

# Intra- and Interobserver Reliability of the Eaton Classification for Trapeziometacarpal Arthritis

## A Systematic Review

Aaron J. Berger MD, PhD, Arash Momeni MD,  
Amy L. Ladd MD

Published online: 6 August 2013  
© The Association of Bone and Joint Surgeons® 2013

### Abstract

**Background** Trapeziometacarpal, or thumb carpometacarpal (CMC), arthritis is a common problem with a variety of treatment options. Although widely used, the Eaton radiographic staging system for CMC arthritis is of questionable clinical utility, as disease severity does not predictably correlate with symptoms or treatment recommendations. A possible reason for this is that the classification itself may not be reliable, but the literature on

this has not, to our knowledge, been systematically reviewed.

**Questions/purposes** We therefore performed a systematic review to determine the intra- and interobserver reliability of the Eaton staging system.

**Methods** We systematically reviewed English-language studies published between 1973 and 2013 to assess the degree of intra- and interobserver reliability of the Eaton classification for determining the stage of trapeziometacarpal joint arthritis and pantrapezial arthritis based on plain radiographic imaging. Search engines included: PubMed, Scopus®, and CINAHL. Four studies, which included a total of 163 patients, met our inclusion criteria and were evaluated. The level of evidence of the studies included in this analysis was determined using the Oxford Centre for Evidence Based Medicine Levels of Evidence Classification by two independent observers.

**Results** A limited number of studies have been performed to assess intra- and interobserver reliability of the Eaton classification system. The four studies included were determined to be Level 3b. These studies collectively indicate that the Eaton classification demonstrates poor to fair interobserver reliability (kappa values: 0.11–0.56) and fair to moderate intraobserver reliability (kappa values: 0.54–0.657).

**Conclusions** Review of the literature demonstrates that radiographs assist in the assessment of CMC joint disease, but there is not a reliable system for classification of disease severity. Currently, diagnosis and treatment of thumb CMC arthritis are based on the surgeon's qualitative assessment combining history, physical examination, and radiographic evaluation. Inconsistent agreement using the current common radiographic classification system suggests a need for better radiographic tools to quantify disease severity.

---

Each author certifies that he or she, or a member of his or her immediate family, has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

One of the authors (ALL) certifies that she, or a member of her immediate family, has received or may receive funding, during the study period, from The Ronald and Ann Williams Foundation Charitable Trust and the NIH (Bethesda, MD, USA) (Grants 1R01AR059185-01A1 and R43 EB003067-02A1).

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request.

---

A. J. Berger, A. Momeni  
Division of Plastic and Reconstructive Surgery, Stanford  
University Medical Center, Palo Alto, CA, USA

A. J. Berger  
e-mail: ab.mdphd@gmail.com

A. L. Ladd (✉)  
Robert A. Chase Hand & Upper Limb Center, Department  
of Orthopaedic Surgery, Stanford University Medical Center,  
770 Welch Road, Suite 400, Palo Alto, CA 94304, USA  
e-mail: alad@stanford.edu

## Introduction

The unique shape of the trapeziometacarpal, or thumb carpometacarpal (CMC), joint affords a variety of movements in multiple planes. Perhaps as a consequence of its shape and the demands placed on this joint, the CMC joint is prone to the development of osteoarthritis. Precise causes of thumb CMC osteoarthritis remain unclear. Many risk factors have been suggested, including factors such as advanced age, female sex, congenital joint malformation, and genetic predisposition and acquired or environmental factors such as medical conditions, acute trauma, and occupational requirements (long-term low-level stress, eg, repetitive thumb use) [10]. Factors characterizing CMC arthritis include changes in the shape of joint surfaces and deterioration of the ligamentous support system.

Osteoarthritis of the first CMC joint is a common, painful problem, with prevalence reported as high as 15% in adults older than 30 years; as much as 1/3 of all postmenopausal women may be afflicted (based on data from the *National Health and Nutrition Survey in 2006*) [1, 7, 11, 14]. The condition is diagnosed using history, physical examination, and plain radiographs.

Eaton and Littler [9] introduced the most commonly used classification system for staging the severity of CMC joint arthritis in 1973. Eaton and Glickel [8] adapted this to include degenerative changes to the scaphotrapezoid joint in 1987. The Eaton classification remains the most common system employed to determine the stage of the disease.

