Ultrasound in bone fracture diagnosis – a comparative meta-analysis and systematic review

Eric Matschiner1,2, Oana Şerban3, Daniela Fodor3, Michael Blaivas4, Rudolf Horn5, Jonas Koch1, Marie-Lise Jakobi1, Lara Grevelding6, Joseph Osterwalder7, David Srivastava2, Christoph Frank Dietrich1

1Department General Internal Medicine (DAIM), Hospitals Hirslanden Bern Beau Site, Salem and Permanence, Bern, Switzerland, 2Department of Emergency Medicine Inselspital, University Hospital Bern, University of Bern, Switzerland, 3”Iuliu Hatieganu” University of Medicine and Pharmacy, 2nd Internal Medicine Department, Cluj-Napoca, Romania, 4Blaivas Michael. University of South Carolina School of Medicine, Department of Internal Medicine, Columbia, South Carolina, USA, 5Center da sandà Val Müstair, Sta. Maria, Switzerland, 6Department of Pediatrics, Division of Pneumology, Allergology, Infectious diseases and Gastroenterology, University Hospital, Frankfurt, Goethe University, Frankfurt, Germany, 7Executive Board, Polipraxis Group, St. Gallen, Switzerland

Introduction

Diagnosing bone fractures promptly and accurately is critical in trauma care, where functional impairment and pain necessitate rapid and precise decision-making for effective treatment. Additionally, fractures in austere or remote environments may benefit from rapid diagnosis for timely reduction/treatment or decision making regarding expensive or costly patient evaluation, especially when higher level medical care is days away. With an estimated 178 million fractures globally in 2019, translating to approximately 2296.2 incidents per 100,000 individuals [1], the demand for an efficient and standardized diagnostic approach is unequivocal. Traditionally, plain radiography has served as the cornerstone of initial fracture assessment in most clinical guidelines [2,3]. However, the complexity of certain fractures may neces-

Abstract

Aim: This meta-analysis evaluates the diagnostic accuracy of ultrasound (US) for bone fractures over the past 47 years, comparing it to established imaging standards. Material and methods: We adhered to PRISMA 2020 guidelines to search Medline, EMBASE, and the Cochrane Library using tailored search strategies. The primary outcome, US diagnostic performance, was analyzed across various subgroups including clinical relevance, patient age, and anatomical considerations. The QUADAS-2 tool was employed to assess study quality and minimize bias. Results: From 5,107 initially identified studies, 75 met the inclusion criteria, encompassing 7,769 participants and 3,575 diagnosed fractures. The majority of studies were prospective (79%) and compared US primarily with plain radiography (76%) and CT scans (19%). Of these, 61 studies were amenable to systematic analysis, revealing US to have a sensitivity and specificity of 91% (95% CI: 90%-92%) and 91.3% (95% CI: 90.5%-92.1%), respectively. Likelihood ratios were favorable, with a positive value of 9.955 and a negative value of 0.087, and an odds ratio of 132.67. The area under the curve stood at 0.9715, indicating high diagnostic accuracy despite significant heterogeneity (I²=81.3% for sensitivity, 89.3% for specificity). Conclusion: The evidence supports US as a highly accurate diagnostic tool for bone fractures, rivalling standard imaging methods like CT and radiography. Its notable diagnostic efficacy, combined with advantages in reducing pain, wait times, and radiation exposure, advocates for its broader application. Further validation in large-scale, randomized trials is essential to integrate US more fully into clinical guidelines for fracture management.

Keywords: ultrasound imaging; bone fractures; diagnostic accuracy
sitate the use of more advanced imaging modalities like computer tomography (CT) scans or magnetic resonance imaging (MRI) [4], known for their enhanced accuracy in diagnosing fractures in anatomically complex regions such as the hip, pelvis, skull, spine and thorax [5-12].

