

Diagnostic accuracy of clinical tests to rule out elbow fracture: a systematic review

Giorgio Breda^{1,2}, Gianluca De Marco¹, Pierfranco Cesaraccio², Paolo Pillastrini¹

¹Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum University of Bologna, Bologna, Italy

²Department of Neurosciences (DNS), University of Padua, Padua, Italy

Elbow traumas represent a relatively common condition in clinical practice. However, there is a lack of evidence regarding the most accurate tests for screening these potentially serious conditions and excluding elbow fractures. The purpose of this investigation was to analyze the literature concerning the diagnostic accuracy of clinical tests for the detection or exclusion of suspected elbow fractures. A systematic review was performed using the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines. Literature databases including PubMed, Cumulative Index to Nursing and Allied Health Literature, Diagnostic Test Accuracy, Cochrane Library, the Web of Science, and ScienceDirect were searched for diagnostic accuracy studies of subjects with suspected traumatic elbow fracture investigating clinical tests compared to imaging reference tests. The risk of bias in each study was assessed independently by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies 2 checklist. Twelve studies (4,485 patients) were included. Three different types of index tests were extracted. In adults, these tests were very sensitive, with values up to 98.6% (95% confidence interval [CI], 95.0%–99.8%). The specificity was very variable, ranging from 24.0% (95% CI, 19.0%–30.0%) to 69.4% (95% CI, 57.3%–79.5%). The applicability of these tests was very high, while overall studies showed a medium risk of bias. Elbow full range of motion test, elbow extension test, and elbow extension and point tenderness test appear to be useful in the presence of a negative test to exclude fracture in a majority of cases. The specificity of all tests, however, does not allow us to draw useful conclusions because there was a great variability of results obtained.

Keywords: Physical examination; Sensitivity and specificity; Elbow joint; Fractures bone

INTRODUCTION

Elbow fractures represent 7% of all body fractures [1]. Extra-articular fractures of the elbow are typical of childhood (60% of cases occur in children) [2,3], while articular fractures are more frequent in those >50 years of age and have a low incidence (0.09% of total fractures). Cases peak between the ages of 12–19 years, usually in boys, and those aged ≥80 years, characteristically in women. In young adults, the fractures are typically caused by high-energy injuries, such as motor vehicular collisions, falls from a height, sports, industrial accidents, and firearms. In con-

trast, >60% of distal humeral fractures in the elderly occur from low-energy injuries, such as a fall from a standing height [4]. In general, for all joints, the most common signs of fracture are hemarthrosis, swelling, and loss of mobility. To the best of our knowledge, there have been no studies evaluating the diagnostic accuracy of X-ray in the identification of elbow fractures, although this type of instrumental investigation is used, in 95% of the cases, as an elite method for diagnosis. In situations of clinical doubt, or when the X-ray images are not clear, a diagnosis is made with computerized tomography or magnetic resonance imaging [5]. In recent years, studies have begun to consider the

Received: March 24, 2022 Revised: May 11, 2022 Accepted: May 23, 2022

Correspondence to: Paolo Pillastrini

Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum University of Bologna, via Massarenti 9, Bologna 40138, Italy
Tel: +39-51-2142496, Fax: +39-51-6362609, E-mail: paolo.pillastrini@unibo.it, ORCID: <https://orcid.org/0000-0002-8396-2250>

Copyright© 2023 Korean Shoulder and Elbow Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

diagnostic accuracy of ultrasonography compared to other instrumental investigations [6], reporting sensitivity and specificity values of >90% in children following elbow trauma. However, there is a low rate of positive radiography findings when assessing for suspected extremity fractures as evidenced by many studies, in which only 50% of patients with upper-extremity injuries [7] and 15% of patients with ankle injuries [8-10] had documented fractures on X-ray. This has led to the development and validation of clinical decision rules to safely reduce radiographic imaging for suspected lower-extremity fractures [11,12]. These rules are usually followed by clinicians to refer or not the patient toward an in-depth instrumental diagnostic exam in the emergency department. On the other hand, to date, there are still few clinical trials or clusters available that were built and designed to be able to “rule out” patients with suspected elbow fractures. In the last few years, several studies [13,14] have investigated the use of some elbow physical tests, which could help clinicians to make a correct decision in cases of suspected bone fractures without resorting to additional diagnostic instrumental investigations. To be clinically useful, a diagnostic test must be valid, reliable, safe, and simple. However, to our knowledge, no review has yet examined all the available literature on the diagnostic accuracy of clinical testing for detecting elbow fractures. The only study with a similar scope was the review by Joshi et al. [15] in which the authors wanted to provide a global overview of the different ways of diagnosing upper-extremity fractures, also through the use of the ultrasound. Therefore, the purpose of the present review was specifically to analyze the literature considering the diagnostic accuracy of these recently proposed clinical tests for the detection or exclusion of suspected elbow fractures.