Reported shortcomings of this system include its limited intra- and interobserver reproducibility; high interobserver reliability is a critical quality measure for any classification system [12]. Perhaps as a consequence of inconsistent reproducibility of radiographic grading techniques, radiographic severity does not predictably correlate with symptoms or treatment recommendations [16]. To our knowledge, the literature on the Eaton classification system has not been systematically reviewed to determine the extent or quality of intra- and interobserver reliability and therefore no major conclusions have been developed on the subject.

Accordingly, we performed a systematic review to assess the degree of intra- and interobserver reliability of the Eaton radiographic classification of CMC joint arthritis.

## Search Criteria and Strategy

We performed a systematic review of the following electronic databases: PubMed, Scopus®, and CINAHL (the Cumulative Index to Nursing and Allied Health Literature). Search terms were developed utilizing services provided through Lane Library at Stanford University School of Medicine (Table 1). We supplemented the search for additional relevant studies

not identified by the electronic database from the references obtained in the reviewed studies. Search terms included: Eaton, thumb, carpometacarpal joints, metacarpophalangeal joint, hand joints, trapeziometacarpal joint, thumb carpometacarpal joint, CMC, basal joint, observer variation, classification, intrarater, interrater, interobserver, arthritis, osteoarthritis\*, spondylarthritis\*.

Two independent reviewers (AJB, AM) screened retrieved titles and abstracts for articles of potential interest. Disagreement was resolved by consensus. Inclusion criteria were original studies investigating the degree of intra- and interobserver reliability of the Eaton classification for CMC arthritis and published in English language up to and including March 2013. Only studies that used plain radiographs as the mode of imaging were included for analysis. Studies using alternate modes of imaging for assessment of intra- and/or interobserver reliability (eg, CT, MRI, or ultrasonography) were excluded, unless they also included plain radiographs. Furthermore, we excluded editorials, commentaries, and letters to the editor.

The electronic search yielded 274 citations with 32 duplicate studies (Fig. 1). We assessed the potential relevance of the 242 unique citations and excluded 235 based on review of title and abstract. Of the seven remaining articles, three [2, 14, 17] were excluded after full-text review, as they were ultimately deemed not relevant to the study question. This process left four articles [5, 12, 13, 15] that met criteria for final analysis, which together included analysis of 163 patients' radiographs, and these were critically reviewed. Each study included distributions of mild, moderate, and advanced cases. The observers in each report included hand surgeons, radiologists, and residents.

Parameters of interest included the level of experience of study participants reviewing the images, number of reviewers involved in the study, number of radiographs per subject/patient reviewed, additional and special views introduced, and limitations discussed in the respective studies. Furthermore, intra- and interobserver reliability calculations (kappa) were extracted. Intra- and interobserver reliability of the Eaton classification were measured using kappa values in all studies except for one [15], which only assessed interobserver reliability.

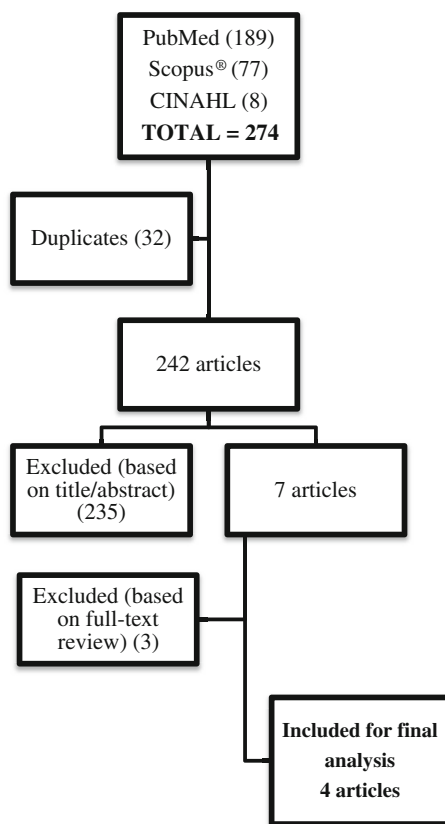
The level of evidence of the studies included in this analysis was determined using the Oxford Centre for Evidence Based Medicine Levels of Evidence Classification by two independent observers (AJB, AM) [4].