Modern medical ultrasound enables a non-invasive approach to fracture evaluation. Typically, ultrasound probes that utilize high-frequency sound waves (for example: emitted by a linear transducer at 4-12 MHz) [13] which can penetrate up to 5 cm in depth [14] are used to distinguish tissue densities by the variance of impedance, thereby enabling the detection of cortical irregularities indicative of fractures [15]. Depending on the clinical application different transducer types and frequency ranges are utilized to evaluate deeper or shallower structures. It has to be taken into account that linear transducers have a wide variety of frequency ranges, and a finger fracture should be evaluated in the 15 MHz range and above, for the top range of the broadband transducer. Conversely, a transducer (so far not available commercially) around the 1 MHz range will image through bone and is of advantage for evaluating injuries in the thigh. The closest we get to this is using a curvilinear probe that will image down to 3 MHz for evaluating the femur for instance.

Since its initial application in fracture diagnosis by Gerlanc et al [16] in 1975, and further classification by Ricciardi et al [17-19] in the early ’90s, documenting the process of post-surgery fracture healing, ultrasound technology has evolved significantly. Its integration into trauma care as an alternative to the invasive diagnostic peritoneal lavage (DPL) [20], particularly through the development of “Focused Assessment with Sonography for Trauma” (FAST) [21], reduced complication rates related to DPL [22]. Further technical progress enabled the creation of affordable mobile, and eventually handheld machines, [23] and helped the already ongoing growth of point-of-care ultrasound (POCUS), which has marked a paradigm shift in immediate bedside diagnostic capabilities [24], including evaluation of bone fractures [25-28]. POCUS’s utility extends across various injury types, from extremity fractures [29-32] to more specialized applications such as pediatric fracture diagnosis and fracture reduction assessments.

Despite these advancements we could find only two comprehensive meta-analysis comparing the diagnostic accuracy of ultrasound with conventional gold standard imaging across different fracture types and patient demographics [15,33]. In both meta-analyses, studies before 2000 were excluded. Schmid et al. ended the literature search in September 2016 [15] and Li E et al. in January 2021 [33]. The latter study lacks a pooled analysis of sensitivity and specificity. This review aims to close these major gaps and to update the two meta-analyses. Through an in-depth analysis segmented by patient age, fracture location, and specific clinical queries – including pediatric fractures, bone reduction procedures, examination duration, non-traumatic and stress fractures, and low-suspicion fracture detection – this study seeks to delineate the scope and limitations of ultrasound in the current diagnostic landscape.

**Material and methods**

**Study design and protocol**

This systematic review and meta-analysis were guided by the PICO framework [34,35] to formulate the research question, detailed in Supplementary Table I. Our protocol adhered to the PRISMA 2020 guidelines [36] and the specific adaptations for diagnostic test accuracy studies outlined by McInnes et al [37]. The protocol was registered on the Inplasy platform.

**Eligibility criteria**

Following the PICOST framework, we established inclusion and exclusion criteria to identify relevant studies comparing ultrasound against various imaging techniques for bone fracture diagnosis. Studies that did not allow for the separate analysis of pediatric and adult data were excluded. Although case series were considered during the preliminary screening, they were ultimately excluded due to the sufficiency of data from peer-reviewed research. Detailed criteria are presented in Supplementary Table II.

**Literature search and study selection**

A comprehensive search was conducted by M on 23rd November 2022, across MEDLINE (via PubMed.gov), Embase (via Elsevier), and the Cochrane Library. The search strategy employed a combination of thesaurus terms, subject headings, text words, MeSH Terms, synonyms, and acronyms, refined until a core list of references was established. The complete search strategy for each database is documented in Supplementary Table III. Selected studies were organized using the Rayyan QCRI tool, with the final inclusion managed via EndNote. Duplication checks were performed manually following Rayyan’s automated suggestions. We also reviewed meta-analyses and systematic reviews identified initially for potentially includable studies, ensuring no overlaps (“via other methods” in the PRISMA diagram).

**Data collection and analysis**

Data extraction followed a pre-defined form containing eight items (Supplementary Table IV), adhering to the Cochrane Handbook’s checklist [38], covering key study information and characteristics (Supplementary Tables V and VI).
Secondary outcomes. The secondary outcomes were not predefined. Secondary outcomes identified in the studies were aggregated and analyzed where comparable data existed among three or more studies.