METHODS

This systematic review was conducted and reported according to the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement guidelines. The 27-item PRISMA diagnostic test accuracy checklist provides specific guidance for reporting systematic reviews and renders the results from systematic reviews of diagnostic test accuracy studies more useful [16]. Institutional review board approval and informed consent from patients was not required for this systematic review.

Registration

The study was registered in April 2020 (CRD42020176511) with the International Prospective Register of Systematic Reviews, an

international database of prospectively registered systematic reviews in health and social care.

Eligibility Criteria

In the phase of planning a review protocol, the inclusion and exclusion criteria were defined a priori for the selection of the studies. Two reviewers (GDM and PC) applied all criteria independently to the full text of articles that passed the initial screening phase. When a disagreement arose, a third author (GB) was consulted to discuss and solve the conflict. Studies, to be eligible, had to satisfy the following inclusion criteria: (1) enrolled patients presented with a suspected elbow fracture, defined as disruption of the bone tissue of the proximal epiphysis of radius and/or proximal epiphysis of the ulna and/or distal epiphysis of the humerus following a trauma; (2) the study investigated diagnostic accuracy without any limitation relating to the language of the report or the publication date; (3) the results of ≥ 1 clinical tests were compared with an acceptable reference standard (X-ray, computed tomography, magnetic resonance imaging); and (4) the study had to report measures of diagnostic accuracy (sensitivity, specificity, positive likelihood ratio, negative likelihood ratio) or to allow their calculation. Articles were excluded if (1) the index test was performed with complex, specific, and technological equipment not easily applicable in daily clinical practice (e.g., dynamometers, electrogoniometers); (2) the index test was performed on a cadaver, under anesthesia, or during or after surgery; and (3) the study included patients who had severe polytrauma and/or temporary loss of consciousness that did not allow investigators to execute the index test.

Information Sources

A literature search was conducted for diagnostic accuracy studies in PubMed, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, Diagnostic Test Accuracy (DiTA), ScienceDirect, the Web of Science, and the Cochrane Library databases by two reviewers acting separately. The search strategy (Supplementary Material 1) has been shared by all reviewers. We also searched for papers in Open SIGLE, Google, and Google Scholar to find grey literature [17]. In the first step, duplicate articles were excluded with the use of the software Zotero (Corporation for Digital Scholarship, Vienna, VA, USA) [18]; then, the articles, consistent with the review question, were screened by two authors (GDM and PC) independently, by reading titles, abstracts, and (if necessary) full-text versions. In the second step, two reviewers (GDM and PC) independently selected studies based on agreed criteria. When a disagreement arose, a third author (GB) was consulted to discuss and solve the conflict.

Risk of Bias

All studies included in the review were evaluated and scored with the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool [19] ([Supplementary Material 2](#)) by two reviewers (GDM and PC) independently. A third author (GB) intervened in the case of any disagreement regarding the score assigned to the individual items. The QUADAS-2 tool is designed to assess the quality of primary diagnostic accuracy studies and consists of four key domains that discuss patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of the risk of bias, and the first three domains are also assessed in terms of concerns about applicability. Signaling questions are included to help judge the risk of bias; these questions flag aspects of study design related to the potential for bias and aim to help reviewers judge the risk of bias. The QUADAS-2 tool can be tailored to each review by adding or omitting signaling questions. In each item, the risk of bias could be classified as “low,” “high,” or “unclear.” If a study is judged to have a “high” or “unclear” risk of bias in ≥ 1 domains, then it may be judged “at risk of bias” or as having “concerns regarding applicability.”

Data Extraction

Two reviewers (GDM and PC) independently extracted information and data regarding all included studies, including first author, publication year, study type, study population, and setting; index test and reference test; diagnostic criteria; prevalence; and the number of true positives, false positives, false negatives, and true negatives for recalculation or calculation, when not provided, of sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (+LR), and negative likelihood ratio (–LR). Extracted data were processed by two reviewers (GDM and PC) using Microsoft Excel (Microsoft Corp., Redmond, WA, USA).