## Results

Intraobserver reliability, as measured by the kappa value, ranged from 0.54 to 0.657 (fair to moderate) (Table 2). In the studies reporting on this end point, it was measured at

**Table 1.** Search terms used to retrieve articles assessing intra- and interobserver reliability of the Eaton classification for trapeziometacarpal arthritis

Electronic database	Search term
PubMed	eaton*[tiab] AND (“thumb”[mesh] OR “Carpometacarpal Joints”[mesh] OR “Metacarpophalangeal Joint”[mesh] OR “Hand Joints”[mesh] AND “trapeziometacarpal joint” OR “thumb carpometacarpal”[tiab] OR “CMC”[tiab] OR thumb[tiab] OR “Basal joint”[tiab] OR joint*[tiab])  (“observer variation”[mesh] OR “classification”[sh] OR intrarater*[tiab] OR interrater*[tiab] OR interobserver*[tiab]) AND ((“thumb”[mesh] OR “Carpometacarpal Joints”[mesh] OR “Metacarpophalangeal Joint”[mesh] OR “trapeziometacarpal”[tiab] OR “carpometacarpal”[tiab] OR thumb*[tiab]) AND (arthritis[tiab] OR osteoarthritis*[tiab] OR spondylarthritis*[tiab] OR “arthritis”[mesh]))
Scopus®	(TITLE-ABS-KEY((interrater* OR intrarater* OR interobserver* OR observer*)) AND TITLE-ABS-KEY((thumb* OR Carpometacarpal OR metacarpophalangeal OR trapeziometacarpal OR carpometacarpal)) AND TITLE((arthritis OR osteoarthritis* OR spondylarthritis*)))
CINAHL	(thumb* OR Carpometacarpal OR metacarpophalangeal OR trapeziometacarpal OR carpometacarpal) AND eaton*

**Fig. 1** A flowchart demonstrates the identification of articles included for analysis.

two different time points ranging from 1 week [5, 13] to 6 weeks [12].

Interobserver reliability, as measured by the kappa value (both within and between groups, as well as being reported overall) ranged from 0.11 to 0.56 (poor to fair) (Table 2).

Based on the Oxford Centre for Evidence Based Medicine Levels of Evidence Classification, all four of the included studies were determined to be Level 3b.

## Discussion

Detailed staging of thumb arthritis was first reported in the 1970s. Burton [3] developed a four-stage disease classification system in 1973 (Table 3) that relies on clinical signs, patient symptoms, and radiographs, including involvement of adjacent joints and the thumb metacarpophalangeal joint. Concurrently, Eaton and Littler [9] introduced a purely radiographic staging system for first CMC osteoarthritis (Table 4) that did not include the qualitative clinical findings associated with the Burton classification. Perhaps because of the more objective radiographic nature of the classification, the Eaton classification has been considered useful in clinical practice and has gained widespread acceptance [15]. However, multiple studies demonstrate large variation in the utility of this descriptive classification; furthermore, absence of agreement exists for correlating recommended treatment to CMC radiographic severity [5, 12, 13, 15]. We therefore evaluated the degree of intra- and interobserver reliability of the Eaton system for staging CMC arthritis.

Among the limitations of this study are the small number of studies available for analysis, as well as variations within each study with respect to observer specialty and level of training; hand surgeons, radiologists, and orthopaedic surgery residents were among the most commonly used observer groups.

The Eaton staging system was intended to describe radiographic severity tied to clinically relevant information, yet a disparity exists between clinical symptoms and radiographic disease. The purpose of this study was to perform a systematic review of the available literature assessing reliability and thus a measure of the strength of this system as it relates to consistency. A question then remains whether the radiographic reliability, or lack thereof, contributes to the discrepancy between clinical and radiographic disease. Furthermore, this study does not examine the validity of the

**Table 2.** Studies of intra- and interobserver reliability of the Eaton classification

Study	Year	Observers	Number of patients	Intraobserver reliability ( $\kappa$ value)	Interobserver reliability ( $\kappa$ value)
Kubik and Lubahn [13]	2002	3 hand surgeons and 3 orthopaedic residents	40	0.657	0.529
Dela Rosa et al. [5]	2004	6 hand surgeons	40	0.59–0.61 (depending on number of views)	0.37–0.56 (depending on number of views)
Spaans et al. [15]	2011	5 radiologists and 8 hand surgeons	40	NA	0.496
Hansen et al. [12]	2013	2 hand surgeons, 2 residents, and 1 radiologist	43	0.54	0.11

NA = not available.

