



Validity, reliability and responsiveness of a goniometer watch to measure pure forearm rotation

Harte, D., Nevill, A., Ramsey, L., & Martin, S. (2023). Validity, reliability and responsiveness of a goniometer watch to measure pure forearm rotation. *British Journal of Hand Therapy*, 29(1), 1-11. Advance online publication. <https://doi.org/10.1177/17589983231211813>

[Link to publication record in Ulster University Research Portal](#)

Published in:
British Journal of Hand Therapy

Publication Status:
Published online: 01/11/2023

DOI:
[10.1177/17589983231211813](https://doi.org/10.1177/17589983231211813)

Document Version
Author Accepted version

General rights

The copyright and moral rights to the output are retained by the output author(s), unless otherwise stated by the document licence.

Unless otherwise stated, users are permitted to download a copy of the output for personal study or non-commercial research and are permitted to freely distribute the URL of the output. They are not permitted to alter, reproduce, distribute or make any commercial use of the output without obtaining the permission of the author(s).

If the document is licenced under Creative Commons, the rights of users of the documents can be found at <https://creativecommons.org/share-your-work/licenses/>.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk



Validity, reliability and responsiveness of a goniometer watch to measure pure forearm rotation

Harte, D., Nevill, A., Ramsey, L., & Martin, S. (2023). Validity, reliability and responsiveness of a goniometer watch to measure pure forearm rotation. *British Journal of Hand Therapy*, 1-11. Advance online publication. <https://doi.org/10.1177/17589983231211813>

[Link to publication record in Ulster University Research Portal](#)

Published in:
British Journal of Hand Therapy

Publication Status:
Published online: 01/11/2023

DOI:
[10.1177/17589983231211813](https://doi.org/10.1177/17589983231211813)

Document Version
Author Accepted version

General rights
Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.



Validity, Reliability and Responsiveness of a Goniometer Watch to Measure Pure Forearm Rotation

Journal:	<i>Hand Therapy</i>
Manuscript ID	HTH-23-0032.R2
Manuscript Type:	Original article
Keywords:	Range of motion, articular, Wrist fractures, Therapeutics, Supination, Outcome measures
Abstract:	<p>Introduction Innovative instruments have been designed to assess forearm rotation, an anatomically challenging motion to measure. This study assessed the validity, reliability and responsiveness of a novel goniometer watch (GoWatch) to measure pure forearm rotation. The modified finger goniometer (MFG) was the gold standard reference.</p> <p>Methods Forty participants with restricted forearm rotation were recruited. Two raters measured supination and pronation using the GoWatch and MFG before and after a hand therapy session. Repeated-measures ANOVA assessed for systematic bias with an apriori residual error of 10° deemed as acceptable. Secondary analysis used intraclass coefficients (ICCs) to categorise interrater reliability. Responsiveness of the GoWatch was calculated using Cohen's <i>d</i>.</p> <p>Results The GoWatch demonstrated acceptable agreement with the MFG with mean difference for supination 1.19° and pronation 0.20°. Interrater reliability was also within acceptable limits with mean difference GoWatch supination 4.43° and pronation 2.23°. Interrater reliability for GoWatch supination and pronation were categorized as excellent (ICC = 0.94) and good (ICC = 0.85) respectively. Systematic bias was observed in the instrument by rater interaction with rater 2 consistently underestimating GoWatch measures ($p < 0.05$). GoWatch supination showed small to medium responsiveness (Rater 1: $d = 0.14$; Rater 2: $d = 0.29$) and pronation very small to medium responsiveness (Rater 1: $d = 0.29$; Rater 2: $d = 0.05$).</p> <p>Discussion The GoWatch is a viable and user-friendly alternative to measure forearm rotation with demonstrable validity, interrater reliability and responsiveness. Further research is required to ensure systematic bias is not endemic when used across multiple raters.</p>

VALIDITY, RELIABILITY AND RESPONSIVENESS OF A GONIOMETER WATCH TO MEASURE PURE FOREARM ROTATION

ABSTRACT

Introduction

Innovative instruments have been designed to assess forearm rotation, an anatomically challenging motion to measure. This study assessed the concurrent validity, interrater reliability and responsiveness of a novel goniometer watch (GoWatch) to measure pure forearm rotation. The modified finger goniometer (MFG) was the criterion reference.

Methods

Forty participants with restricted forearm rotation were recruited. Two raters measured supination and pronation using the GoWatch and MFG before and after a hand therapy session. Repeated-measures ANOVA assessed for systematic bias with an *a priori* residual error of 5° deemed as acceptable. Secondary analysis used intraclass coefficients (ICCs) to categorise interrater reliability. Responsiveness of the GoWatch was calculated using Cohen's *d*.

Results

The GoWatch demonstrated acceptable agreement with the MFG with mean difference for supination 1.19° and pronation 0.20°. Interrater reliability was also within acceptable limits with mean difference GoWatch supination 4.43° and pronation 2.23°. Interrater reliability for GoWatch supination and pronation were categorized as excellent (ICC =0.94) and good (ICC=0.85) respectively. Systematic bias was observed in the instrument by rater interaction with rater 2 consistently underestimating GoWatch measures ($p<0.05$). GoWatch supination showed small to medium responsiveness (Rater 1: $d=0.14$; Rater 2: $d=0.29$) and pronation very small to medium responsiveness (Rater 1: $d=0.29$; Rater 2: $d=0.05$).

Conclusion

1
2
3 The GoWatch is a viable and user-friendly alternative to measure forearm rotation with
4 demonstrable validity, interrater reliability and responsiveness. Further research is required to
5 ensure systematic bias is not endemic when used across multiple raters.
6
7
8
9

10 11 12 **INTRODUCTION**

13
14 Forearm rotation requires unrestricted mobility of the proximal radioulnar, distal radioulnar and
15 humeroradial joints making it possible to orientate the palm up or down in functional tasks(1).

16
17 When injury occurs to these structures, function can be impaired (2). Supination in particular can
18 be affected due to restrictions imposed by casts and splints, protective posturing and as a result
19 of a natural tendency to perform most activities of daily living with the forearm in pronation(3).