RESULTS

Through research on biomedical databases from September 2020 to May 2021, 1,506 articles were obtained ([Supplementary Material 3](#)), including 899 studies in Medline, 62 in CINAHL, 46 in Scopus, 11 DiTA, 311 in the Web of Science, 55 in the Cochrane Library, and 122 in ScienceDirect in addition to an American degree thesis obtained through a search of the grey literature. After removing duplicates, 901 articles and a thesis were obtained to be reviewed independently by the two authors. After titles and abstracts were read, we excluded 888 articles and 1 thesis because they did not respect the inclusion criteria previously defined. Thirteen articles remained after this last step, and their full texts

were read. At this stage, only one article [20] was considered irrelevant because the index test consisted of a clinical test and an X-ray and therefore required complex instrumentation. In the end, 12 diagnostic accuracy studies ([Fig. 1](#)) deemed suitable for the purpose of this systematic review were evaluated using the QUADAS-2 scale.

Article and Clinical Characteristics

Five studies [13,14,21-23] (1,050 total patients) compared the elbow’s range of motion (ROM) as an index test versus X-ray as a reference standard, four studies [24-27] (654 patients with fracture, 2,024 total) compared the ROM in elbow extension as an index test to X-ray as a reference standard, and three studies [28-30] (1,411 total patients) compared clinical clusters that included the evaluation of ROM in elbow extension and tenderness points to X-ray as a reference standard. In four studies [14,23,24,27], the values of +LR and –LR were also present, while, in eight studies [14,21-25,27,29], VPP and VPN values were reported; in other cases, the data were extrapolated by the researchers (GDM and PC) using specific online calculators and added to the summary table of diagnostic accuracy values. Two studies [22,29] investigated this condition in children, and 10 studies investigated the same in children and in adults [13,14,21,24-26,28,30]. The total number of fractures was not calculated in relation to the global sample divided by subgroups of different index tests. In cases where the +LR and –LR were indirectly calculated from the data present in the studies, it was not possible to associate them with relative confidence intervals. No articles were found that investigated index tests that included resistive muscle tests or specific special tests. The main features of the studies are collected in [Tables 1](#) and [2](#). In all studies, the tests were done by doctors or nurses. Most of the included studies did not provide training for learning how to perform the clinical tests, considering them easy to perform. Full results of the included studies are available in [Supplementary Material 4](#). As can be seen from [Table 3](#), all studies are characterized by a low risk of bias in the domains related to applicability; the study with the lowest risk of bias appears to be that by Amiri et al. [21].

DISCUSSION

Diagnostic Accuracy of the Studies

[Tables 2](#) and [3](#) summarize the main data concerning the diagnostic accuracy of the selected studies. Its rapid analysis shows that, overall, the studies have high sensitivity values that, in several cases, were close to 100%, making these tests useful to exclude possible elbow fractures. In contrast, the specificity was much