**Table 3.** Burton classification [3]

Stage	Characteristics
I	Pain Ligamentous laxity Dorsoradial subluxation of the trapeziometacarpal joint
II	Instability Chronic subluxation Radiographic degenerative changes
III	Involvement of the scaphotrapezoid joint or less commonly the trapezotrapezoid or trapeziometacarpal to the index finger
IV	Stage II or III with degenerative changes at the metacarpophalangeal joint

**Table 4.** Eaton classification [9, 15]

Stage	Characteristics
I	Synovitis phase, slight widening of the joint space (joint capsule distension), normal articular contours; less than $\frac{1}{3}$ subluxation in any projection
II	Significant capsular laxity, at least $\frac{1}{3}$ subluxation of the joint; small bone or calcific fragments < 2 mm in diameter are present, usually adjacent to the volar or dorsal facets of the trapezium
III	Greater than $\frac{1}{3}$ subluxation; fragments > 2 mm are present dorsally or volarly, usually in both locations; slight joint space narrowing
IV	Advanced degenerative changes; major subluxation and very narrow joint space, with cystic and sclerotic subchondral bone changes; the margins of the trapezium show lipping and osteophyte formation, and there is significant erosion of the dorsoradial facet of the trapezium

Eaton classification as it relates to clinical utility or severity. Specific features that could enhance the clinical utility of the classification as it relates to surgical planning or outcomes are beyond the scope of this examination; this would require an enhancement of the reliability of the Eaton classification as it is currently exists.

Intraobserver reliability of radiographic CMC arthritis staging was fair to moderate, and interobserver reliability was poor to fair. Even when familiarity with the classification system is reported, the kappa values have been poor to moderate [5, 12, 13, 15]. Increasing the number of views and adding three-dimensional views of the CMC joint only modestly improved interobserver reliability of the staging system in two of the studies we evaluated [5, 12]. One of the variables for examination in the Burton [3], Eaton [9], and modified Eaton [2, 8] classifications is subluxation [3] and extent of subluxation [8, 9, 15] of the first metacarpal on the trapezium. Recent studies have confirmed that subluxation is present in asymptomatic subjects, measured with precise stress radiographs [17] and CT [6]. Thus, using degree of subluxation as criteria for staging, rather than a supporting variable, suggests current Eaton stratification is an inaccurate measure of disease severity. A concern with the current system is that observers tend to form their own interpretation of the classification system. This is especially true as it relates to Stage I, which describes joint widening. This is rarely seen in the experience of the senior author (ALL), and a common modified Stage I among hand surgeons describes minimal joint narrowing. Furthermore, since subluxation is probably not relevant, the senior author and colleagues often dismiss the required degree of subluxation in staging advanced disease. This type of informal modification predicts poor interobserver reliability.

Current literature suggests that Eaton stage does not correlate well with clinical findings, and treatment depends on patient response to nonoperative measures, a concept that has been confirmed in a recent Cochrane review evaluating different surgical strategies for various stages of osteoarthritis. That review indicated that patients with Stage I osteoarthritis are likely to benefit from nonoperative interventions, and treatment of those with Stage II to IV osteoarthritis depends on the severity of patient symptoms and their functional demands [16].

Regardless of rater experience level, number of available views, and availability of CT images, radiographic staging of

CMC arthritis with the current standard, the Eaton classification, demonstrates intra- and interobserver agreement that is at best moderate and is generally even worse in the middle ranges of disease severity (Stages II and III). Additionally, the choice of treatment strategy tends to vary from surgeon to surgeon, and a single surgeon may recommend different therapies for patients with the same stage of osteoarthritis. This suggests that other qualitative factors, such as reported symptoms, physical examination, and surgeon preference, are more important for choosing treatment modalities than Eaton radiographic staging [15]. Variation in interpretation of the current staging system indicates the need for more accurate and reproducible radiographic measurements that quantify disease severity. Quantified radiographic reference standards may provide a preliminary step for reaching a consensus concerning the treatment of the common thumb CMC arthritis.

