdominal soft tissues.*

Quality assessment

Figure 2 summarises the results of the risk of bias assessment of each study part across the seven studies using QUADAS-2. There was a high risk of bias for the patient selection domain across four study parts.[54,55,56,57] This is because these studies did not sufficiently describe the sample of TXRs to judge whether the spectrum of patients was selected consecutively or randomly. Regarding flow and timing, there was an unclear risk of bias in three study parts, as the studies[55,56,57] did not provide enough information to judge whether all patients were included in the analysis. There was one study part with applicability concerns to the research question regarding the conduct of the index test.[18] This is because, although this review focused on the initial independent interpretation of TXRs by JDs, the study allowed JDs to seek assistance from other medical staff if they wanted to, potentially compromising the independence of their interpretation.

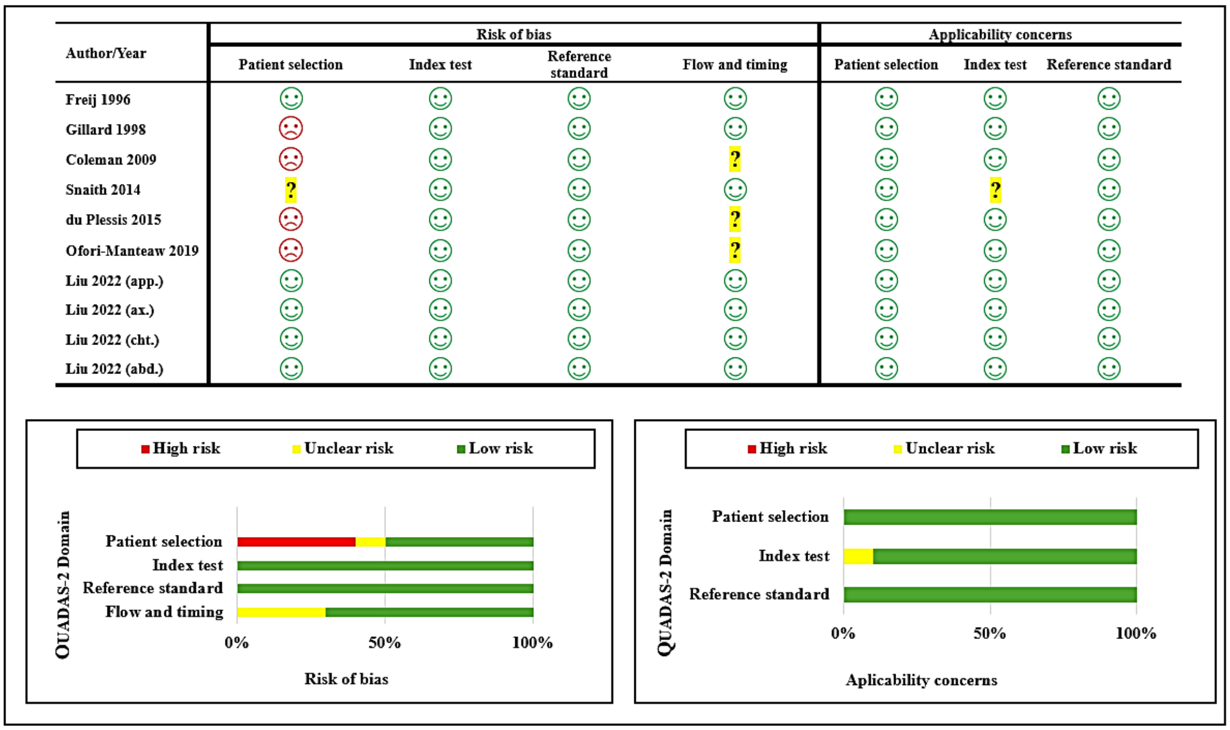

Figure 2. QUADAS-2 assessment results for the risk of bias and applicability.

NB: ☹ = high risk ☺ =low risk **?** = unclear risk.

Accuracy of junior doctors' TXR interpretation

Overall, the pooled sensitivity, specificity and AUC of JDs' interpretation of plain TXRs across the seven studies (10 study parts) were 0.65 (95% CI= 0.47 – 0.80), 0.89 (95% CI= 0.77 – 0.95), and 0.86 (95% CI= 0.83 – 0.89), respectively. The coupled forest plot and SROC curve representing this analysis have been presented in Figures 3 and 4, respectively.