20
21 One example of an injury that can present with restrictions in forearm rotation are distal radius
22 fractures. The incidence of this injury is rising globally(4) and as such patients with distal radius
23 fractures are commonly assessed and treated by healthcare practitioners.
24
25

26
27 Outcome measures are vital for appropriate goal setting and to ascertain patient progress and
28 subsequently must be reliable, valid and responsive(5). It is recommended when using the
29 International Classification of Functioning, Disability, and Health (ICF) as a framework that
30 assessment should incorporate *Body Functions and Structures, Participation, Activities* and the
31 *Environment*(6). In the context of hand therapy, one example of *Body Functions and Structures*
32 assessment is goniometry to objectively measure range of movement (ROM). Guidelines to
33 standardise methods of goniometer measurement(7) recommend measuring forearm rotation
34 using a two-arm goniometer with one arm placed on the wrist and the other aligned parallel to
35 the humerus(7). This standard method measures pure forearm rotation and has excellent test-
36 retest reliability and inter-rater reliability(8). However the circumference of the distal wrist is
37 oval-like in shape posing issues with accuracy in placing the level surface of the goniometer arm.
38 Additionally, the assessor is required to assume that the second arm is truly vertical using vision
39 only.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 An alternative approach is referred to as functional rotation, taking into consideration hand
4 orientation and therefore may be more meaningful to everyday function. Two techniques to
5 measure functional rotation have been described(9, 10). One involves the patient gripping a
6 pencil while the therapist aligns one arm of the goniometer with the pencil(9). The second
7 describes a tubular handle attached perpendicular to the horizontal arm of a standard goniometer
8 and a plumb line attached to its axis(10). The patient holds the tubular handle to define the plane
9 of the palm so when the forearm is rotated the weighted plumb line establishes the vertical plane.
10 These measures have reported high intra-rater and inter-rater reliability(10-12), though several
11 disadvantages to both approaches are noted. When measuring functional rotation, the patient
12 must be able to grasp the handle or pencil to define the palmar plane, however commonly
13 observed in patients with distal radius fractures is the inability to make a fist in the early phase of
14 rehabilitation. Further, compensatory movements at the wrist and the fourth and fifth
15 metacarpals need to be considered along with proximal positions.

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32 A design by Szekeres *et al.* demonstrated how measurement of forearm rotation can be improved
33 by attaching a plumb line to a finger goniometer with flat arms(13) (Figure 1). The weight on the
34 goniometer uses gravity to achieve a true vertical position. The authors propose that the flat
35 arms of the finger goniometer allow the therapist to secure it more firmly to the flatter surface of
36 the dorsal aspect of the wrist during measurement, gaining a more stable position. This design
37 was further tested and demonstrated the same inter-rater reliability as the standard approach in
38 measuring supination and slightly higher reliability in measuring pronation(14). However this
39 method requires continual manipulation by the clinician during measurement, limiting
40 observation of the upper extremity as the therapist is unable to move back from the patient to
41 observe for any compensatory postures.

42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
In this study we tested an alternative design to measure pure forearm rotation which considers
the issues raised with existing measuring approaches. Unlike the recent method proposed by
Szekeres *et al.*(13), this method does not require continual manipulation by the clinician, allowing

1
2
3 the therapist to observe for any compensatory postures. This is achieved by applying consistent
4 pressure to the wrist using a “snap” bracelet with a mounted goniometer watch. A "snap"
5
6 bracelet is a bi-stable object that can be manipulated into two different configurations(15): it can
7
8 transit from a straight shape with a groove along its entire width to a coiled shape where the
9
10 groove disappears.
11

12
13
14 The GoWatch was designed in-house using a novelty toy watch purchased from a commercial
15
16 outlet. The strap of the watch was a “snap” bracelet with a silicone cover that also had a circular
17
18 mount where the small toy watch face was inserted. The GoWatch dial face was produced using
19
20 a high resolution 3D-printer so that its base could fit into the existing mount.
21

22
23 Previous attempts to design goniometers that attach to the wrist(16, 17) were impractical as they
24
25 involved the assembly of a large apparatus. Further, the design in this study uses a ball bearing to
26
27 act as a gravity-assisted dial, improving linear motion by reducing friction between the
28
29 goniometer components. Therefore this design may theoretically offer further improvements to
30
31 the recent method described and potentially be more user-friendly.
32

33
34 The purpose of this study was to test the concurrent validity, inter-rater reliability and
35
36 responsiveness of this new goniometer design in measuring pure forearm rotation with the
37
38 modified finger goniometer as the criterion instrument for reference.
39
40

41 42 43 **METHODS**

44 45 **Design**

46
47 This was a prospective repeated measures study examining the validity, inter-rater reliability and
48
49 responsiveness of a new goniometer device to measure pure forearm rotation. The Guidelines
50
51 for Reporting Reliability and Agreement Studies (GRRAS)(18), COnsensus-based Standards for
52
53 the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist were used to
54
55 ensure quality procedures and reporting(19).
56
57
58
59
60

Participants

Participants with upper limb trauma or disease that had impeded forearm rotation were consecutively recruited if they met the inclusion criteria. The availability of the chief investigator and both raters was necessary if participation was to occur on the day of recruitment otherwise the participant could be involved at a future therapy session.

Ethics

This study was approved by the Office of Research Ethics Committees, Northern Ireland (ORECNI) and the Medicines and Healthcare products Regulatory Authority (MHRA). MHRA recommended that the goniometer watch was labelled “exclusively for clinical investigations” as it was a new medical device. All participants provided written, informed consent.

Recruitment

Potential participants were identified by surgeons and therapists through the Trust’s fracture clinic, physiotherapy and occupational therapy departments. Each eligible patient had a "cooling off" period to consider if they wished to participate in the study. Consent and participation in the study did not occur until participants saw the goniometer watch, gauged its acceptability and read the study information sheet. Alternatively, patients could choose to participate at their next planned hospital review. Nobody who met the study criteria and was invited declined to participate in the study.

Sample Size

Bland and Altman recommend that agreement is necessary to establish if the precision of a new measuring device is acceptable compared to the gold standard. They highlighted this using a sample size of only 17 participants(20). A sample of 30 is commonly employed as a benchmark to rely on the central limit theorem, as statistical research has found that with a sample size of 30

1
2
3 the sampling distribution of the mean is approximately normal(21). Therefore 40 participants
4
5 were proposed as an adequate and cost-effective sample size.
6
7
8
9

10 **Inclusion Criteria:**

11
12 The study aimed to recruit people aged 18 years and over with a history of upper limb trauma or
13
14 disease affecting forearm rotation and that active ROM of the forearm was permitted by the
15
16 treating physician.
17
18
19
20

21 **Exclusion Criteria:**

22
23 Any patient with an open wound around the distal portion of the wrist, hyperaesthesia or
24
25 allodynia to the wrist or a cognitive impairment or learning difficulty that prevented
26
27 understanding of verbal instructions or informed consent were excluded from the study. Also
28
29 patients were excluded if ROM was contra-indicated due to phase of healing.
30
31
32
33
34

35 **Materials**

- 36
37 • Modified Finger Goniometer (MFG) (Device 1):

38
39 The weighted finger goniometer was placed on the dorsal aspect of the wrist, the long arm
40
41 crossing both Lister's tubercle and the ulnar head with the weight over the ulnar side of the wrist.
42
43 It is flipped to measure either supination or pronation (Figure 1).
44
45
46
47

- 48
49 • Goniometer Watch (Device 2):