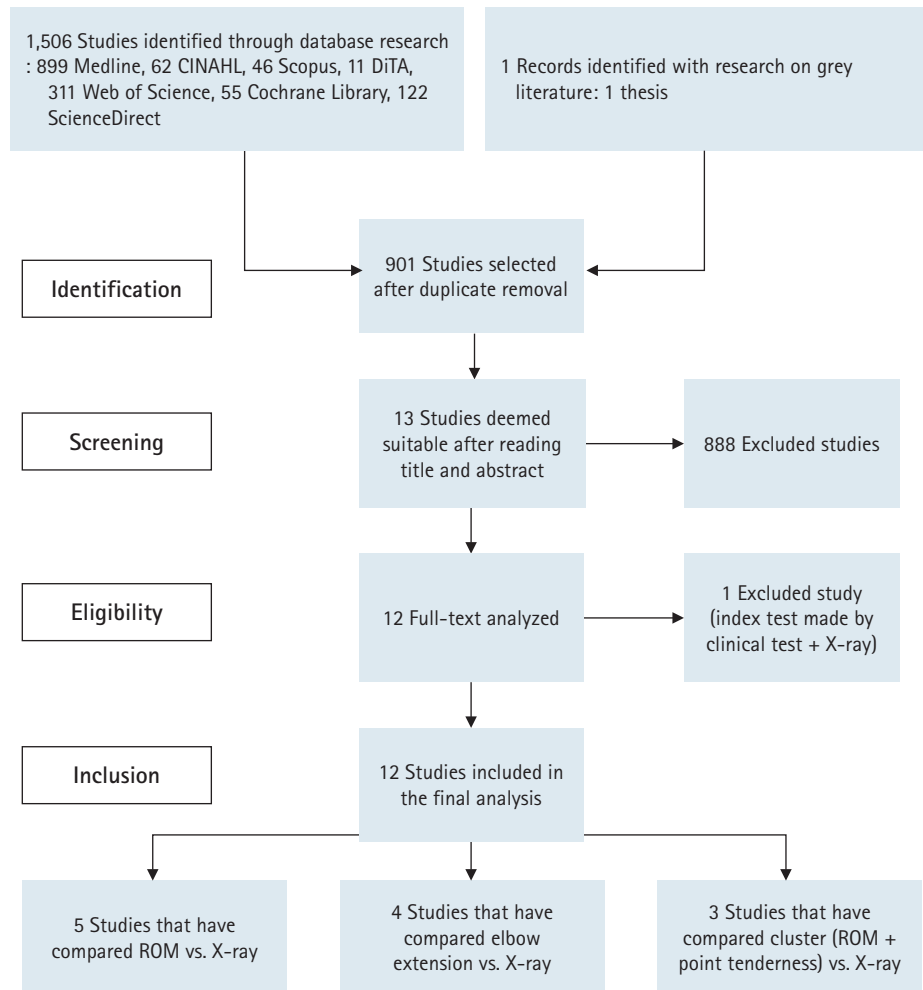


Fig. 1. Flowchart and study selection process. CINAHL: Cumulative Index to Nursing and Allied Health Literature, DiTA: Diagnostic Test Accuracy, ROM: range of motion.

lower, with a very wide range of variability (24%–69.4%).

Comparing our results to those obtained by Joshi et al. [15], it was possible to find that, in both studies, the mobility tests obtained very similar values, although more studies than Joshi et al.'s review [15] review were added to our study and the reviews had different objectives. In particular, in the previous review, the average sensitivity was 0.84 (0.82–0.87) and the specificity was 0.57 (0.54–0.59) for the elbow extension test.

Overall, three main types of index tests could be identified. The first evaluates the elbow extension test, the second evaluates the ROM in different directions of the elbow, and the third involves a cluster formed by an extension test and palpation test. In each test, the reference standard used was the X-ray.

Elbow Extension Test

Four studies [24–27] analyze the diagnostic accuracy of the elbow extension test. Overall sensitivity showed values of >90% with a

maximum value of 97.3% (84.6%–99.9%) in Docherty et al.'s study [25]. In contrast, in these four studies [24–27] specificity and +LR displayed much lower values (48.5%–69.4%), thus reducing the test's ability to correctly classify subjects with a positive test.

Elbow ROM Test

Among the five studies [13,14,21,22] included in this subgroup, very high sensitivity values emerged, reaching 100% in the studies by Darracq et al. [13], Vinson et al. [14]. Conversely, specificity achieved high values only in the studies of Amiri et al. [21] and Darracq et al. [13] (88%–97%). In Lennon et al.'s study [23], diagnostic accuracy values emerged that strongly disagree with those of the rest of the studies, showing a high specificity and low sensitivity of the elbow ROM test. This is, however, due to the outcome being studied, which was the detection of the absence of fracture in this study but the presence of fracture in all others.

Table 1. Information relating to studies and values of diagnostic accuracy of clinical test in adults