Study flow characteristics. For study flow characteristics, we examined the sequence of clinical examination, index testing (US), and reference testing across studies, ensuring a consistent approach to fracture visibility comparison. This process information was summarized in three main questions: 1) Clinical examination performed in the beginning?; 2) Reference test results were available; 3) Were all fractures seen in the index test in comparison to the reference test?

Risk of bias assessment
An initial risk of bias assessment utilized the Critical Appraisal Skills Programme (CASP) checklist [39], followed by a comprehensive evaluation using the QUADAS-2 tool [40]. Results are summarized in a detailed table with visual bar charts for clarity provided by the form of the University of Bristol [41].

Statistical analysis
Diagnostic accuracy metrics – sensitivity (Se), specificity (Sp), likelihood ratios, odds ratios (OR), as well as the receiver operator characteristic (ROC) plot with the area under the curve (AUC) as well as the forest plot of the sensitivities – were extracted or calculated for each study. A random-effects model was employed for meta-analysis using MetaDiSc 1.4 software from the Ramón y Cajal Hospital in Madrid in Spain [42], that created pooled Se, Sp, likelihood ratio, OR, as well as heterogeneity assessment via the I² statistic, interpreted according to Higgins JPT et al [43] – I² values between 0 and 40%: nearly no significant, between 40 and 60%: moderate, between 60 and 80%: substantial and at least between 80 and 100%: considerable heterogeneity. The diagnostic utility was further explored through ROC curves, OR, and likelihood ratios. ROC curve were interpreted being more accurate closer to 1 as the perfect accuracy and 0.5 as showing no apparent accuracy [44]. OR were interpreted by the definition of Szumilas et al. [45]: “OR=1 Exposure does not affect odds of outcome, OR>1 Exposure associated with higher odds of outcome and OR<1 Exposure associated with lower odds of outcome”. Likelihood ratios were interpreted according to Jaeschke et al. [46]: greater than 10 or less than 0.1 as significant and conclusive change, between 5 and 10 or 0.1 and 0.2 moderate, between 2 and 5 or 0.2 small and between 1 and 2 or 0.5 and 1 as insignificant.

Subgroup and additional analyses
Studies were categorized based on patient age and fracture location using the 2018 OTA/AO Fracture and Dislocation Classification [47] as presented in Table I, with further subgroup analyses conducted as warranted by the data. Attention was given to studies highlighting ultrasound’s comparative effectiveness or efficiency. Detailed subgroup characteristics are delineated in Supplementary Table VII.

Results

Study selection and characteristics
Our comprehensive search yielded 5,107 manuscripts, culminating in the inclusion of 75 studies [28,48-121] after a rigorous screening process detailed in Figure 1. The final cohort emerged from an initial pool of 51 articles (13 meta-analyses and 38 individual studies) and an additional 37 “new studies” identified through systematic reviews. Exclusion criteria were applied systematically, with common reasons for exclusion being a focus solely on secondary outcomes, lack of clear age differentiation, and absence of primary outcome measurement.

The included studies span from 1996 to 2021, with a notable peak of 452 publications in 2021, indicating a growing interest in the field (Supplementary Fig 1 and 2). A total of 7,769 subjects were involved, predominantly pediatric patients (73.8%), with a slight underrepresentation of female participants in both pediatric and adult cohorts (43% and 38%, respectively). The majority of studies (79%) employed a prospective design, with diagnostic comparisons largely centered on US versus plain radiography (76%). Of these studies 61% refer to the children and 39% to the adult subgroup.

In the CT-scan subgroup, the standard technique for skull (including facial bones) and the chest cage bone fractures, 14 studies were included in the analysis. One study used MRT [121], one study physical examination
[65] and one further study surgical examination [83], each corresponding to 1%.

Nine secondary outcomes were found that qualified for evaluation (Supplementary Table VIII). Three common inclusion criteria (Supplementary Table IX) and 10 common exclusion criteria (Supplementary Table X) were identified. In 87% of the included studies a clinical examination was performed before the index test and in 58% of the studies the index test was performed before the reference test. Nearly half of the studies (49%) described that all fractures were seen in the index test in comparison to the reference test (Supplementary Table XI).

