How Reliable is the Radiographic Diagnosis of Mild Madelung Deformity?

Sebastian Farr, MD[1]   Thierry G. Guitton, MD, PhD[2]   David Ring, MD, PhD[3]
Science of Variation Group

[1] Department of Pediatric Orthopaedics and Adult Foot and Ankle Surgery, Orthopaedic Hospital Speising, Vienna, Austria
[2] Department of Plastic Surgery, University Medical Center Groningen, Groningen, The Netherlands
[3] Dell Medical School, The University of Texas at Austin, Austin, Texas

Address for correspondence Sebastian Farr, MD, Department of Pediatric Orthopaedics and Adult Foot and Ankle Surgery, Orthopaedic Hospital Speising, Speisingerstrasse 109, Vienna 1130, Austria (e-mail: sebastian.farr@oss.at).

Abstract

Background   Patients with Madelung deformity exhibit a spectrum of mild to severe deformity and distortion of wrist geometry. It may be difficult to reliably distinguish mild Madelung deformity from normal.

Purpose   This study thus tested the reliability of the diagnosis of mild Madelung deformity on a single posteroanterior (PA) radiograph.

Materials and Methods   An online survey was sent to hand and wrist surgeons of the Science of Variation Study Group for evaluation of 25 PA wrist radiographs comprising five adults with suspected mild Madelung deformity and 20 radiographs without any evident wrist pathology. Interobserver agreement was evaluated both via average percent agreement and Fleiss' kappa. To evaluate the relationship of rater characteristics and accuracy, a linear regression model was computed.

Results   The interobserver agreement among the 69 participating surgeons was low ($K = 0.12$). The overall sensitivity, specificity, and accuracy were 0.30, 0.86, and 0.75, respectively. The mean confidence was 7.4 ± 0.4 for mild Madelung and 7.8 ± 0.5 for normal ($p = 0.112$). The observers' confidence level was the only factor which had a mild but significant effect on the accuracy of the ratings.

Conclusion   The diagnosis of mild Madelung deformity on a single PA radiograph is unreliable.

Level of Evidence   The level of evidence is II, diagnostic study.

Keywords
► Madelung deformity
► ulnar tilt
► lunate fossa angle
► interobserver reliability
► wrist pain

Patients with Madelung deformity exhibit a spectrum of mild to severe deformity and distortion of wrist geometry.[1,2] McCarroll et al noted that some patients with mild Madelung deformity have values within the normal range for each of the four most commonly used indices—the ulnar tilt, lunate subsidence, palmar carpal displacement, and lunate fossa angle.[3] Hence, it may be difficult to reliably distinguish mild Madelung deformity from normal.

Given the high prevalence of nonspecific wrist pain, it is important that mild Madelung deformity may not be subjected to overtreatment. The continuum of Madelung disease might cause some symptoms that are generally well adapted. In other words, the influence of mild pathophysiology on symptoms is likely overwhelmed by the influence of resiliency. In this setting, addressing the pathophysiology is not likely to be as helpful as addressing the resiliency. If the diagnosis of mild Madelung deformity on radiographs is unreliable, that may further decrease the relevance of these minor anatomical variations.

This study tested the primary null hypothesis that the diagnosis of mild Madelung deformity on a single posteroanterior (PA) radiograph among a research group of mostly

academic hand surgeons is unreliable. Our secondary null hypotheses were:

1. There is a low sensitivity and specificity for the detection of mild Madelung deformity with regards to the radiographs presented,
   and
2. There is no difference in confidence about diagnosis of mild Madelung deformity compared with diagnosis of normal,
   and
3. There are no factors associated with reliability of diagnosis of mild Madelung deformity.

Materials and Methods

Ethical approval was waived by the local ethics committee since this article does not contain any studies with human or animal subjects. An online survey (Survey Monkey, CA) was sent to a subset of hand and wrist surgeons of the Science of Variation Study Group (SOVG)—a group of orthopaedic, trauma, and plastic surgeons that studies the variation. The participants were asked to evaluate 25 PA wrist radiographs of adults with suspected mild Madelung deformity (5 patients) and radiographs without any evident wrist pathology or symptoms (20 radiographs). The five radiographs with suspected mild Madelung deformity were chosen based on established measurements of radiographic wrist deformity for this particular entity.[4] As mentioned before, there exists an overlap with normal wrists and thus the five cases included were within this transition zone (ulnar tilt, 20°–30°; lunate fossa angle, 20°–30°). Inclusion was confirmed by expert opinion by the first and senior authors of this study. In addition, a consecutive series of normal wrist PA radiographs was retrieved by searching the first author's hospital radiology database. Only patients without any reported wrist symptoms, in whom radiographs were obtained for reasons of comparison with the contralateral side, were eligible for inclusion. No lateral wrist radiographs were evaluated for this study.