Study	Test used	Sample size (n)	Prevalence	Sensitivity (95% CI, %)	Specificity (95% CI, %)	+LR (95% CI)	-LR (95% CI)
Appelboom et al. (2008) [24]	Extension	960	0.30	98.4 (96.3–99.5)	47.7 (43.7–51.6)	1.88 (1.75–2.03)	0.03 (0.01–0.08)
Hawksworth et al. (1991) [26]	Extension	100	0.54	90.7 (79.7–96.9)	58.7 (43.2–73.0)	2.20 (1.54–3.13)	0.16 (0.07–0.38)
Arundel et al. (2014) [28]	Extension	348	NA	86.0 (81.8–89.3)	48.7 (43.4–54.1)	1.68 (1.45–1.94)	0.29 (0.18–0.46)
Darracq et al. (2008) [13]	Extension	113	0.47	100.0 (93.0–100.0)	100.0 (94.0–100.0)	Maximum	0.00 (NA)
Docherty et al. (2002) [25]	Extension	114	0.34	97.3 (84.6–99.9)	69.4 (57.3–79.5)	3.19 (2.24–4.53)	0.04 (0.01–0.26)
Jie et al. (2014) [30]	Extension	587	0.63	93.5 (88.7–96.7)	47.6 (40.2–55.0)	1.78 (1.55–2.06)	0.14 (0.08–0.25)
Lamprakis et al. (2007) [27]	Extension	70	0.34	91.7 (73.0–99.0)	60.9 (45.4–74.9)	2.34 (1.60–3.43)	0.14 (0.04–0.53)
Lennon et al. (2007) [23] ^{a)}	Extension	407	NA	54.1 (43.6–64.3)	92.4 (85.7–96.1)	7.11 (NA)	0.50 (NA)
Darracq et al. (2008) [13]	Flexion	113	0.47	64.0 (50.0–69.0)	100.0 (94.0–100.0)	Maximum	0.36 (NA)
Lennon et al. (2007) [23] ^{b)}	Flexion	407	NA	61.2 (50.5–70.8)	74.3 (65.2–81.7)	2.38 (NA)	0.52 (NA)
Darracq et al. (2008) [13]	Pronation	113	0.47	34.0 (22.0–48.0)	100.0 (94.0–100.0)	Maximum	0.66 (NA)
Darracq et al. (2008) [13]	Supination	113	0.47	43.0 (30.0–58.0)	97.0 (88.5–100.0)	14.33 (NA)	0.58 (NA)
Lennon et al. (2007) [23] ^{a)}	Pronosupination	407	NA	51.8 (41.3–62.1)	80.0 (71.4–86.5)	2.59 (NA)	0.60 (NA)
Darracq et al. (2008) [13]	Elbow full ROM	113	0.47	100.0 (93.0–100.0)	97.0 (88.5–100.0)	33.33 (NA)	0.00 (NA)
Lennon et al. (2007) [23] ^{b)}	Elbow full ROM	407	NA	25.9 (17.8–36.1)	96.2 (90.6–98.5)	6.82 (NA)	0.77 (NA)
Amiri et al. (2012) [21]	Elbow full ROM	102	0.10	90.0 (55.5–99.8)	88.0 (79.6–93.9)	7.53 (4.17–13.60)	0.11 (0.02–0.73)
Vinson et al. (2016) [14]	Elbow full ROM	251	0.39	99.0 (94.5–100.0)	59.9 (51.6–67.7)	2.47 (2.03–3.00)	0.02 (0.00–0.12)
Arundel et al. (2014) [28]	Elbow extension+bruising+tenderness	348	NA	100.0 (97.0–100.0)	24.0 (19.0–30.0)	1.32 (NA)	0.00 (NA)
Jie et al. (2014) [30]	Elbow extension+point tenderness	587	0.63	98.6 (95.0–99.8)	11.1 (6.2–17.9)	1.11 (1.04–1.18)	0.13 (0.03–0.55)
Darracq et al. (2008) [13]	Point tenderness	113	0.47	100.0 (93.0–100.0)	67.0 (53.0–78.0)	3.03 (NA)	0.00 (NA)
Jie et al. (2014) [30]	Point tenderness	587	0.63	91.5 (85.6–95.5)	18.3 (11.9–26.1)	1.12 (1.02–1.23)	0.47 (0.24–0.90)

CI: confidence interval, +LR: positive likelihood ratio, -LR: negative likelihood ratio, NA: not applicable, ROM: range of motion.

^{a)}This study used an inverse diagnostic question respect to the others for which the sensitivity and specificity values are to be considered inverted for the purpose of the review question.

Table 2. Information relating to studies and values of diagnostic accuracy of clinical test for children

Study	Test used	Sample size (n)	Prevalence	Sensitivity (95% CI, %)	Specificity (95% CI, %)	+LR (95% CI)	-LR (95% CI)
Appelboom et al. (2008) [24]	Extension	780	0.30	94.6 (90.7–97.2)	49.5 (45.2–53.7)	1.87 (1.72–2.05)	0.11 (0.06–0.19)
Arundel et al. (2014) [28]	Extension	144	NA	78.0 (66.0–87.0)	58.0 (51.0–62.0)	1.86 (NA)	0.38 (NA)
Baker et al. (2011) [22]	Extension	177	0.60	80.2 (71.3–87.3)	64.8 (52.5–75.8)	2.28 (NA)	0.31 (NA)
Dubrovsky et al. (2015) [29]	Extension	322	0.55	82.5 (75.2–89.9)	47.2 (40.7–53.6)	1.56 (1.34–1.82)	0.37 (0.24–0.58)
Jie et al. (2014) [30]	Extension	587	0.63	83.3 (71.5–91.7)	37.6 (30.3–45.2)	1.33 (1.14–1.57)	0.44 (0.24–0.81)
Lamprakis et al. (2007) [27]	Extension	70	0.34	100.0 (15.8–100.0)	57.1 (18.4–90.1)	2.33 (0.99–5.49)	0.00 (0.00–0.00)
Lennon et al. (2007) [23] ^{a)}	Extension	407	NA	40.8 (30.4–52.0)	90.2 (80.2–95.4)	4.16 (NA)	0.66 (NA)
Baker et al. (2011) [22]	Flexion	177	0.60	88.7 (81.1–94.0)	45.1 (33.2–57.3)	1.62 (NA)	0.25 (NA)
Lennon et al. (2007) [23] ^{a)}	Flexion	407	NA	51.3 (40.3–62.2)	83.6 (72.4–90.8)	3.13 (NA)	0.58 (NA)
Lennon et al. (2007) [23] ^{a)}	Pronosupination	407	NA	50.0 (39.0–61.0)	78.7 (66.9–87.1)	2.35 (NA)	0.64 (NA)
Baker et al. (2011) [22]	Elbow full ROM	177	0.60	93.4 (86.9–97.3)	33.8 (23.0–46.0)	1.41 (1.19–1.68)	0.20 (0.09–0.43)
Lennon et al. (2007) [23] ^{a)}	Elbow full ROM	407	NA	15.8 (9.3–25.6)	100.0 (94.1–100.0)	Maximum	0.84 (NA)
Dubrovsky et al. (2015) [29]	Point tenderness	322	0.55	95.1 (91.0–99.3)	23.1 (17.7–28.6)	1.24 (1.14–1.35)	0.21 (0.09–0.51)
Jie et al. (2014) [30]	Point tenderness	587	0.63	93.6 (82.5–98.7)	16.3 (10.2–24.0)	1.12 (1.00–1.25)	0.39 (0.12–1.26)
Dubrovsky et al. (2015) [29]	Points tenderness+elbow extension	322	0.55	99.0 (97.1–100.0)	14.0 (9.5–18.5)	1.15 (1.09–1.22)	0.07 (0.01–0.50)
Jie et al. (2014) [30]	Points tenderness+elbow extension	587	0.63	97.9 (88.7–100.0)	5.7 (2.3–11.4)	1.04 (0.98–1.10)	0.37 (0.05–2.96)