50
51 A 360° goniometer watch (GoWatch) with measurements in 2° increments was used (Figure 1).
52
53 Zero degrees are indicated at the bottom of the watch face and incorporates gravity to influence
54
55 motion of the ball-bearing. The GoWatch is attached to the radial border of the wrist just
56
57 proximal to the ulnar styloid using the snap bracelet mechanism. The strap is attached so the 0°
58
59 marking on the watch face is aligned with the centre of the anatomical snuffbox. The arm is
60

1
2
3 positioned to the side of the trunk in the midline position and elbow flexed to 90°. The forearm
4
5 moves either into pronation or supination and a record is made of the degrees reached on the
6
7 GoWatch face.
8
9

10 11 12 **Procedure**

13
14 A pilot study was conducted with ten participants to familiarise therapists with the process and
15
16 address any safety concerns. No safety concerns were highlighted. Some guidance was provided
17
18 to both raters by the Chief Investigator on the application of both measuring devices during the
19
20 pilot. Two occupational therapists each with over ten years' experience were designated raters.
21
22 Two trials of each goniometry method were performed by each therapist before and after a
23
24 session of hand therapy. The first trial data were used to compare measures between raters
25
26 (inter-rater reliability) and also measures between devices (concurrent validity). The second trial
27
28 was used to examine the responsiveness of the devices.
29
30
31
32
33

34 35 **Randomisation and Blinding**

36
37 A box of 90 cards, indicating which instrument, therapist, and direction of rotation would be
38
39 measured first was used. The second therapist followed the same sequence of instrument and
40
41 direction. An online random sequence generator (www.random.org) determined the sequence of
42
43 cards. After each measure, the rater gave their result verbally to the Chief Investigator who
44
45 recorded it. Each rater did not have access to the measurement sheets to assist blinding to the
46
47 first set of results. Each rater was also blinded to the other's results during the course of the
48
49 study. Participants were not blinded to the results as they would hear the results verbally after
50
51 each measure.
52
53

54
55 Participants were assessed over one session with approximately 30-60 minutes between trials. A
56
57 therapy session including various active and passive movements and/or heat therapy occurred
58
59
60

1
2
3 during the interval. Participants sat on a plinth with their feet flat on the ground and were
4
5 instructed to keep their elbow to the side of their trunk. Therapists observed that the
6
7 participant's shoulders were level, the elbow positioned at 90° and the forearm was in mid-
8
9 position before movement in either direction occurred. Participants were then instructed to
10
11 move their forearm arm into maximum pronation or supination.
12
13
14
15

16 **Statistical analysis**

17 Concurrent Validity and Inter-rater Reliability

18
19 Bland Altman plots were populated on Microsoft Excel (2016) to assess agreement between each
20
21 device (concurrent validity) and each rater (interrater reliability) with *a priori* limits of agreement
22
23 set at 5° and that the confidence interval (CI) of the mean difference included the line of equality
24
25 (i.e. no difference in mean scores). A 5 to 10° margin of error has been deemed acceptable for
26
27 intra and interrater reliability of wrist range of movement (22). The authors deemed a threshold
28
29 of 5° difference as an acceptable benchmark for interrater reliability. Szekeres *et al.*(14) do not
30
31 provide details on the level of agreement between the two-arm goniometer and the MFG
32
33 however provide the mean difference between raters using Bland Altman which reported
34
35 pronation as 5° [95%CI(-6, 16)] and supination as 3° [95%CI(-9, 14)]. In the absence of any *a*
36
37 *priori* values in the literature for concurrent validity, the authors also determined within 5°
38
39 difference as an acceptable level of agreement.
40
41
42
43
44
45

46 Concurrent validity is the extent two instruments agree in measuring the same construct. One
47
48 instrument is typically an established tool used as a reference when assessing the accuracy of a
49
50 new measurement instrument(23). Inter-rater reliability is the extent to which the measures by
51
52 two or more raters agree(24). Bland Altman enables visualization of accuracy of these properties.
53
54
55

56
57 Bias
58
59
60

1
2
3 Repeated-measures ANOVA using Statistical Package for Social Sciences (v.25 SPSS) measured
4 the extent to which values recorded the similarity and differences between raters and
5 instruments. The primary benefit of using repeated-measures ANOVA instead of Intraclass
6 Correlation Coefficients (ICC) is that it can provide a more precise estimate of agreement(20)
7 and more detailed information about the difference or relationship between raters and
8 instruments, separating bias from unexplained residual error. Significant bias (systematic
9 differences across variables) was set at $p < 0.05$. The square root of within-subject mean square
10 errors was used to calculate unexplained error.
11
12
13
14
15
16
17
18
19
20
21
22

23 Responsiveness

24
25 Responsiveness of each device was assessed by measuring the effect size between trial 1 and 2
26 using Cohen's d to interpret if the effect size with small, medium or large with (d) equal to 0.2,
27 0.5 and 0.8 respectively(25). Responsiveness is the extent to which a device can detect change
28 over time(26). No previous analysis of responsiveness of the MFG has been conducted.
29
30
31
32
33
34
35
36

37 Interrater Reliability (Secondary Analysis)

38
39 ICCs were performed as a secondary analysis of inter-rater reliability to compare with key results
40 reported in the literature. Inter-rater reliability using ICCs (type 2.2) was calculated at 95% CI. A
41 two-way random effects model with the mean scores ($k = 2$) was used to ascertain absolute
42 agreement between raters for both techniques(27). An ICC score greater than 0.9 is deemed
43 excellent while scores less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9 are poor,
44 moderate and good respectively(28). ICC scores were reported as average measures. Standard
45 error of measurement and 95% CI (SEM_{95}) were also calculated to determine the precision of
46 each device and Shapiro-Wilk test ($p > 0.05$) was used to ascertain normality of data sets used.
47
48
49
50
51
52
53
54
55
56
57
58

59 RESULTS

1
2
3 From November 2019 until December 2022 40 participants were recruited (Table 1). One upper
4 extremity was assessed for each participant and all had restrictions in forearm rotation due to
5 injury with most participants being female. Descriptive statistics on both measures are provided
6 in Table 2.
7
8
9
10
11
12
13

14 **Concurrent Validity**

15
16 The mean difference between the MFG and GoWatch values for supination and pronation in
17 trial 1 followed a normal distribution as determined by the Shapiro-Wilk test ($p > 0.05$). Bland
18 Altman plots illustrate the mean difference between both instruments for supination was 1.19°
19 [95%CI(-1.77, 4.15)] and for pronation was 0.20° [95%CI(-3.10, 2.70)] both of which were
20 within the 95% CI and the 5° acceptable range indicating agreement (Figure 2).
21
22
23
24
25
26
27
28
29

30 **Inter-rater Reliability**

31
32 Agreement between raters was also assessed using Bland Altman plots. Normality of the mean
33 difference scores between raters was determined using the Shapiro-Wilk tests however values for
34 the mean difference in MFG pronation values were non-normal ($p < 0.05$) and Log10
35 transformations could not be performed due to zero and negative values.
36
37
38
39