The participants were asked the following two questions: (1) Do you think that this patient has a mild Madelung deformity? (Yes/No). (2) Please indicate the degree of confidence you had in rating this radiograph (scale, 1–10). The invitations were sent in February 2017 and a reminder was sent after 4 weeks. No measurements or remarks could be made on the actual radiographs of the survey. A total of 268 invitations were sent to the group members, of whom 71 participants (26%) opened the survey. This response rate, however, is not precise since some surgeons included in the mailing list may not treat congenital disorders, some of the emails may no longer be active, and nonparticipating surgeons have not been deleted from the email list. Two participants with incomplete responses were eliminated from further analysis leaving 69 complete responses (►Table 1).

Statistical Analysis

To summarize the basic features of the data, descriptive statistics were calculated. Depending on the scaling of the

Table 1 Demographic data of participating hand and wrist surgeons ($N = 69$)[a]

	N (%)
Sex	
Male	66 (95.7)
Female	3 (4.3)
Area of practice	
Asia	1 (1.5)
Europe	13 (18.8)
North America	49 (71.0)
Other	6 (8.7)
Years in practice	
< 5 y	20 (29.0)
5–10 y	14 (20.3)
10–20 y	25 (36.2)
20–30 y	10 (14.5)
Supervising function	
Yes	61 (88.4)
No	8 (11.6)

[a]Surgeons who fully completed the survey.

variables, means and standard deviations (SD), medians, and interquartile ranges (IQR), or percentages were computed. Interobserver agreement was evaluated both via average percent agreement and Fleiss' kappa.[5] The kappa coefficient is a measure of chance corrected agreement for nominal data among multiple raters. The values were interpreted using the guidelines provided by Landis and Koch, where 0.01 to 0.2 indicate slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and 0.81 to 1.00 almost perfect agreement.[6] Average percent agreement between groups (Madelung yes/no) as well as differences in confidence was tested for significance using t-tests for independent samples. Standard formulas were used to compute accuracy, sensitivity, and specificity. We used the prevalence of mild Madelung in this study (20%) to calculate the positive and negative predictive values according to Bayes' theorem. To evaluate the relationship of rater characteristics and accuracy, a linear regression model was computed. Place of work, years of independent practice, supervision of surgical trainees, and the mean confidence in the ratings were entered as factors. Overall fit was tested by means of the R^2 statistic. The independent effect of each factor was expressed via beta coefficients. The level of significance was set to 5%.

Results

There was slight interobserver agreement ($K = 0.12$) on the diagnosis of mild Madelung deformity among the participating surgeons (►Table 2).

Table 2 Rate of agreement among surgeons and accordance of ratings with radiographs (mild Madelung versus normal)

	Fleiss' kappa (95% CI)	Average pairwise agreement (%)	Average accuracy
All radiographs	0.12 (0.04–0.20)	74.74	0.75 (0.26)
Madelung: Yes	0.08 (−0.01 to 0.17)	61.47	0.30 (0.16)
Madelung: No	0.10 (−0.00 to 0.21)	78.06	0.86 (0.12)

Abbreviation: CI, confidence interval.

The overall sensitivity, specificity, and accuracy were 0.30, 0.86, and 0.75, respectively. Positive and negative predictive values (PPV and NPV) were 0.38 and 0.84, respectively.

The average confidence in diagnosis was 7.7 ± 0.5 on a scale of 1 to 10. The mean confidence was 7.4 ± 0.4 for mild Madelung and 7.8 ± 0.5 ($p = 0.112$) for normal. Overall, a mean value of 4.3 ± 3.9 radiographs was diagnosed with mild Madelung deformity; however, 15 observers stated that none of the 25 radiographs had Madelung deformity (►**Table 3**).

Regression analysis shows poor model fit ($R^2 = 0.111$). The only factor that could significantly contribute to the model was rater confidence (►**Table 4**).

Discussion

Madelung deformity can be very mild and unrecognized by surgeons and radiologists. The prevalence of mild Madelung deformity is uncertain. It is not clear if mild Madelung deformity causes symptoms or measurable impairment. Given the prevalence of nonspecific pain in wrists with no identifiable pathophysiology, combined with the fact that wrists with more obvious objective pathophysiology (e.g.,

Table 3 Number of positive Madelung ratings among observer group

	N (%)
None	15 (21.7)
4	8 (11.6)
1	7 (10.1)
5	6 (8.7)
2	5 (7.2)
3	5 (7.2)
8	5 (7.2)
6	4 (5.8)
11	4 (5.8)
7	3 (4.3)
10	3 (4.3)
9	2 (2.9)
13	1 (1.4)
16	1 (1.4)
Total	69 (100.0)

Table 4 Predictors of accuracy via multiple linear regression among the raters

Predictor	Standardized coefficient beta	t	p-Value
Rater confidence	0.352	3.077	0.003
Years in independent practice	0.047	0.383	0.703
Supervision of trainees	−0.008	−0.071	0.944
Place of practice	−0.171	−1.425	0.159

wrist arthritis) often produce few symptoms or limitations, we expect mild Madelung deformity to be an incidental finding that does not benefit from specific treatment. This supportive approach would be particularly appealing if the diagnosis of mild Madelung deformity is unreliable. Hence, this study measured the reliability of radiographic diagnosis of mild Madelung deformity in a sizable cohort of hand and wrist surgeons.