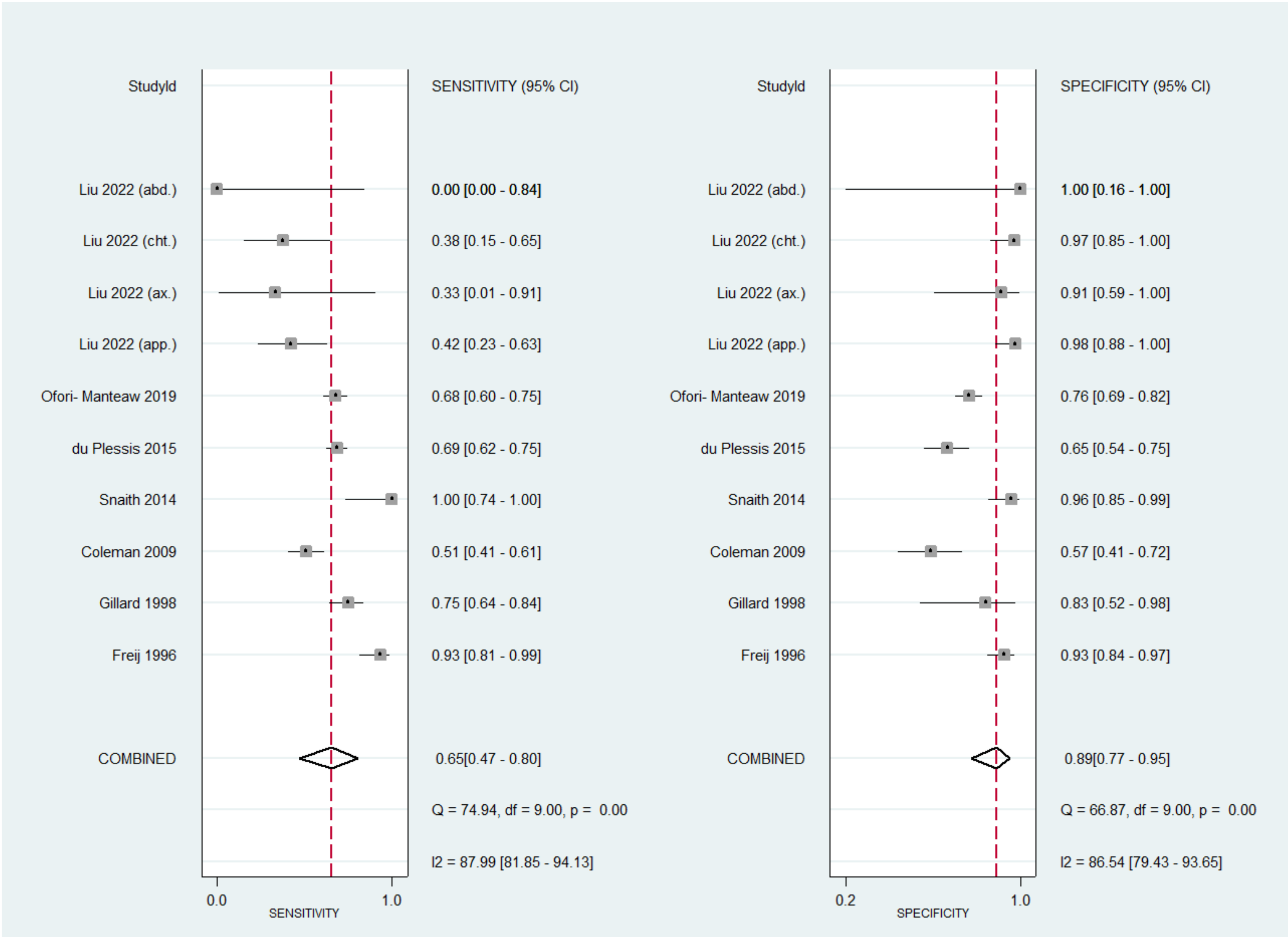

Figure 3. Coupled forest plot of junior doctors’ overall accuracy in plain trauma X-ray interpretation.

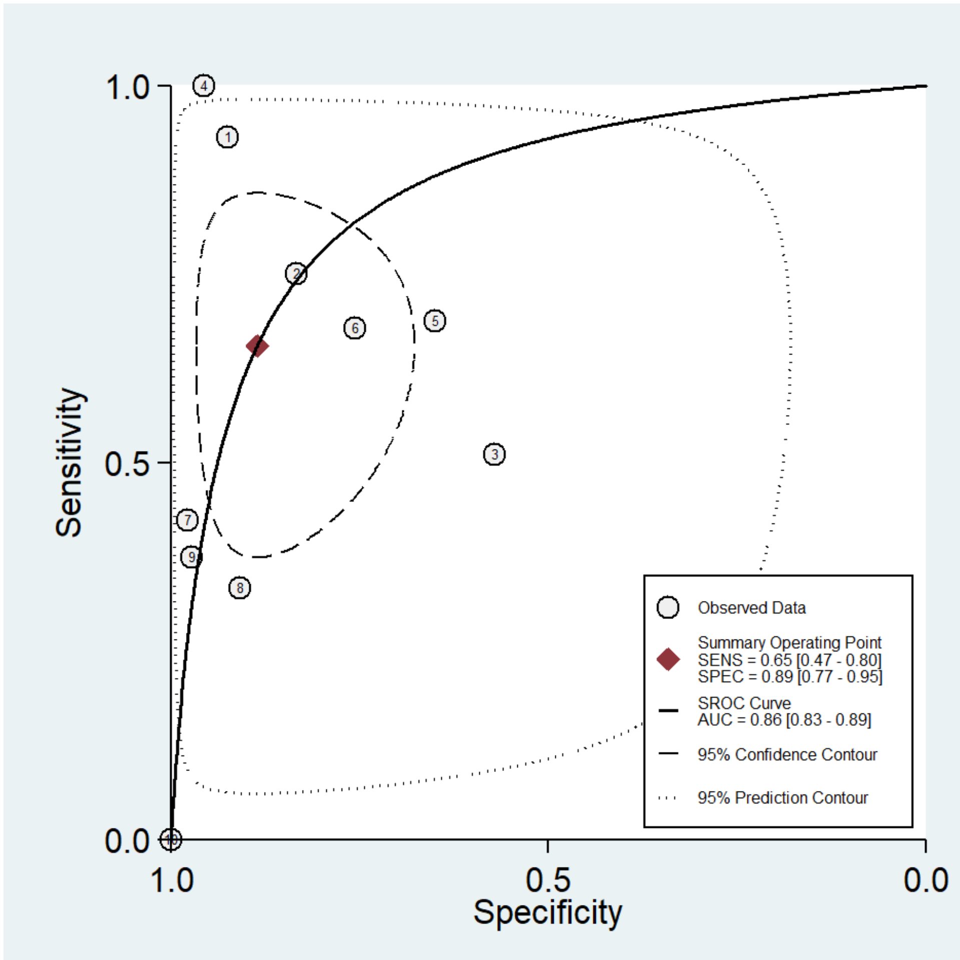

Figure 4. Summary receiver operating curve of junior doctors' overall plain trauma X-rays interpretation accuracy.