CI: confidence interval, +LR: positive likelihood ratio, -LR: negative likelihood ratio, NA: not applicable, ROM: range of motion.

^{a)}This study used an inverse diagnostic question respect to the others for which the sensitivity and specificity values are to be considered inverted for the purpose of the review question.

Table 3. Bias risk assessment

Study	Risk of bias			Applicability concern			
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Appelboom et al. (2008) [24]	😊	😊	😞	😞	😊	😊	😊
Hawksworth et al. (1991) [26]	😞	😊	😊	😊	😊	😊	😊
Arundel et al. (2013) [28]	😊	😊	😞	😞	😊	😊	😊
Baker et al. (2011) [22]	😞	😊	😊	😊	😊	😊	😊
Darracq et al. (2008) [13]	😞	😊	😊	😊	😊	😊	😊
Docherty et al. (2002) [25]	😊	😊	😊	😞	😊	😊	😊
Dubrovsky et al. (2015) [29]	😞	😊	😞	😞	😊	😊	😊
Jie et al. (2014) [30]	😊	😊	?	😊	😊	😊	😊
Lamprakis et al. (2007) [27]	?	😊	😊	😊	😊	😊	😊
Lennon et al. (2007) [23]	😊	😊	?	😞	😊	😊	😊
Amiri et al. (2012) [21]	😊	😊	😊	😊	😊	😊	😊
Vinson et al. (2016) [14]	😞	😊	😊	😊	😊	😊	😊

Considering this aspect, the results are therefore in line with the rest of the studies.

Cluster of Elbow ROM and Tenderness Test

In the three studies [28-30] that conducted a mobility test associated with the palpation of some bone points of the elbow, the sensitivity was very high (>97% in all cases). Meanwhile, the specificity of contrast had very low values never exceeding 24%.

Adults and Children

In five studies [23-25,28,30] diagnostic accuracy values were calculated specifically in adults (aged >16 years). In all studies, a high sensitivity of the tests was found that reached a maximum value of 98.6%, while there was a wide range of specificity (24%–69.4%). In contrast, seven studies [2,23,24,27-30] analyzed tests in a specific population of subjects under the age of 14–16 years. In these studies, the test with a better sensitivity was the elbow extension test associated with the palpation of 5 specific points with values of 97.1%–100% [29]. The specificity was also very variable in this case, passing from 14% of the elbow extension and palpation cluster [29] to 64% of the elbow extension test [22].

Risk of Bias

The methodological quality of the studies analyzed was very good in relation to the applicability of the tests, while, in almost all studies, the presence of bias was assessed, in particular relating to the selection of the patient sample [13,14,22,26], or related to the timing of administration of the index test [24,28,29]. This may have partially influenced the obtained results. In children, it is possible to make a similar argument, with similar results, but the sample is not sufficiently numerous. All included studies were identified as cross-sectional, although three studies [24,28,29] conducted telephone follow-up to 7–10 days in patients with a negative index test and, only in the case of persistent symptoms, these same patients were recalled for X-ray. This means that these studies cannot be defined as real cross-section analyses, and it would be appropriate to consider the results based only on the first standard reference made. However, this has not been done as it was decided to fully analyze all the data from the 12 included studies so that a homogeneous analysis of the latter could be performed. The external applicability of these tests in an outpatient setting is very high: as we mentioned above, these tests are very simple to perform and easily reproducible, although they were performed in an emergency setting in the included studies.