40
41 The mean difference between rater 1 and 2 measuring supination with the GoWatch was 4.43°
42 [95%CI(.52, 8.33)] and for pronation was 2.23° [95%CI(-2.22, 6.65)] whereby the line of equality
43 fell within the 95% CI and also within the 5° acceptable range indicating agreement. The mean
44 difference between rater 1 and 2 measuring supination with the MFG was -1.30° (CI95: -5.01,
45 2.41) whereby the line of equality fell within the 95% CI, also within the 5° acceptable range and
46 indicating agreement (Figure 3).
47
48
49
50
51
52
53

54
55 Secondary analysis using ICCs was then performed. MFG data sets did not show linearity along
56 the QQ plot line which was confirmed by the Shapiro-Wilks test ($p < 0.05$). GoWatch supination
57 in trial 1 was 0.94 [95%CI(.86-.97)] and pronation was 0.85 [95%CI(.71-.92)] indicating excellent
58
59
60

1
2
3 and good interrater reliability respectively. The SEM_{05} for the GoWatch in trial 1 was 6.44° for
4
5 supination and 7.33° for pronation, both outside the acceptable limits ($<5^\circ$).
6
7
8
9

10 **Systematic Bias**

11 Repeated-measures ANOVA data from trial 1 detected significant systematic bias between all
12
13 supination and pronation measures ($p<0.01$) with higher pronation values and a residual error of
14
15
16 36.21° (Mean square error = 1311.08).
17

18 Figure 4 illustrates a significant rater-by-device interaction with systematic bias between rater 1
19
20 and 2 using the GoWatch (Device 2) when analysing average measures of combined pronation
21
22 and supination in trial 1 ($p=0.02$) with a residual error of 5.95° (Mean square error = 35.37).
23
24

25 Rater 1 and Rater 2 closely agreed on MFG (Device 1) and GoWatch measures (interaction
26
27 between error bars) however systematic bias was observed with Rater 2 consistently
28
29 underestimating measures with the GoWatch compared to Rater 1. This is represented by the
30
31 diagonal line between rater 1 and rater 2.
32
33

34 Repeated-measures ANOVA agreement between each rater, each device and each component of
35
36 ROM (supination and pronation) in trial 1, detected no significant bias ($p=0.18$) however the
37
38 residual error of 8.26° (Mean square error = 68.28) was outside the acceptable *a priori* level of 5° .
39
40
41
42

43 **Responsiveness**

44 Normality testing of all rater 1 MFG measures and rater 2 MFG supination measures were
45
46 significant using the Shapiro-Wilk test ($p<0.05$) however both raters' GoWatch measures
47
48 followed a normal distribution and therefore were analysed for responsiveness using Cohen's *d*.
49
50 Using measures collated by rater 1 a very small effect size was observed between trial 1 and 2
51
52 GoWatch supination ($d=0.14$ 95% CI [-.30, .58]) and a small to medium effect for pronation
53
54 ($d=0.29$ 95% CI [-.17, .72]). For rater 2 a small to medium effect size was observed between trial
55
56 1 and 2 GoWatch supination ($d=0.29$ 95% CI [-.16: .72]) and a very small effect for pronation
57
58
59
60

1
2
3 ($d = 0.05$ 95% CI [-.38, .49]). Repeated-measures ANOVA also confirm that there was a positive
4
5 effect between trial 1 and 2 when analysing average combined measures (i.e.
6
7 pronation+supination) ($p < 0.01$) with a residual error of 7.66° (Mean square error = 58.74).
8
9

10 11 12 **DISCUSSION**

13
14 The GoWatch demonstrated acceptable concurrent validity and interrater reliability for both
15
16 supination and pronation measures ($< 5^\circ$) however systematic bias was found when the rater by
17
18 instrument interaction was analysed. The MFG demonstrated superior agreement between raters
19
20 compared to the GoWatch. Results from the study by Szekeres *et al.*(14) further demonstrates
21
22 superior agreement between raters using the MFG than the results obtained on the GoWatch in
23
24 this study.
25
26

27
28 Bland Altman plots (Figure 3) illustrate that the difference between raters measuring supination
29
30 using the GoWatch increased as higher ROM measures were obtained, while differences outside
31
32 the *apriori* limits of agreement for GoWatch pronation and MFG supination were evenly spread.
33
34 Rater 2's total measures (supination+pronation) produced on average significantly lower
35
36 GoWatch measurements compared to their MFG measures in trial 1 (Figure 4). GoWatch
37
38 responsiveness in detecting change between trial 1 and 2 was significant using both statistical
39
40 methodologies i.e. repeated-measures ANOVA and Cohen's *d*. Each rater detected some change
41
42 using the GoWatch though this varied from a very small to small / medium effect. This
43
44 treatment effect between trials was hypothesized as it was intended to illicit an improvement in
45
46 ROM. However the ability for the GoWatch to measure responsiveness should be considered
47
48 with caution due to the systematic bias detected in this study and also that the procedure was
49
50 unable to completely "blind" both raters and participants from the results. Further mean
51
52 difference in scores between trial 1 and 2 for both raters measuring supination and pronation are
53
54 low relative to the acceptable level of agreement ($< 5^\circ$) except for rater 2 measuring supination
55
56 which had a mean difference of 6.45° . Further Armstrong *et al.*(8) have recommended that a
57
58
59
60

1
2
3 meaningful change in rotation measures would be at least 10°. Future studies on the GoWatch
4
5 should consider measuring responsiveness over a longer duration to determine if it can detect
6
7 meaningful clinical change.
8

9
10 Secondary analysis of GoWatch scores using ICCs, demonstrated good interrater reliability for
11
12 measuring pronation and excellent interrater reliability for measuring supination. However this
13
14 criteria can be misleading as the use of correlation coefficients to test agreement between
15
16 measurement instruments and raters is cautioned(20). ICCs imply the GoWatch is fit for purpose
17
18 in a clinical setting however fail to identify systematic bias of the GoWatch measures. GoWatch
19
20 ICCs imply excellent interrater reliability for measuring supination and good interrater reliability
21
22 for measuring pronation yet Bland Altman plots indicate mean difference was smaller for
23
24 pronation. This disparity between ICCs and mean difference have previously been
25
26 demonstrated(29). These authors advise that ICCs “do not always reflect the clinical implications
27
28 of measurement errors” and recommend the utilization of other statistical methods such as those
29
30 used in this study. The results further highlight the risk of basing assumptions on clinical
31
32 acceptability of a measuring instrument solely on ICCs. Szekeres *et al.*(14) reported that the MFG
33
34 had excellent inter-rater reliability for pronation (ICC: 0.86) and supination (ICC: 0.95)(14),
35
36 similar to results for the GoWatch, however no systematic bias was identified with the MFG
37
38 using linear regression analysis. SEM₉₅ scores for GoWatch pronation was 7.33° and for
39
40 supination was 6.44°, both outside the acceptable *a priori* limits and a larger error than previously
41
42 reported for the MFG (SEM₉₅ pronation = 2.1° and supination = 1.2°). SEM₉₅ will decrease as
43
44 sample size increases and Szekeres *et al.*(14) measured ICCs and SEM₉₅ using pooled samples of
45
46 60 participants over six sessions producing a sample size of 360 possibly explaining smaller
47
48 SEM₉₅ values. With systematic bias SEM₉₅ values may not reduce much further regardless of a
49
50 larger sample size. Results were not pooled from both trials because the second trial measured
51
52 effect size and therefore pooling to assess agreement and reliability would be a flaw in the
53
54
55
56
57
58
59
60