Covariate analysis of junior doctors' interpretation accuracy per anatomical region of TXRs

Covariate analysis was performed to assess JDs' TXR interpretation accuracy by anatomical region of TXRs; this was possible for skeletal and appendicular anatomical subdivisions only. This is because for axial, chest and abdomen, the study parts were one each. For skeletal TXR interpretation, JDs' pooled sensitivity, specificity, and AUC were 0.72 (95% CI= 0.55 – 0.85), 0.86 (95% CI= 0.73 – 0.93), and 0.87 (95% CI= 0.83 – 0.89), respectively. Again, for appendicular plain TXR interpretation, JDs' pooled sensitivity, specificity, and AUC were 0.68 (95% CI= 0.49 – 0.82), 0.82 (95% CI= 0.62 – 0.93), and 0.81 (95% CI= 0.77 – 0.84), respectively. From Table 4, these results are presented as well as their respective diagnostic odds ratios.

Table 4. Summary estimates of combined effect sizes.

Anatomical region of TXRs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	AUC (95% CI)
Skeletal	0.72 (0.55 – 0.85)	0.86 (0.73 – 0.93)	16 (4 - 56)	0.87 (0.83 – 0.89)
Appendicular	0.68 (0.49 – 0.82)	0.82 (0.62 – 0.93)	10 (2 – 39)	0.81 (0.77 – 0.84)

NB: DOR= Diagnostic Odds Ratio, CI= confidence interval, AUC= Area under the receiver operating characteristics curve, skeletal= either axial or appendicular regions, Appendicular= only the appendicular region.

Heterogeneity and threshold effect

From Figure 3, there was substantial heterogeneity in sensitivity and specificity in the meta-analysis conducted. Specifically, the Higgins I^2 for sensitivity and specificity was 87.99% and 86.54% respectively. Moreover, the considerable difference observed between the 95% confidence and prediction contour in the SROC curve affirms these substantial heterogeneities[58] (Figures 4).

On investigating whether heterogeneity was due to threshold effect using Spearman correlation, it was found that the correlation coefficient was 0.29 (p= 0.49), indicating that heterogeneity was not due to threshold effects. Table 5 further summarises these estimates.

Table 5. Summary estimates for heterogeneity and threshold effect.

	Values
I^2 (sensitivity)	87.99% (95% CI= 81.85 – 94.13)
I^2 (Specificity)	86.54% (79.43 – 93.65)
Spearman correlation	0.2857 (p= 0.4927)

I^2= Higgins I^2 CI= confidence interval

Meta-regression and subgroup analysis

Considering the absence of a threshold effect in this study, meta-regression and subgroup analysis was performed to determine the sources of heterogeneity. Specifically, meta-regression was used to assess whether study-level characteristics influenced variations in sensitivity and specificity across the included studies. The variables included anatomical region of TXRs, geographic location of studies, and assessment technique. For the anatomical region of TXRs, two parameters were examined: skeletal region (thus either appendicular or axial regions) and appendicular region only. Studies that reported the initial TXR interpretive accuracy of JDs on skeletal regions were coded as 1 (yes) and all others as 0 (no). Similarly, studies that reported the initial TXR interpretive accuracy of JDs on the appendicular region were coded as 1 (yes) and all others as 0 (no). For geographic location, the parameter was Africa. Studies conducted in Africa were coded as 1 (yes) and studies from outside Africa were coded as 0 (no). For the assessment technique, the parameter was test bank approach. Studies that employed a test bank approach in the assessment of JDs' interpretive accuracy were coded as 1 (yes), while studies using a prospective or retrospective analysis approach were coded as 0 (no).

The results revealed that the anatomical region of TXRs and the technique employed in the assessment of JDs' accuracy caused heterogeneity. JDs had a higher sensitivity in interpreting appendicular TXRs than non-appendicular TXRs. Moreover, their sensitivity was higher in studies that employed the test bank assessment technique than in studies that employed either a prospective or retrospective analysis approach. For specificity, there were no significant differences identified across these variables. Despite this, heterogeneity remained substantial across the majority of the combined effect sizes. Table 6 presents the results.

Table 6. Summary of the meta-regression and subgroup analysis.

Variable	**Parameter**	**Category/subgroup**	**No. of studies**	**Sensitivity [95% CI]**	**p-value**	**Specificity [95% CI]**	**p-value**	**I^2**
Anatomical region of TXRs	Skeletal	Skeletal (Yes)	8	0.47 [0.26 - 0.67]	0.10	0.67 [0.52 - 0.83]	0.20	0.69
		Non-skeletal (No)	2	0.11 [-0.10 - 0.32]		0.93 [0.78 - 1.00]		
	Appendicular	Appendicular (Yes)	5	0.50 [0.32 - 0.67]	0.01*	0.69 [0.56 - 0.81]	0.15	0.94
		Non-appendicular (No)	3	0.10 [-0.03 - 0.24]		0.88 [0.70 - 1.00]		
Geographic location	Africa	Africa (Yes)	6	0.29 [0.06 - 0.52]	0.27	0.78 [0.60 - 0.97]	0.51	0.35
		UK (No)	4	0.54 [0.25 - 0.84]		0.63 [0.34 - 0.91]		
Assessment technique	Test bank approach	Test bank approach (Yes)	4	0.70 [0.58 - 0.83]	0.00*	0.74 [0.57 - 0.91]	0.74	0.86
		Prospective or retrospective assessment (No)	6	0.21 [0.11 - 0.31]		0.73 [0.52 - 0.94]		

NB: ** = Statistically significant, I^2= heterogeneity, CI= Confidence interval, TXR= Trauma X-rays, UK= United Kingdom.*

Publication bias

Figure 5 shows Deeks' test results for JDs' overall plain TXR interpretation accuracy. The p-value of 0.41 indicated that there was no publication bias.