Risk of bias and study quality

A detailed risk of bias assessment revealed generally low concerns across most studies, with some exceptions in patient selection and study flow (fig 2). Each study’s specific risk profile is extensively documented in the QUADAS-2 assessment (Supplementary Table XII).

Diagnostic accuracy and meta-analysis

The meta-analysis focused on 61 studies [28,48-53, 56-60,62-65,67-72,74,75,78,80-88,90-94,96-104,107, 108,110-119,121] providing complete diagnostic data, revealing an average Se of 91% across a range of 32.4% to 100%. Specificity averaged at 91.3%, with pooled likelihood ratios (average positive and negative likelihood ratio of 9.955 and 0.087), ORs (average OR 132.67) and AUC (average AUC 0.9715) indicating strong diagnostic performance (Supplementary Tables XIII-XIX, fig 3, fig 4). However, considerable heterogeneity was observed (I²=81.3% for Se, 89.3% for Sp).

Subgroup analyses

Subgroup analyses by age showed negligible differences in Se but notable variability in Sp between pediatric (93.6%) and adult (81.7%) populations (I² 82.3% for the children and 80.4% for the adults). The likelihood ratio is defined conclusive for children and moderate for the adult population. Additional subgroup investigations, particularly those focusing on plain radiography and CT scans as comparators, reinforced ultrasound’s robust diagnostic utility (Se of 91.8% and 90.7% respectively, likelihood ratios significant and conclusive when compared to plain radiography but only moderate when compared to CT scan), albeit with substantial heterogeneity within each subgroup (I² 78.2% for radiography and 73.9% for CT subgroups).

Further analyses highlighted the high diagnostic Se of US in specific contexts (Supplementary Table XV), such
Fig 3. Forest plots of a) sensitivities, b) specificities, c) positive and d) negative likelihood ratios and e) the diagnostic odds ratio of the included 61 studies.
as fracture reduction evaluations (96.2%) and the assessment of special pediatric fractures (94.9%), with varying degrees of heterogeneity observed across different clinical scenarios.

**Additional analyses**

Clinical relevance extended into subgroups delineated by fracture location and type, with significant findings detailed in Supplementary Tables XV-XXII, suggesting nuanced applications of ultrasound across a broad spectrum of clinical needs.

**Discussion**

**Overview of findings**

This meta-analysis underscores the high diagnostic accuracy of US in identifying bone fractures, aligning with previous studies but extending the insights to specific body regions and clinical scenarios [15,29-32,33]. The comparative analyses, predominantly against the standard initial imaging technique of plain radiography, highlight US’s potential to surpass traditional methods in certain contexts [52,61,65,74,76,79,80,97,103,105,109,110,112,120]. Focusing on studies directly comparing US and plain radiography against other standards – MRT [121], physical examination [65] and surgical examination [83] – US was always noted to be more accurate. Notably, our findings reveal instances where US demonstrates superior diagnostic capabilities, especially in studies focusing on complex fractures where standard radiography’s two-dimensional limitations are apparent [52,61,76,79,97,105,112,120,122].

**Advantages of ultrasound**

Ultrasound’s multiplanar imaging capacity offers a significant advantage, enabling dynamic evaluation of the fracture site from various angles, thereby enhancing diagnostic precision [48,49,52,68,69,71,78,81,100,105,119]. This is particularly evident in assessing fracture displacement and alignment post-reduction, with studies indicating US’s comparability to radiography in evaluating reduction success [52,53,75,79,90,101]. In scenarios where radiography falls short, such as inconclusive findings or the need for further detail, US stands out as a rapid, effective alternative to more time-consuming and resource-intensive modalities like CT or MRI [57,62,64,74,82,85,88,91,93,99,100,103,107,114]. Beyond bone fracture identification, ultrasound offers detailed insights into associated soft tissue trauma and hematoma, areas where plain radiography may fall short [61,88,93,94,98]. It also holds the potential to identify stress fractures that often elude radiographic detection [54]. Incorporating US into the diagnostic toolkit enriches fracture assessments, providing a more comprehensive overview of injury and informing more nuanced treatment pathways [117].