1
2
3 statistical approach. Non-normality of MFG data meant comparisons with GoWatch
4
5 responsiveness could not be made.
6

7 MFG supination measures had non-normal platykurtic kurtosis meaning that the values were
8
9 uniform with few outliers. Rater 1 pronation measures had non-normal leptokurtic kurtosis
10
11 meaning that there were frequent outliers. Similar trends were observed in the GoWatch
12
13 measures and MFG rater 2 pronation measures but kurtosis fell within acceptable limits (± 1.0).
14
15 Systematic bias was also observed between all measures of supination and pronation with a
16
17 residual error of 36.21° . This was hypothesized as most people have limitations in supination
18
19 rather than pronation after a distal radius fracture(3) and this phenomenon was observed in this
20
21 cohort (82.5% of study participants). This may also explain the negative skewness of the
22
23 pronation measures which were observed in both raters though with rater 2's measures just
24
25 within normal distribution.
26
27

28
29 Systematic bias between raters may be an isolated event, due to weaknesses in blinding or the
30
31 potentially arbitrary reference point for placing the GoWatch. The wide area of the anatomical
32
33 snuffbox may produce variability in placement and therefore in measurement readings. In larger
34
35 or oedematous limbs, the anatomical snuffbox may be less defined and further affect accuracy.
36
37 Participants were on average assessed in this study at around three months from injury or
38
39 surgery though details on oedema or visibility of the anatomical snuffbox was not collated for
40
41 this study.
42
43

44
45 Several suggestions on future iterations may help reduce bias. Adjusting the position of the
46
47 watch face whereby it is positioned on the dorsum of the wrist where the flatter and wider
48
49 surface area may offer a more stable base for the watch and limit unwanted movement. Aligning
50
51 the centre of the watch face with the third metacarpal base by a reference line on the snap
52
53 bracelet could further improve precision, and a serrated edge on the interior of the watch face
54
55 may minimise oscillating movements when end ROM is held.
56
57
58
59
60

1
2
3 Both raters in this study were experienced in hand therapy and were inducted via a pilot study on
4 both methods which were novel to them. Both raters did not find the MFG intuitive, with the
5 Chief Investigator interjecting to provide instruction to flip the goniometer to enable
6 measurements in opposing directions. Formal feedback on each rater's preference of instrument
7 was however beyond the scope of this study.
8
9

10
11
12
13
14 Limitations included systematic bias of rater-instrument interactions restricting a thorough
15 understanding of the reliability, validity and responsiveness of the GoWatch. Using only two
16 raters limited generalisation of the results to hand therapy practitioners. Some datasets were not
17 normally distributed and as such some important aspects of analysis could not be completed.
18
19

20
21
22
23 The study commenced in November 2019 and delays in recruitment were encountered due to
24 COVID-19.
25
26

27
28 Systematic bias may be resolved by further research with more than two raters and provision of
29 more extensive training on each instrument. Future study procedures should also consider
30 strategies to ensure both participants and raters are completely blind to measurements. Further
31 iterations to the GoWatch may also improve precision and reduce systematic bias. The hinge at
32 the base of the GoWatch allows it to fold 90° degrees parallel to the forearm and it can also
33 rotate 360° in its fulcrum. Subsequently the GoWatch may be used to measure other joint ROM
34 (e.g. elbow flexion). The raters in this study found the GoWatch more intuitive to use than the
35 MFG however this is not particularly reflected in the results. Further, exploration of clinician
36 perceptions would be an important area of research. This study highlights the need for caution
37 when interpreting ICC scores of measuring instruments. Using the categories defined by Shrout
38 and Fleiss(28) can be misleading when choosing an instrument for clinical use.
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53 The GoWatch is easy to use and further iterations may help improve its precision so that it
54 becomes an instrument of choice for many clinicians. Thereafter, further research will be
55 necessary to evaluate its measurement properties and to ascertain if systematic bias is endemic
56 across multiple raters.
57
58
59
60

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest

Funding

The authors disclose receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the British Association of Hand Therapists (BAHT) [2018].

REFERENCES

1. Soubeyrand M, Assabah B, Bégin M, Laemmel E, Dos Santos A, Crézé M. Pronation and supination of the hand: Anatomy and biomechanics. *Hand Surg Rehabil.* 2017;36(1):2-11.
2. Kleinman WB. Distal radius instability and stiffness: common complications of distal radius fractures. *Hand Clin.* 2010;26(2):245-64.
3. Naughton N, Algar L. Therapy Management of Distal Radius Fractures. In: Skirven TM, Osterman AL, Fedorczyk JM, Amadio PC, Feldscher SB, Shin EK, editors. *Rehabilitation of the Hand and Upper Extremity* 7th ed. Elsevier Health Sciences: Philadelphia, 2021. p. 833-850
4. MacIntyre NJ, Dewan N. Epidemiology of distal radius fractures and factors predicting risk and prognosis. *J Hand Ther.* 2016;29(2):136-45.
5. Yang W, Houtrow A, Cull DS, et al. Quality and Outcome Measures for Medical Rehabilitation. In: Cifu DX, editor. *Braddom's Physical Medicine and Rehabilitation* 6th ed. Elsevier: Philadelphia 2021. p. 100-14.e2.
6. World Health Organisation. International classification of functioning, disability, and health, <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health> (2018, accessed 12 October 2023)