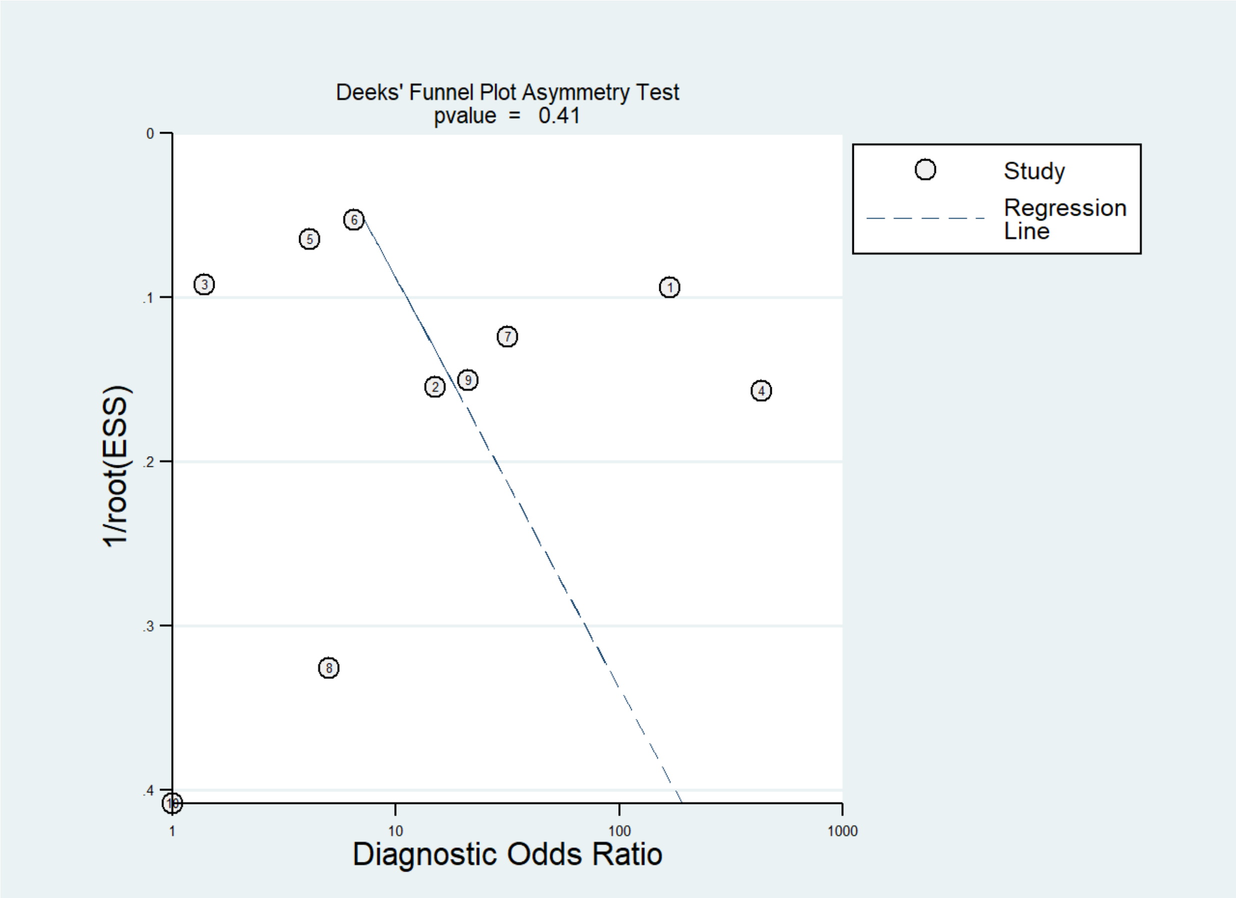

Figure 5. Deeks' test for publication bias across the included studies.

Discussion

JDs in some jurisdictions play a role in assessing and interpreting plain X-ray findings in trauma-related cases within emergency departments, often making immediate management decisions based on those interpretations.[15,18,23,55] This typically occurs in contexts where immediate access to a radiologist or reporting radiographer interpretation is not feasible,[15,55] particularly in resource-limited settings lacking sufficient numbers of trained radiologists and where the role of reporting radiographers is not yet established.[15] While this form of task-shifting can facilitate rapid decision-making during acute presentations,[56] concerns remain about interpretation errors, which could negatively impact patient care. Accordingly, the present study was aimed at evaluating the accuracy of JDs' interpretation of TXRs.

This meta-analysis involving seven studies (10 eligible study parts) indicated that compared with a reference standard, JDs' pooled sensitivity, specificity, and AUC in interpreting TXRs (across all studies, skeletal TXR interpretation and appendicular TXR interpretation) were 0.65 – 0.72, 0.82 – 0.89, and 0.81 – 0.87, respectively (Figure 4 and Table 4). These findings consistently indicate that JDs had a moderate accuracy when interpreting plain TXRs. Specifically, they demonstrated a high ability to classify true negative radiographs correctly (11%-18% false positive rate) compared to a moderate ability to classify true abnormal radiographs (28%-35% false negative rate), suggesting a considerable likelihood of missed diagnosis.

Our findings align with a Danish study on JDs' interpretation of a set of chest X-rays, which found that their sensitivity was lower than their specificity.[40] The current evidence, which suggests that JDs may be able to rule out abnormalities when there are none present on a TXR rather than to detect them, raises safety concerns, particularly in trauma care, where missed fractures or internal injuries can lead to significant morbidity. By far, this is especially troubling.