Implications for Clinical Practice and Future Research

It was not possible to carry out a meta-analysis of the results obtained in this study as there were several studies with different and non-homogeneous characteristics, and only a descriptive analysis was performed. These tests seem to be useful to exclude the presence of fractures, especially in cases where they are negative, given the low number of false negatives found. As for specificity, in all three types of proposed index tests, discordant values were obtained, but the evaluation of the complete ROM, if positive, could lead us to think of a fracture following an important traumatic event. Among all movements of the elbow, the extension movement is that which, as values of diagnostic accuracy, is more reliable for assessing possible elbow fractures, compared to the flexion and pronosupination movements, but also with respect to the tenderness points.

Study Limitations

The limitations of this review are characterized by the relatively few studies and their heterogeneity, which, as in many revisions, are different for the type of sample, characteristics of administration of the index tests, and comparison with the respective reference tests. Furthermore, the lack of some data in the individual studies excluded the possibility of making definitive conclusions from all 12 selected studies.

In the analysis and interpretation of our results, we must also consider that radiographs are not a perfect reference standard. As indicated above, X-ray is not a sensitive and 100% specific exam for elbow fractures. In a minority of cases, but still a significant number given the detection of the condition, it may happen that occult fractures are detected only by computerized axial tomography or magnetic resonance. In some of these studies, non-fractures identified by the X-ray may not correspond to the actual clinical condition. This is an important limitation of all the studies analyzed, as the search for a gold standard for elbow fractures has not yet been sufficiently investigated.

Finally, an important aspect to consider is age used to define the pediatric subgroup. It was not uniform in the studies included but had a certain range of variability. The comparison between the data obtained in these studies may therefore not be as accurate.

CONCLUSIONS

Considering the results of the studies with the lowest number of biases, the elbow mobility tests appear to be useful, in case of a negative test, to rule out an elbow fracture. The specificity of all the index tests proposed at the moment does not allow us to

draw useful conclusions. Further studies are needed to investigate more deeply the diagnostic accuracy of these clinical tests and to confirm the results of this review.

NOTES

ORCID

Giorgio Breda <https://orcid.org/0000-0002-7076-2380>
 Gianluca De Marco <https://orcid.org/0000-0002-7215-7467>
 Pierfranco Cesaraccio <https://orcid.org/0000-0002-6789-7321>
 Paolo Pillastrini <https://orcid.org/0000-0002-8396-2250>

Author contributions

Conceptualization: GB, PC, PP. Data curation: GB, GDM. Investigation: GDM. Methodology: GB, PC, PP. Supervision: PP. Writing – original draft: GB, GDM. Writing – review & editing: PC.

Conflict of interest

None.

Funding

None.

Data availability

None.

Acknowledgments

None.

Supplementary materials

Supplementary materials can be found via <https://doi.org/10.5397/cise.2022.00948>.