**Time efficiency and economic considerations**

The efficiency of US, evidenced by its shorter examination times (averaging 1.5 minutes per examination compared to 27 and 37 minutes for plain radiography and CT scans, respectively) and immediate availability in emergency settings, underscores its value in acute situations, contributing to improved patient management [91,105,107]. Time management is not only essential for unstable patients, who cannot be transferred to the radiology department, but fast and reliable diagnosis impacts quality, cost effectiveness and leads to higher patient satisfaction, as referral to treatment and activation of necessary resources is implemented in a timely fashion. The widespread availability of US equipment in emergency and clinical settings [15] further contributes to its efficiency, though the speed of examination can depend on the availability and skill level of the operator [50,72,97] or the operability of the US device itself. Moreover, the relatively short training required for clinicians to become proficient in US use supports its practicality and reproducibility in diverse clinical settings [92]. Economic analyses highlight US’s cost-effectiveness, factoring in both the initial investment in equipment and ongoing training costs, presenting US as a financially viable diagnostic tool [123,124].

**Secondary outcomes and radiation exposure**

Secondary outcomes, such as reduced patient discomfort and the absence of ionizing radiation with US, reinforce its utility in pediatric populations and scenarios where radiation exposure is a concern [52,58,60,63,72,95,104,105,108,113,119,125]. The sensitivity and specificity of US in detecting common pediatric fractures highlight its potential as a frontline diagnostic tool in pediatric trauma care [49,50,53,58,60,68,72,78,94,110,113,115,119].
**Limitations and considerations**

Despite the advantages, the limitations of US, particularly its depth penetration and field of view constraints, necessitate caution in some applications, especially in joint and intraarticular fracture assessments, where visibility is restricted by the reflection of ultrasound waves at the level of the most superficial bony surface [14,52,53,56,81,86,101,115]. The variability in US performance based on operator skill also emphasizes the need for standardized training protocols to ensure reliability across diverse clinical settings [73,92,113].

**Implications and future directions**

Recent guidelines and randomized controlled trials affirming US’s non-inferiority to plane radiography in specific fracture types advocate for its broader acceptance and integration into clinical practice [126,127]. The evolving landscape of diagnostic ultrasound for fracture assessment, supported by technological advancements and increasing clinical adoption, underscores its growing significance in modern trauma care, particularly in conditions where rapid and accurate diagnosis is essential [81,91,114].

**Study limitations**

While this review highlights the promising utility of US in fracture diagnosis, the observed heterogeneity, probably resulted from the broad research question, and potential selection biases, notably due to the sole reliance on studies with German or English abstracts and a single reviewer’s (EM) categorization, warrant cautious interpretation. The literature search, conducted in November 2022, may have overlooked late-year publications, and certain studies presented challenges in quality, particularly in study flow, timing, and patient selection. Despite these limitations, the consistent high sensitivity observed in a majority of studies reaffirms ultrasound’s diagnostic utility, albeit with caution due to the noted methodological constraints.

**Conclusions**

This meta-analysis highlights US’s high diagnostic accuracy in bone fracture detection, underscoring its benefits such as reduced patient discomfort, lower costs, and absence of ionizing radiation. However, its effectiveness heavily relies on the operator’s expertise, necessitating standardized training and protocols. The presence of methodological heterogeneity calls for further validation through large-scale studies and randomized controlled trials (RCTs) to ensure reliable integration of US into fracture diagnosis guidelines and clinical practice.

**Conflict of interest:** none

**Acknowledgement:** The authors thank Dr. Ole Ackermann for very helpful advice.

**References**


44. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.


79. Huang V, Cohen A, Underwood P, McCarty B, Jalali N. The diagnostic utility of musculoskeletal ultrasound in evaluat-
Ultrasound in bone fracture diagnosis – a comparative meta-analysis and systematic review

10 Eric Matschner et al