Most patients presenting to EDs are initially evaluated by JDs before senior clinician involvement. However, undergraduate training and early clinical exposure often do not adequately prepare JDs for independent practice at the ED.[31] Yet the ED is often one of the first environments where junior doctors are personally responsible for initial management and the discharge of patients,[23] for which, if the presentation were trauma, their management decision may be informed by their initial interpretation of a patient's TXR. Given their demonstrated sensitivity, reliance on such interpretation could lead to inadequate treatment and, inevitably, the recall of a discharged patient where misinterpreted findings are clinically significant. This can also pose some medico-legal consequences for JDs.

Professional ethics demand that patients are given the best possible care[59,] and hence, consultant and senior clinicians at the ED have a responsibility to ensure that patients attending the emergency department are correctly diagnosed and managed. While this can take several forms, it can also include encouraging JDs to seek reviews from senior doctors at the ED when uncertain about plain TXR findings.[14,38,60] Also, targeted interventions, such as incorporating TXR interpretation training for JDs, may be helpful.[29,57]

Moreover, radiographers could also be useful in this regard, with studies having demonstrated the effectiveness of hot reporting by advanced practice radiographers on junior doctors' interpretation accuracy at the ED.[18] In jurisdictions where this may not be feasible, junior doctors can collaborate with radiographers to avoid misinterpretations, as demonstrated in a UK study.[61] As a backup, teleradiology services can also be employed to obtain a prompt radiologist's report.[62] Additionally, commercially available AI fracture detection systems, which have been demonstrated to have the potential to improve fracture detection when used alongside humans,[63] could serve as an additional tool to help reduce misinterpretation among JDs.

Substantial heterogeneity was observed in the pooled sensitivities and specificities in our analyses. On the one hand, this may come from differences in the experiences, training or case exposure among JDs across studies. On the other hand, this may come from factors related to the TXRs, like complexity or the presence of multiple findings on one single radiograph.

By conducting meta-regression and subgroup analyses of JDs TXR interpretation accuracy, this review explored the sources of heterogeneity and found that studies that reported the interpretation accuracy of JDs on appendicular TXRs recorded higher sensitivity. For this result, the authors believe that this may be due to familiarity with appendicular trauma cases, which are a common trauma-related presentation to the emergency department.[15,64] Again, it was identified that studies that reported the use of test bank approach in the assessment of JDs' interpretive accuracy recorded higher sensitivity among JDs compared with studies that reported a prospective or retrospective analysis. Notably, studies employing test bank assessments in this review tended to include a disproportionately higher number of abnormal radiographs compared to normal radiographs (Table 3), of which, according to Hardy et al.[64] have the potential to overestimate X-ray interpretation accuracy, especially sensitivity. Therefore, it is recommended that assessments aiming to evaluate interpretive competency using a test bank approach should be carefully constructed so that cases selected will reflect local clinical realities, ensuring a balanced and representative case matrix.[65]

Although it is unlikely that the overall conduct of this review introduced bias, one notable methodological shortcoming of the studies included was that only two of the studies (five study parts) described the sampled trauma X-rays to enable the judgment that the spectrum of patients was representative of those in practice, i.e. consecutive or random sampling of patients. Thus, there was a likelihood of patient selection bias in the remaining five studies. Again, in one study, there was the likelihood of JDs seeking assistance during their initial interpretation, which may have influenced the independence of the index test and raised applicability concerns. Additionally, not all included studies reported full 2×2 contingency values (i.e., TP, FP, TN, FN). For such studies, the authors were contacted, but raw data were not retrieved. As a result, these values were mathematically calculated from reported sensitivity, specificity, case counts and prevalence. Again, despite conducting meta-regression and subgroup analysis due to substantial heterogeneities identified, significant heterogeneity persisted in the combined effect sizes. However, this is not unusual in diagnostic test reviews and may be attributed to variations in demographic characteristics or study design. Also, the studies did not provide enough information to explore other JD-specific or plain TXR-related factors that could have influenced heterogeneity, like experience and case complexity. Therefore, it remains unclear the influence of these factors on JDs' accuracy. Moreover, the number of studies meta-analysed was small and from only three countries, affecting the generalisability of findings. Therefore, caution should be taken when generalising findings.

One methodological strength among the studies included was that the reference standard employed in each included study was applied to all the trauma radiographs employed in each respective study, and hence eliminated any potential for differential verification bias, of which evidence suggests can overestimate accuracy.[49,66] Also, the reference interpretations were generated without the knowledge of JDs. This review and meta-analysis employed a comprehensive and rigorous methodology, adherence to PRISMA-DTA guidelines for reporting, and guidelines for conducting diagnostic test accuracy reviews. This, coupled with a review protocol registered before full conduct, ensured transparency. Furthermore, there was no publication bias in the meta-analysis conducted.

Conclusion

The accuracy of ED JDs in interpreting TXRs was moderate. Specifically, they demonstrated a high ability to classify true negative findings correctly compared to a moderate ability to classify true abnormalities. Given that JDs often make initial clinical decisions in emergency departments based on these interpretations, especially in settings lacking immediate access to radiologists or reporting radiographers, these limitations may have direct consequences for patient safety and quality care. Incorporating structured support systems, including second-reader protocols by senior clinicians, targeted training in trauma radiograph interpretation, teleradiology services, or collaboration with radiographers at the ED could improve their accuracy and possibly their initial clinical decisions. Future studies can consider JD interpretation accuracy in trauma axial, chest or abdomen, the clinical and medicolegal implications of interpretation errors, and explore factors that may influence accuracy, as well as evidence-based strategies for improving accuracy.