- 1
2
3 7. MacDermid J, Solomon G, Valdes K. *Clinical Assessment Recommendations* 3rd ed. American
4 Society of Hand Therapists: Mount Laurel, NJ, 2015
- 5
6
7 8. Armstrong AD, MacDermid JC, Chinchalkar S, Stevens RS, King GJ. Reliability of
8 range-of-motion measurement in the elbow and forearm. *J Shoulder Elbow Surg.* 1998;7(6):573-80.
- 9
10
11 9. McRae R. *Clinical Orthopaedic Examination*. 1st ed. Churchill Livingstone: Edinburgh;
12 1981.
- 13
14
15 10. Flowers KR, Stephens-Chisar J, LaStayo P, Galante BL. Intrarater reliability of a new
16 method and instrumentation for measuring passive supination and pronation: a preliminary
17 study. *J Hand Ther.* 2001;14(1):30-5.
- 18
19
20 11. McGarry G, Gardner E, Muirhead A. Measurement of forearm rotation: an evaluation of
21 two techniques. *J Hand Surg Br.* 1988;13(3):288-90.
- 22
23
24 12. Karagiannopoulos C, Sitler M, Michlovitz S. Reliability of 2 functional goniometric
25 methods for measuring forearm pronation and supination active range of motion. *J Orthop Sports*
26 *Phys Ther.* 2003;33(9):523-31.
- 27
28
29 13. Szekeres M, MacDermid JC, Rooney J. A new method for measuring forearm rotation
30 using a modified finger goniometer. *J Hand Ther.* 2015;28(4):429-31; quiz 32.
- 31
32
33 14. Szekeres M, MacDermid JC, Birmingham T, Grewal R. The inter-rater reliability of the
34 modified finger goniometer for measuring forearm rotation. *J Hand Ther.* 2016;29(3):292-8.
- 35
36
37 15. Guo Q, Zheng H, Chen W, Chen Z. Modeling bistable behaviors in morphing structures
38 through finite element simulations. *Biomed Mater Eng.* 2014;24(1):557-62.
- 39
40
41 16. Wagner C. Determination of the rotary flexibility of the elbow joint. *Eur J Appl Physiol*
42 *Occup Physiol.* 1977;37(1):47-59.
- 43
44
45 17. Darcus HD, Salter N. The amplitude of pronation and supination with the elbow flexed
46 to a right angle. *J Anat.* 1953;87(2):169-84.
- 47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 18. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al.
4
5 Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin*
6
7 *Epidemiol.* 2011;64(1):96-106.
8
9
- 10 19. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al.
11
12 COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of
13
14 outcome measurement instruments: a Delphi study. *BMC Medical Research Methodology.*
15
16 2020;20(1):293.
17
18
- 19 20. Bland JM, Altman DG. Statistical methods for assessing agreement between two
20
21 methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.
22
23
- 24 21. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. *Korean J*
25
26 *Anesthesiol.* 2017;70(2):144-56.
27
28
- 29 22. Marx RG, Bombardier C, Wright JG. What do we know about the reliability and validity
30
31 of physical examination tests used to examine the upper extremity? *J Hand Surg Am.*
32
33 1999;24(1):185-93.
34
35
- 36 23. Field A. *An Adventure in Statistics: The Reality Enigma.* 2nd ed. Sage Publications Ltd:
37
38 London, 2022. p 61
39
40
- 41 24. Lange RT. Inter-rater Reliability. In: Kreutzer JS, DeLuca J, Caplan B, editors.
42
43 *Encyclopedia of Clinical Neuropsychology.* Springer: New York, 2011. p. 1348.
44
45
- 46 25. Cohen J. *Statistical Power Analysis for the Behavioral Sciences* 2nd ed. Hillsdale, NJ: Lawrence
47
48 Earlbaum Associates, 1988.
49
50
- 51 26. Terwee CB. Responsiveness to Change. In: Michalos AC, editor. *Encyclopedia of Quality of*
52
53 *Life and Well-Being Research.* Springer: Netherlands, 2014. p. 5547-5550.
54
55
- 56 27. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation
57
58 Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155-63.
59
60
28. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological*
bulletin. 1979;86(2):420.

1
2
3 29. Lee KM, Lee J, Chung CY, Ahn S, Sung KH, Kim TW, et al. Pitfalls and important
4 issues in testing reliability using intraclass correlation coefficients in orthopaedic research. *Clin*
5
6
7 *Orthop Surg.* 2012;4(2):149-55.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Figure 1: Measuring pronation (top left) and supination (top right) using the goniometer watch (GoWatch) and measuring pronation (bottom left) and supination (bottom right) using the modified finger goniometer (MFG)