Our results indicate that the diagnosis of mild Madelung disease, defined by a borderline range of deformity according to McCarroll et al,[4] is unreliable. The overall agreement among the observers was low ($K = 0.12$) as was the sensitivity (0.30) for the entire cohort. However, with regards to correctly identify those without mild pathology, we observed more acceptable values (specificity, 0.86; NPV, 0.84).

Despite the low overall agreement, the raters were quite confident with their ratings as shown by a mean confidence level of more than 7 points out of 10 for all 69 surgeons. The confidence level was furthermore the only factor that had a mild but significant effect on the reliability/accuracy of the ratings. All other factors included in the model (place of work, years of independent practice, supervision of surgical trainees) did not yield an increased accuracy in regression analysis.

McCarroll et al previously highlighted that there is a relatively large span of deformity within this distinct entity.[4] The ulnar tilt was found to vary from 14° to 73°, lunate subsidence from –5 to 20 mm, and palmar carpal displacement from 10 to 36 mm. The lunate fossa angle has in contrast been found to show poorer correlation coefficients between observers, most likely due to the difficulty in establishing this measurement in a poorly defined area

Fig. 1 Two cases of a suspected mild Madelung deformity and a normal anteroposterior (AP) wrist radiograph are presented. Note the relatively steep radial plateau (**A**, **B**) and the slight sloping of the volar lunate facet (arrows) compared with the normal wrist radiograph (**C**).

(lunate fossa). The deformity ranges for Madelung wrists compared with normal wrists were for ulnar tilt 16° to 63° and 6° to 28°, for lunate subsidence −4.5 to 19 mm and −6 to 3 mm, for lunate fossa angle 27 to 85 and −7.5 to 29, and for palmar carpal displacement 9.5 to 35 mm and 4 to 19 mm, respectively.[3] Hence, there is a wide overlap between what seems to be pathologic and what is healthy (►**Fig. 1**). The same study group thereafter established cut-off values for unanimous agreement between four observers.[1] These were: ulnar tilt 33°, lunate subsidence 4 mm, lunate fossa angle 40°, and palmar carpal displacement 20 mm. However, only one of these four parameters had to be above the threshold to lead to unanimous agreement between observers. In total, 14 of 48 wrists did not reach the threshold value in any category.

The results of the previous and current studies indicate that there are currently no reliable radiographic measurement parameters to define and establish mild Madelung deformity. The so-called distal radius type comprises very mild types, which are definitely hard to recognize in standard radiographs given the minimal anatomic changes. Tuder et al defined them as a "forme fruste" mild Madelung type.[2] Skeletally immature adolescents with nonspecific wrist pain and mild Madelung deformity might have radiographs checked a second time after 6 to 12 months, but in other patients no additional monitoring or treatment is likely to be helpful. Until there is better evidence that mild Madelung deformity might cause symptoms amenable to surgery, adult patients with minor deformity are best treated supportively.

This study has several limitations. The diagnostic performance characteristics were based on a reference standard of consensus by two surgeons. Consequently, we consider the reliability data more reproducible than the diagnostic performance characteristics. Additionally, the prevalence of Madelung deformity in the population is likely much lower than in this study, which would consequently make the PPV lower and the NPV higher. Finally, we decided to exclude lateral radiographs from the survey since three out of four of the most reliable Madelung parameters according to previous studies are measured solely on PA images.[1,3,4] Inclusion of these, however, may have theoretically resulted in a higher percentage of interobserver agreement.

In summary, mild adult Madelung deformity is not reliably diagnosed on single PA radiographs. Mild Madelung disease probably goes undiagnosed in a majority of patients and never causes any harm. And wrist pain is common and often nonspecific, so a radiographic finding of mild Madelung deformity is likely incidental and not the cause of symptoms or limitations. Until there is good evidence that mild Madelung can be reliably and accurately diagnosed, and that we have an intervention that relieves symptoms better than sham surgery, it seems wise to err on the side of caution and treat mild Madelung deformity supportively.

Note
Ethical approval was waived by the local ethics committee since this article does not contain any studies with human or animal subjects.

Conflict of Interest
Sebastian Farr, MD, and Thierry G. Guitton, MD, PhD, declare that there is no conflict of interest. David Ring, MD, PhD, reports grants from Skeletal Dynamics, other from Wright Medical, personal fees from Biomet, personal fees from Acumed, other from Illuminos, personal fees from Deputy Editor for *Journal of Hand Surgery*, personal

References

1 McCarroll HR Jr, James MA, Newmeyer WL III, Manske PR. Madelung's deformity: diagnostic thresholds of radiographic measurements. J Hand Surg Am 2010;35(05):807–812
2 Tuder D, Frome B, Green DP. Radiographic spectrum of severity in Madelung's deformity. J Hand Surg Am 2008;33(06):900–904
3 McCarroll HR, James MA, Newmeyer WL III, Manske PR. Madelung's deformity: quantitative radiographic comparison with normal wrists. J Hand Surg Eur Vol 2008;33(05):632–635
4 McCarroll HR Jr, James MA, Newmeyer WL III, Molitor F, Manske PR. Madelung's deformity: quantitative assessment of x-ray deformity. J Hand Surg Am 2005;30(06):1211–1220
5 Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–382
6 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(01):159–174