List of abbreviations

A&E Accident and Emergency

ED Emergency department

TXR Trauma X-ray

UK United Kingdom

JD Junior doctor

TP True positive

FP False positive

TN True negative

FN False negative

SROC Summary Receiver Operating Characteristic

AUC Area under the receiver operating characteristic curve

SHO senior house officer

CI confidence interval

DOR Diagnostic Odds Ratio

Declarations

Ethical approval

Since this was a review, no ethical approval was needed

Consent for publication

N/A

Availability of data and materials

The data sets used and analysed during the current study are available from the corresponding author upon reasonable request.

Competing interest

The authors have no competing interest to declare

Funding

This research did not receive any specific funding.

Author's contributions

GA contributed to the conceptualisation, design, data acquisition, analysis, interpretation of data, final draft and revision. ICA-K contributed to the conceptualisation, design, data acquisition, analysis, interpretation of data, final draft and revision. AD contributed to the analysis, interpretation of data, final draft and revision. YAW contributed to the analysis, interpretation of data, final draft and revision. BOB contributed to the analysis, interpretation of data, final draft and revision. MJN contributed to the data analysis, interpretation of data, final draft and revision. PCB contributed to the data analysis, interpretation of data, final draft and revision.

References

1. World Health Organization, Injuries and Violence. https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence, 2024 (Accessed 6 February 2025).

2. Craig HA, Lowe DJ, Khan A, Paton M, Gordon MW. Exploring the impact of traumatic injury on mortality: An analysis of the certified cause of death within one year of serious injury in the Scottish population. Injury. 2024 Jun 1;55(6):111470. https://doi.org/10.1016/j.injury.2024.111470

3. Rossiter ND. Trauma—the forgotten pandemic?. International orthopaedics. 2022 Jan;46(1):3-11. https://doi.org/10.1007/s00264-021-05213-z

4. Abebe T, Boru Y, Belay E, Abebe A, Kefyalew M, Kifle F. Clinical profile and outcomes of trauma patients visiting the emergency department of a trauma

center Addis Ababa, Ethiopia. African journal of emergency medicine. 2022 Dec 1;12(4):478-83. https://doi.org/10.1016/j.afjem.2022.10.013

5. World Health Organization, Fragility fractures. https://www.who.int/news-room/fact-sheets/detail/fragility-fractures, 2024 (Accessed 6 February 2025).

6. Wu AM, Bisignano C, James SL, Abady GG, Abedi A, Abu-Gharbieh E, Alhassan RK, Alipour V, Arabloo J, Asaad M, Asmare WN. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. The Lancet Healthy Longevity. 2021 Sep 1;2(9):e580-92. https://doi.org/10.1016/s2666-7568(21)00172-0

7. Court-Brown CM, Caesar B. Epidemiology of adult fractures: a review. Injury. 2006 Aug 1;37(8):691-7. https://doi.org/10.1016/j.injury.2006.04.130

8. Polinder S, Haagsma J, Panneman M, Scholten A, Brugmans M, Van Beeck E. The economic burden of injury: Health care and productivity costs of injuries in the Netherlands. Accident analysis & prevention. 2016 Aug 1;93:92-100. https://doi.org/10.1016/j.aap.2016.04.003

9. Sadat-Ali M, Alomran AS, Azam Q, Al-Sayed HN, Al-Dhafer BA, Kubbara AF, AlShaikh SH. Epidemiology of fractures and dislocations among urban communities of eastern Saudi Arabia. Saudi Journal of Medicine & Medical Sciences. 2015 Jan 1;3(1):54-7. https://10.4103/1658-631X.149682

10. Bergh C, Wennergren D, Möller M, Brisby H. Fracture incidence in adults in relation to age and gender: a study of 27,169 fractures in the Swedish Fracture Register in a well-defined catchment area. PloS one. 2020 Dec 21;15(12):e0244291. https://doi.org/10.1371/journal.pone.0244291

11. Van Staa TP, Dennison EM, Leufkens HA, Cooper C. Epidemiology of fractures in England and Wales. Bone. 2001 Dec 1;29(6):517-22. https://doi.org/10.1016/s8756-3282(01)00614-7

12. Cooper C, Dennison EM, Leufkens HG, Bishop N, van Staa TP. Epidemiology of childhood fractures in Britain: a study using the general practice research database. Journal of bone and mineral research. 2004 Dec 1;19(12):1976-81. https://doi.org/10.1359/jbmr.040902

13. Rupp M, Walter N, Pfeifer C, Lang S, Kerschbaum M, Krutsch W, Baumann F, Alt V. The incidence of fractures among the adult population of Germany: An analysis from 2009 through 2019. Deutsches Ärzteblatt International. 2021 Oct 8;118(40):665. https://doi.org/10.3238/arztebl.m2021.0238

14. Guly HR. Diagnostic errors in an accident and emergency department. Emergency Medicine Journal. 2001 Jul 1;18(4):263-9. https://doi.org/10.1136/emj.18.4.263

15. Liu YY, O'Hagan S, Holdt FC, Lahri SA, Pitcher RD. After-hour trauma-radiograph interpretation in the emergency centre of a District Hospital. African Journal of Emergency Medicine. 2022 Sep 1;12(3):199-207. https://doi.org/10.1016/j.afjem.2022.04.001

16. York T, Franklin C, Reynolds K, Munro G, Jenney H, Harland W, Leong D. Reporting errors in plain radiographs for lower limb trauma—a systematic review and meta-analysis. Skeletal Radiology. 2022 Jan;51(1):171-82. https://doi.org/10.1007/s00256-021-03821-9