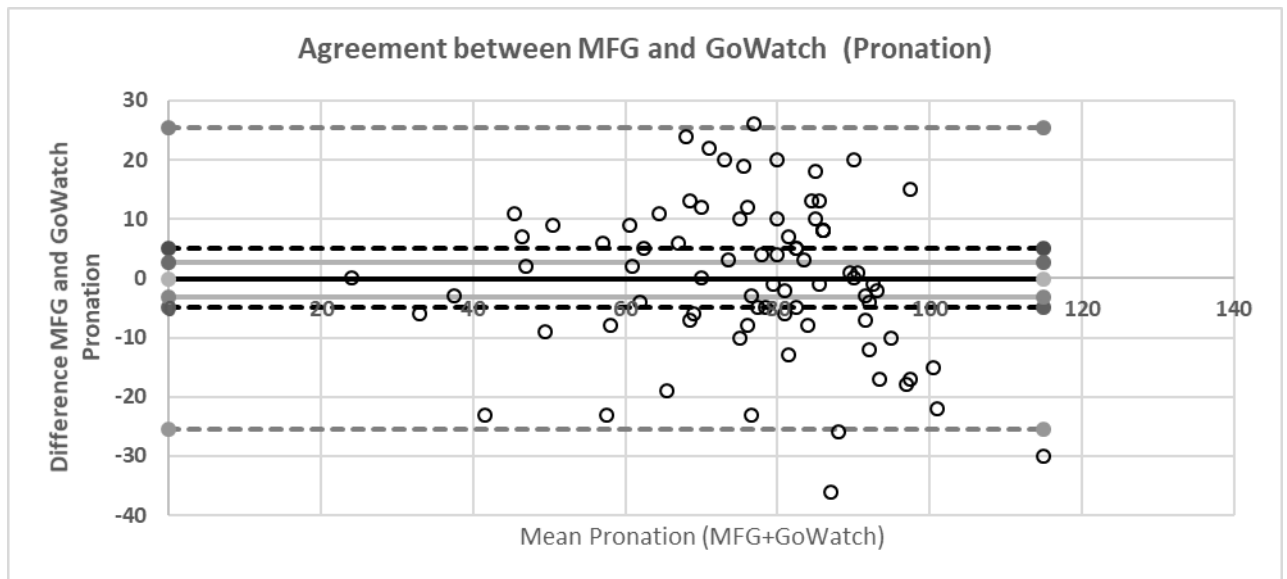
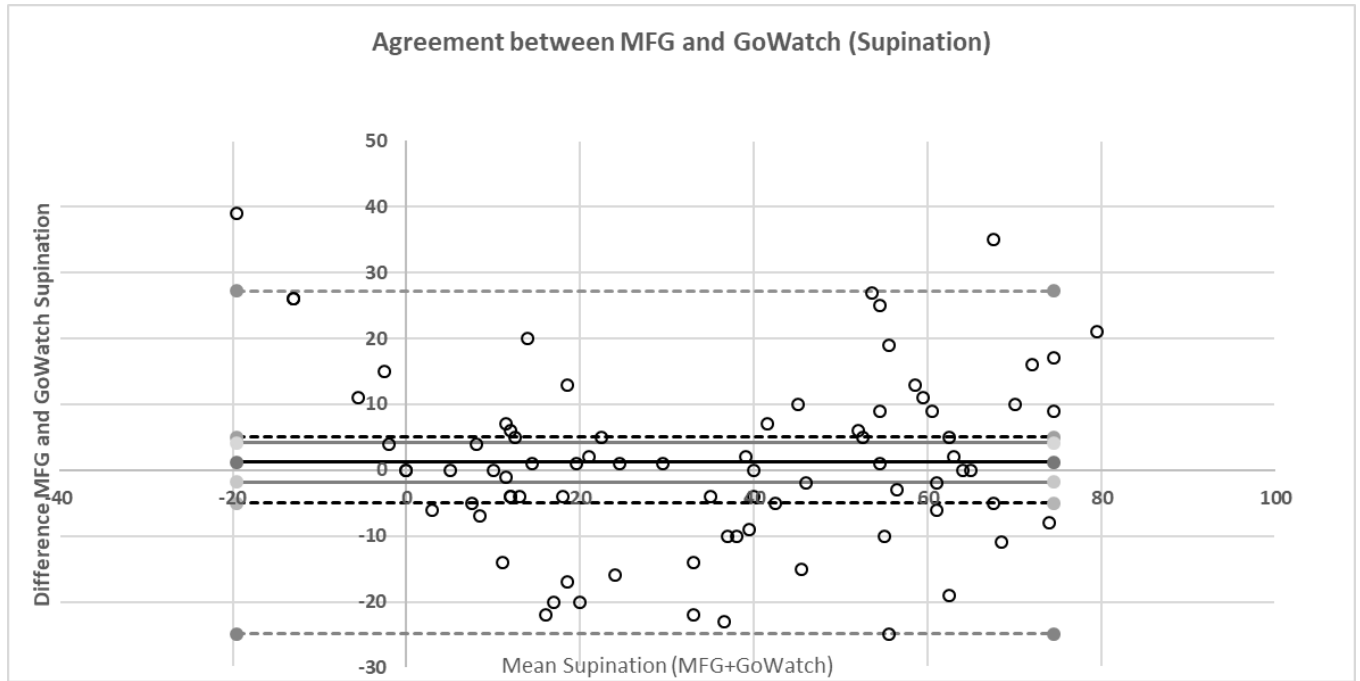
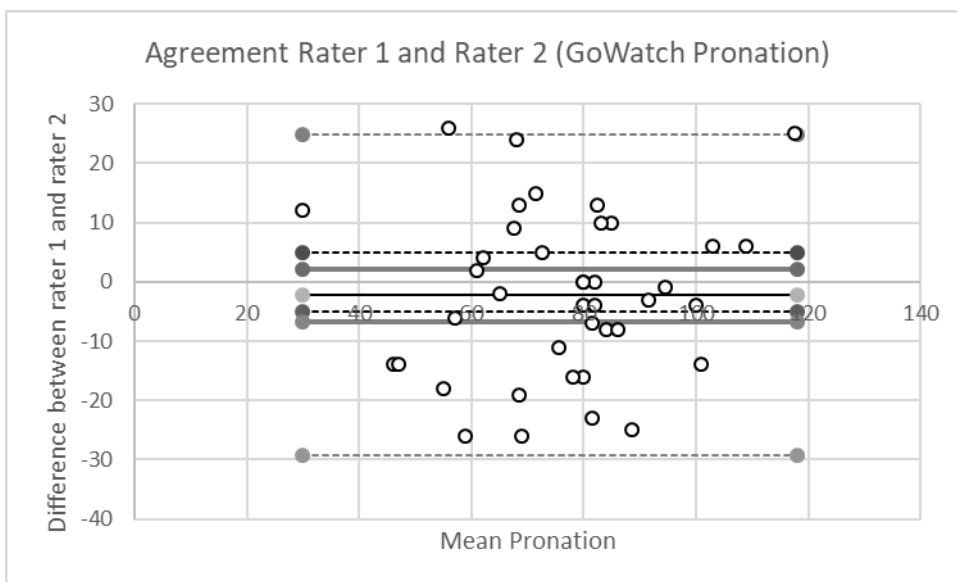
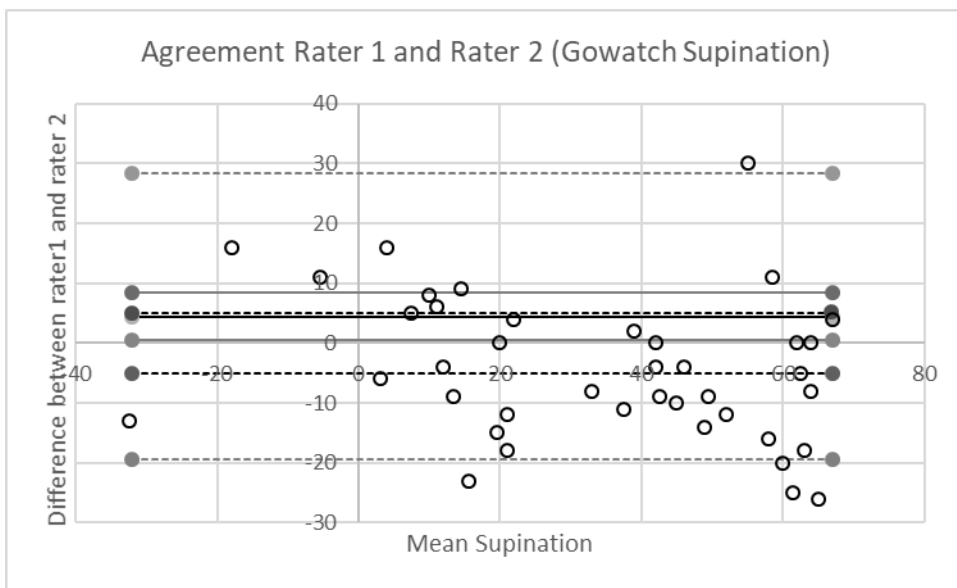


Figure 2: Agreement between MFG and GoWatch in measuring pronation and supination. Black central solid line = mean difference; two solid grey lines = 95% CI; black dotted lines = *a priori* Limits of Agreement (LOA); Outside grey dotted lines = LOA based on mean of the two values, minus and plus 1.96 standard deviations



1
2
3 Figure 3: Agreement between Rater 1 and 2 in measuring supination and pronation using the GoWatch
4 and supination using the MFG. Black central solid line = mean difference; two solid grey lines = 95% CI;
5
6 black dotted lines = *apriori* Limits of Agreement (LOA); Outside grey dotted lines = LOA based on mean
7
8 of the two values, minus and plus 1.96 standard deviations
9
10

11
12 Note: MFG supination upper *apriori* LOA black dotted line not visible as overlaps with upper 95% CI
13
14 (solid grey line)
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