**Acknowledgments** The authors thank the librarians at Lane Library in Stanford University School of Medicine, especially Christopher Stave, for ensuring academic rigor and thorough search strategies for this systematic review.

## References

1. Armstrong AL, Hunter JB, Davis TR. The prevalence of degenerative arthritis of the base of the thumb in post-menopausal women. *J Hand Surg Br.* 1994;19:340–341.
2. Belt E, Lehtivuori J, Kaarela K, Kauppi M, Lehtinen J, Lehto MU. Larsen grades in evaluating the first carpometacarpal joint. *Scand J Rheumatol.* 1999;28:305–307.
3. Burton RI. Basal joint arthrosis of the thumb. *Orthop Clin North Am.* 1973;4:331–338.
4. Centre for Evidence Based Medicine. Levels of Evidence 1. 2013. Available at: <http://www.cebm.net/?O=1025>. Accessed June 11, 2013.
5. Dela Rosa TL, Vance MC, Stern PJ. Radiographic optimization of the Eaton classification. *J Hand Surg Br.* 2004;29:173–177.
6. de Raedt S, Stilling M, van de Giessen M, Streekstra GJ, Vos FM, Hansen TB. A three-dimensional analysis of osteoarthritic changes in the thumb carpometacarpal joint. *J Hand Surg Eur Vol.* 2012 November 7 [Epub ahead of print].
7. Dillon CF, Hirsch R, Rasch EK, Gu Q. Symptomatic hand osteoarthritis in the United States: prevalence and functional impairment estimates from the third U.S. National Health and Nutrition Examination Survey, 1991–1994. *Am J Phys Med Rehabil.* 2007;86:12–21.
8. Eaton RG, Glickel SZ. Trapeziometacarpal osteoarthritis: staging as a rationale for treatment. *Hand Clin.* 1987;3:455–471.
9. Eaton RG, Littler JW. Ligament reconstruction for the painful thumb carpometacarpal joint. *J Bone Joint Surg Am.* 1973;55:1655–1666.
10. Fontana L, Neel S, Claise JM, Ughetto S, Catilina P. Osteoarthritis of the thumb carpometacarpal joint in women and occupational risk factors: a case-control study. *J Hand Surg Am.* 2007;32:459–465.
11. Haara MM, Heliövaara M, Kroger H, Arokoski JP, Manninen P, Karkkainen A, Knekt P, Impivaara O, Aromaa A. Osteoarthritis in the carpometacarpal joint of the thumb: prevalence and associations with disability and mortality. *J Bone Joint Surg Am.* 2004;86:1452–1457.
12. Hansen TB, Sørensen OG, Kirkeby L, Homilius M, Amstrup AL. Computed tomography improves intra-observer reliability, but not the inter-observer reliability of the Eaton-Glickel classification. *J Hand Surg Eur Vol.* 2013;38:187–191.
13. Kubik NJ, Lubahn JD. Intrarater and interrater reliability of the Eaton classification of basal joint arthritis. *J Hand Surg Am.* 2002;27:882–885.
14. Sodha S, Ring D, Zurakowski D, Jupiter JB. Prevalence of osteoarthritis of the trapeziometacarpal joint. *J Bone Joint Surg Am.* 2005;87:2614–2618.
15. Spaans AJ, van Laarhoven CM, Schuurman AH, van Minnen LP. Interobserver agreement of the Eaton-Littler classification system and treatment strategy of thumb carpometacarpal joint osteoarthritis. *J Hand Surg Am.* 2011;36:1467–1470.
16. Wajon A, Carr E, Edmunds I, Ada L. Surgery for thumb (trapeziometacarpal joint) osteoarthritis. *Cochrane Database Syst Rev.* 2009:CD004631.
17. Wolf JM, Oren TW, Ferguson B, Williams A, Petersen B. The carpometacarpal stress view radiograph in the evaluation of trapeziometacarpal joint laxity. *J Hand Surg Am.* 2009;34:1402–1406.