17. Dallaudière B, Larbi A, Lefere M, Perozziello A, Hauger O, Pommerie F, Fraboulet B, Jacob D. Musculoskeletal injuries in a resource-constrained environment: comparing diagnostic accuracy of on-the-spot ultrasonography and conventional radiography for bone fracture screening during the Paris–Dakar rally raid. Acta radiologica open. 2015 May 27;4(5):2058460115577566. https://doi.org/10.1177/2058460115577566

18. Snaith B, Hardy M. Emergency department image interpretation accuracy: The influence of immediate reporting by radiology. International emergency nursing. 2014 Apr 1;22(2):63-8. https://doi.org/10.1016/j.ienj.2013.04.004

19. Wood K. How is the reporting radiographer role portrayed in published studies? A scoping review. Radiography. 2022 Feb 1;28(1):215-21. https://doi.org/10.1016/j.radi.2021.07.027

20. Guly HR. Missed diagnoses in an accident & emergency department. Injury. 1984 May 1;15(6):403-6. https://doi.org/10.1016/0020-1383(84)90207-9

21. Beggs I, Davidson JK. Accident and emergency reporting in UK teaching departments. Clinical radiology. 1990 Apr 1;41(4):264-7. https://doi.org/10.1016/s0009-9260(05)81662-9

22. De Lacey G, Barker A, Harper J, Wignall B. An assessment of the clinical effects of reporting accident and emergency radiographs. The British Journal of Radiology. 1980 Apr 1;53(628):304-9. https://doi.org/10.1259/0007-1285-53-628-304

23. Gompels B, Rusby T, Limb R, Ralte P. Diagnostic accuracy and confidence in management of forearm and hand fractures among foundation doctors in the accident and emergency department: survey study. JMIR Formative Research. 2023 Aug 18;7(1):e45820. https://doi.org/10.2196/45820

24. Galasko CS, Monahan PR. Value of re-examining x-ray films of outpatients attending accident services. Br Med J. 1971 Mar 20;1(5750):643-4. https://doi.org/10.1136/bmj.1.5750.643

25. Sica G, Guida F, Bocchini G, Codella U, Mainenti PP, Tanga M, Scaglione M. Errors in imaging assessment of polytrauma patients. InSeminars in Ultrasound, CT and MRI 2012 Aug 1 (Vol. 33, No. 4, pp. 337-346). WB Saunders. https://doi.org/10.1053/j.sult.2012.01.012

26. Pinto A, Reginelli A, Pinto F, Lo Re G, Midiri F, Muzj C, Romano L, Brunese L. Errors in imaging patients in the emergency setting. The British journal of radiology. 2016 May 1;89(1061):20150914. https://doi.org/10.1259/bjr.20150914

27. Gwynne A, Barber P, Tavener F. A review of 105 negligence claims against accident and emergency departments. Emergency Medicine Journal. 1997 Jul 1;14(4):243-5. https://doi.org/10.1136/emj.14.4.243

28. McLauchlan CA, Jones K, Guly HR. Interpretation of trauma radiographs by junior doctors in accident and emergency departments: a cause for concern?. Emergency Medicine Journal. 1997 Sep 1;14(5):295-8. https://doi.org/10.1136/emj.14.5.295

29. Brealey S, Scally A, Hahn S, Thomas N, Godfrey C, Coomarasamy A. Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis. Clinical radiology. 2005 Feb 1;60(2):232-41. https://doi.org/10.1016/j.crad.2004.07.012

30. Christiansen JM, Gerke O, Karstoft J, Andersen PE. Poor interpretation of chest X-rays by junior doctors. Dan Med J. 2014 Jul 1;61(7):A4875. https://ugeskriftet.dk/dmj/poor-interpretation-chest-x-rays-junior-doctors

31. Cheung T, Harianto H, Spanger M, Young A, Wadhwa V. Low accuracy and confidence in chest radiograph interpretation amongst junior doctors and medical students. Internal medicine journal. 2018 Jul;48(7):864-8. https://doi.org/10.1111/imj.13946

32. Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Belfiore MP, Papapietro VR, Carotti M, Pinto F, Giovagnoni A, Romano L. Traumatic fractures in adults: missed diagnosis on plain radiographs in the Emergency Department. Acta Bio Medica: Atenei Parmensis. 2018;89(Suppl 1):111. https://doi.org/10.23750/abm.v89i1-S.7015

33. Berman L, de Lacey G, Twomey E, Twomey B, Welch T, Eban R. Reducing errors in the accident department: a simple method using radiographers. Br

Med J (Clin Res Ed). 1985 Feb 9;290(6466):421-2. https://doi.org/10.1136/bmj.290.6466.421

34. Tachakra SS, Beckett MW. Why do casualty officers miss radiological abnormalities?. Journal of the Royal College of Surgeons of Edinburgh. 1985 Oct 1;30(5):311-3.

35. Wardrope J, Chennells PM. Should all casualty radiographs be reviewed?. Br Med J (Clin Res Ed). 1985 Jun 1;290(6482):1638-40. https://doi.org/10.1136/bmj.290.6482.1638

36. Snow DA. Clinical significance of discrepancies in roentgenographic film interpretation in an acute walk-in area. Journal of general internal medicine. 1986 Sep;1:295-9. https://doi.org/10.1007/BF02596206

37. Gleadhill DN, Thomson JY, Simms P. Can more efficient use be made of x ray examinations in the accident and emergency department?. Br Med J (Clin Res Ed). 1987 Apr 11;294(6577):943-7. https://doi.org/10.1136/bmj.294.6577.943

38. Overton-Brown P, Anthony D. Towards a partnership in care: nurses' and doctors' interpretation of extremity trauma radiology. Journal of Advanced Nursing. 1998 May;27(5):890-6. https://doi.org/10.1046/j.1365-2648.1998.t01-1-00565.x

39. Bell LT, Dick O, Ali N, Little D. Undergraduate radiology education: foundation doctors' experiences and preferences. Clinical radiology. 2019 Jun 1;74(6):480-6. https://doi.org/10.1016/j.crad.2019.01.029

40. Jimah BB, Baffour Appiah A, Sarkodie BD, Anim D. Competency in chest radiography interpretation by junior doctors and final year medical students at a teaching hospital. Radiology Research and Practice. 2020;2020(1):8861206. https://doi.org/10.1155/2020/8861206

41. British Medical Association, "Junior" doctors change name to "resident" doctors. https://www.bma.org.uk/bma-media-centre/junior-doctors-change-name-to-resident-doctors?utm, 2024 (Accessed 8 February 2025).