REFERENCES

- Macdermid JC, Vincent JI, Kieffer L, Kieffer A, Demaiter J, Macintosh S. A survey of practice patterns for rehabilitation post elbow fracture. *Open Orthop J* 2012;6:429-39.
- Barr LV. Paediatric supracondylar humeral fractures: epidemiology, mechanisms and incidence during school holidays. *J Child Orthop* 2014;8:167-70.
- Houshian S, Mehdi B, Larsen MS. The epidemiology of elbow fracture in children: analysis of 355 fractures, with special reference to supracondylar humerus fractures. *J Orthop Sci* 2001;6:312-5.
- Robinson CM, Hill RM, Jacobs N, Dall G, Court-Brown CM. Adult distal humeral metaphyseal fractures: epidemiology and results of treatment. *J Orthop Trauma* 2003;17:38-47.
- Mellema JJ, Janssen SJ, Guitton TG, Ring D. Quantitative 3-dimensional computed tomography measurements of coronoid fractures. *J Hand Surg Am* 2015;40:526-33.
- Rabiner JE, Khine H, Avner JR, Friedman LM, Tsung JW. Accuracy of point-of-care ultrasonography for diagnosis of elbow fractures in children. *Ann Emerg Med* 2013;61:9-17.
- Bentohami A, Walenkamp MM, Slaar A, et al. Amsterdam wrist rules: a clinical decision aid. *BMC Musculoskelet Disord* 2011;12:238.
- Barelds I, Krijnen WP, van de Leur JB, van der Schans CP, Goddard RJ. Diagnostic accuracy of clinical decision rules to exclude fractures in acute ankle injuries: systematic review and meta-analysis. *J Emerg Med* 2017;53:353-68.
- Heyworth J. Ottawa ankle rules for the injured ankle. *Br J Sports Med* 2003;37:194.
- Lau LH, Kerr D, Law I, Ritchie P. Nurse practitioners treating ankle and foot injuries using the Ottawa Ankle Rules: a comparative study in the emergency department. *Australas Emerg Nurs J* 2013;16:110-5.
- Eggl S, Sclabas GM, Eggl S, Zimmermann H, Exadaktylos AK. The Bernese ankle rules: a fast, reliable test after low-energy, supination-type malleolar and midfoot trauma. *J Trauma* 2005;59:1268-71.
- Seaberg DC, Yealy DM, Lukens T, Auble T, Mathias S. Multi-center comparison of two clinical decision rules for the use of radiography in acute, high-risk knee injuries. *Ann Emerg Med* 1998;32:8-13.
- Darracq MA, Vinson DR, Panacek EA. Preservation of active range of motion after acute elbow trauma predicts absence of elbow fracture. *Am J Emerg Med* 2008;26:779-82.
- Vinson DR, Kann GS, Gaona SD, Panacek EA. Performance of the 4-way range of motion test for radiographic injuries after blunt elbow trauma. *Am J Emerg Med* 2016;34:235-9.
- Joshi N, Lira A, Mehta N, Paladino L, Sinert R. Diagnostic accuracy of history, physical examination, and bedside ultrasound for diagnosis of extremity fractures in the emergency department: a systematic review. *Acad Emerg Med* 2013;20:1-15.
- McInnes MD, 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;319:388-96.
- De Vet HC, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Chapter 7: searching for studies. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test*. London: The Cochrane Collaboration; 2008.
- Coar JT, Sewell JP. Zotero: harnessing the power of a personal

- bibliographic manager. *Nurse Educ* 2010;35:205-7.
19. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155:529-36.
 20. Acar K, Aksay E, Oray D, Ilmamoğlu T, Gunay E. Utility of computed tomography in elbow trauma patients with normal X-ray study and positive elbow extension test. *J Emerg Med* 2016;50:444-8.
 21. Amiri H, Shams Vahdati S, Fekri S, Zadegan SA, Shokoohi H, Rahimi-Movaghar V, et al. Does preservation of active range of motion after acute elbow injury rule out the need for radiography. *Ulus Travma Acil Cerrahi Derg* 2012; 18:479-82.
 22. Baker M, Borland M. Range of elbow movement as a predictor of bony injury in children. *Emerg Med J* 2011; 28:666-9.
 23. Lennon RI, Riyat MS, Hilliam R, Anathkrishnan G, Alderson G. Can a normal range of elbow movement predict a normal elbow x ray. *Emerg Med J* 2007;24:86-8.
 24. Appelboam A, Reuben AD, Bengler JR, et al. Elbow extension test to rule out elbow fracture: multicentre, prospective validation and observational study of diagnostic accuracy in adults and children. *BMJ* 2008;337:a2428.
 25. Docherty MA, Schwab RA, Ma OJ. Can elbow extension be used as a test of clinically significant injury. *South Med J* 2002; 95:539-41.
 26. Hawksworth CR, Freeland P. Inability to fully extend the injured elbow: an indicator of significant injury. *Arch Emerg Med* 1991;8:253-6.
 27. Lamprakis A, Vlasik K, Siampou E, Grammatikopoulos I, Lionis C. Can elbow-extension test be used as an alternative to radiographs in primary care. *Eur J Gen Pract* 2007; 13:221-4.
 28. Arundel D, Williams P, Townend W. Deriving the East Riding Elbow Rule (ER2): a maximally sensitive decision tool for elbow injury. *Emerg Med J* 2014; 31:380-3.
 29. Dubrovsky AS, Mok E, Lau SY, Al Humaidan M. Point tenderness at 1 of 5 locations and limited elbow extension identify significant injury in children with acute elbow trauma: a study of diagnostic accuracy. *Am J Emerg Med* 2015;33:229-33.
 30. Jie KE, van Dam LF, Verhagen TF, Hammacher ER. Extension test and ossal point tenderness cannot accurately exclude significant injury in acute elbow trauma. *Ann Emerg Med* 2014; 64:74-8.