42. World Medical Association, Junior doctors/about us. https://www.wma.net/junior-doctors/about-us/, 2025 (Accessed on 8 February 2025).

43. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clinical Microbiology and Infection. 2014 Feb 1;20(2):105-13. https://doi.org/10.1111/1469-0691.12474

44. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editors. Cochrane handbook for systematic reviews of diagnostic test accuracy. John Wiley & Sons; 2023 Jun 13. Available from https://training.cochrane.org/handbook-diagnostic-test-accuracy/current.

45. McInnes MD, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L, Hunt HA. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. Jama. 2018 Jan 23;319(4):388-96. https://doi.org/10.1001/jama.2017.19163

46. Rosner B. *Fundamentals of biostatistics*. 8th ed. Boston (MA): Cengage Learning; 2015.

47. Bečulić H, Begagić E, Džidić-Krivić A, Pugonja R, Softić N, Bašić B, Balogun S, Nuhović A, Softić E, Ljevaković A, Sefo H. Sensitivity and specificity of machine learning and deep learning algorithms in the diagnosis of thoracolumbar injuries resulting in vertebral fractures: A systematic review and meta-analysis. Brain and Spine. 2024 Apr 17:102809. https://doi.org/10.1016/j.bas.2024.102809

48. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011 Oct 18;155(8):529-36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

49. Brealey S, Scally A, Hahn S, Thomas N, Godfrey C, Crane S. Accuracy of radiographers red dot or triage of accident and emergency radiographs in clinical practice: a systematic review. Clinical radiology. 2006 Jul 1;61(7):604-15. https://doi.org/10.1016/j.crad.2006.01.015

50. Dwamena B. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies. 2009

51. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988 Jun 3;240(4857):1285-93. https://doi.org/10.1126/science.3287615

52. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. Statistical methods of meta-analysis. Korean journal of radiology. 2015 Dec 1;16(6):1188-96. https://doi.org/10.3348/kjr.2015.16.6.1188

53. Freij RM, Duffy T, Hackett D, Cunningham D, Fothergill J. Radiographic interpretation by nurse practitioners in a minor injuries unit. Emergency Medicine Journal. 1996 Jan 1;13(1):41-3. https://doi.org/10.1136/emj.13.1.41

54. Gillard JH, Hubbard C, Das R, Sherriff H. Digital radiology in skeletal trauma: assessment of casualty officers' performance. Journal of the Royal Society of Medicine. 1998 Mar;91(3):129-32. https://doi.org/10.1177/014107689809100304

55. Coleman L, Piper K. Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers. Radiography. 2009 Aug 1;15(3):196-202. https://doi.org/10.1016/j.radi.2007.12.001

56. du Plessis J, Pitcher R. Towards task shifting? A comparison of the accuracy of acute trauma-radiograph reporting by medical officers and senior radiographers in an African hospital. Pan African Medical Journal. 2015;21(1). https://doi.org/10.11604/pamj.2015.21.308.6937

57. Ofori-Manteaw BB, Dzidzornu E. Accuracy of appendicular radiographic image interpretation by radiographers and junior doctors in Ghana: Can this be improved by training?. Radiography. 2019 Aug 1;25(3):255-9. https://doi.org/10.1016/j.radi.2019.04.003

58. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

59. Varkey B. Principles of clinical ethics and their application to practice. Medical Principles and Practice. 2021 Jun 4;30(1):17-28.

60. Kinnersly J, Ahmed F, Selman C, Bourke EM. Skeletal radiograph interpretation discrepancies in the emergency department setting: A retrospective chart review. Emergency Medicine Australasia. 2025 Feb;37(1):e14539. https://doi.org/10.1111/1742-6723.14539

61. Kelly BS, Rainford LA, Gray J, McEntee MF. Collaboration between radiological technologists (radiographers) and junior doctors during image interpretation improves the accuracy of diagnostic decisions. Radiography. 2012 May 1;18(2):90-5. https://doi.org/10.1016/j.radi.2011.06.002

62. Agrawal A. Emergency teleradiology-past, present, and, is there a future?. Frontiers in radiology. 2022 Jun 24;2:866643.

63. Husarek J, Hess S, Razaeian S, Ruder TD, Sehmisch S, Müller M, Liodakis E. Artificial intelligence in commercial fracture detection products: a systematic review and meta-analysis of diagnostic test accuracy. Scientific Reports. 2024 Oct 4;14(1):23053.

64. Vincent CA, Driscoll PA, Audley RJ, Grant DS. Accuracy of detection of radiographic abnormalities by junior doctors. Emergency Medicine Journal. 1988 Jun 1;5(2):101-9. https://doi.org/10.1136/emj.5.2.101

65. Hardy M, Flintham K, Snaith B, Lewis EF. The impact of image test bank construction on radiographic interpretation outcomes: a comparison study. Radiography. 2016 May 1;22(2):166-70. https://doi.org/10.1016/j.radi.2015.10.010

66. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, Van Der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. Jama. 1999 Sep 15;282(11):1061-6. https://doi.org/10.1001/jama.282.11.1061