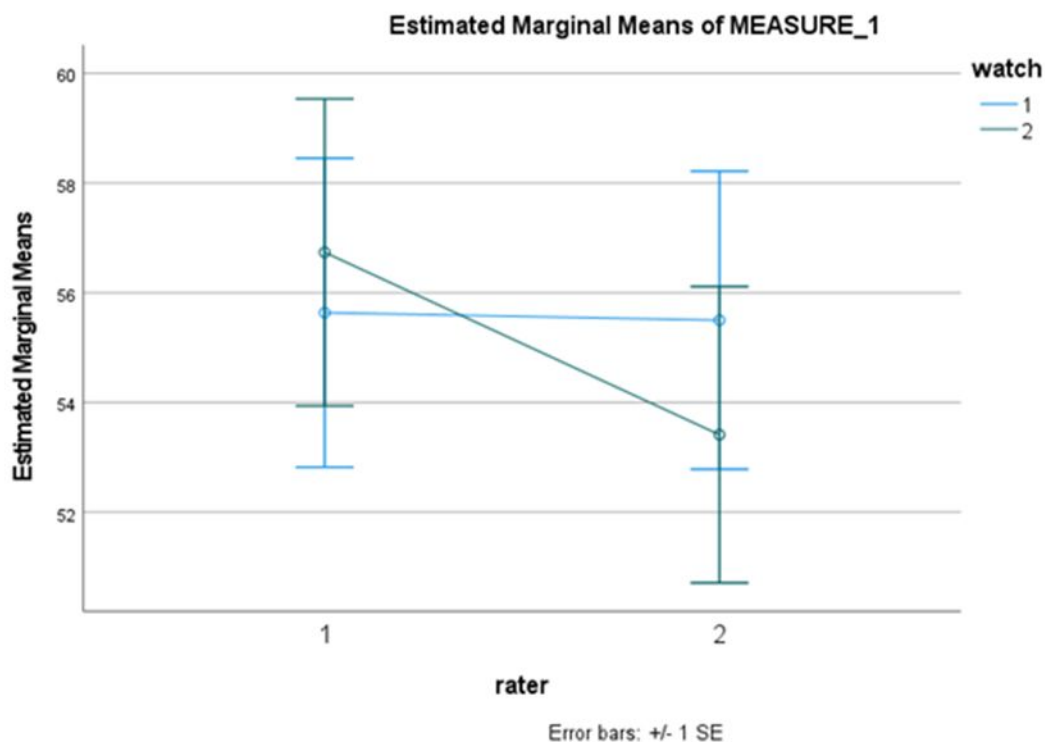


Figure 4: Estimated marginal mean for pooled total ROM (°) as measured by rater 1 and rater 2 in trial 1 using both the MFG and GoWatch. Watch 1 on x axis refers to MFG and Watch 2 refers to GoWatch. Overlapping vertical lines illustrate agreement between devices measured by each rater. Horizontal lines illustrate agreement between raters using each device. Error bars represent standard error. Total ROM refers to total arc of forearm rotation (pronation + supination)

Table 1: Demographic characteristics of participants (n=40)

Characteristics	Mean	SD	Range
Age	53.6	13.58	20-78
Time from Injury/ Surgery* (Days)	98.79	59.78	24-344
	n	(%)	
Female	30	75	
Affected Side			
Left	22	55	
Right	18	45	
Injury Type			
Humeral fracture	3	7.5	
Distal radius fracture	31	77.5	
Distal radius and ulna fracture	2	5	
Radial Head fracture	2	5	
Elbow dislocation	1	2.5	
Triad injury	1	2.5	

n = number of participants, SD = standard deviation

*Days counted from surgery rather than injury in cases where surgery was performed

Table 2. Descriptive statistics of GoWatch and MFG measures. Trial 1 measures carried out prior to therapy. Trial 2 measures carried out after 30-60 minute interval of hand therapy

GoWatch		MFG					
		n	Mean (SD)	Range	n	Mean (SD)	Range
<i>Trial 1</i>							
Rater 1	Pronation	40	77.35(18.40)	24-108	40	76.83 (17.44)	24-100
	Supination	40	36.13 (27.81)	-26-78	40	34.45 (27.44)	0-83
Rater 2	Pronation	40	75.13 (19.61)	36-130	40	75.25 (16.83)	30-105
	Supination	40	31.70 (24.67)	-39-70	40	35.75 (25.30)	0-90
<i>Trial 2</i>							
Rater 1	Pronation	40	82.25 (16.93)	50-119	40	78.08 (15.69)	37-100
	Supination	40	40.25 (27.00)	-21-86	39	37.74 (25.66)	0-84
Rater 2	Pronation	40	76.15 (16.45)	34-110	40	79.88 (17.45)	35-120
	Supination	40	38.15 (23.94)	-26-82	40	39.85 (26.67)	0-100

Note: Trial 1 measures carried out prior to therapy. Trial 2 measures carried out after 30-60 minute interval of hand therapy.

COSMIN Risk of Bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument

Table 1. Elements that make up a comprehensive reached question.

Element of the research question	
1	the name of the outcome measurement instrument
2	the version of the outcome measurement instrument or way of operationalization of the measurement protocol
3	the construct measured by the measurement instrument
4	a specification whether one is interested in a reliability parameter (i.e. a relative parameter such as an ICC, Generalizability coefficient ϕ , or Kappa κ) or a parameter of measurement error (i.e. an absolute parameter expressed in the unit of measurement e.g. SEM, LoA or SDC; or expressed as agreement or misclassification, e.g. the percentage specific agreement).
5	a specification of the components of the measurement instrument that will be repeated (especially when only part of the measurement instrument is repeated, e.g. only assignment of the score based on the same images)
6	a specification of the source(s) of variation that will be varied (e.g. time or occasion, the (level of expertise of) professionals, the machines, or other components of the measurement)
7	a specification of the patient population studied

ICC = Intraclass correlation coefficient; SEM = standard error of measurement; LoA = Limits of Agreement; SDC = smallest detectable change.

Responses to above from “Main Document”:

1. Refer to main manuscript: GoWatch
2. Page 5 Line 109 and Page 6 Line 146
3. Page 4 Line 95
4. Page 7 - 8
5. Page 7 “Procedure”
6. Page 7 “Procedure”
7. Page 4-6

Standards for studies on reliability

Design requirements	very good	adequate	doubtful	inadequate	NA
1 Were patients stable in the time between the repeated measurements on the construct to be measured?	Yes (evidence provided)	Reasons to assume standard was met	Unclear	No (evidence provided)	Na
2 Was the time interval between the repeated measurements appropriate?	Yes		Doubtful, OR time interval not stated	No	Na
3 Were the measurement conditions similar for the repeated measurements – except for the condition being evaluated as a source of variation?	Yes (evidence provided)	Reasons to assume standard was met, OR change was unavoidable	Unclear	No (evidence provided)	Na
4 Did the professional(s) administer the measurement without knowledge of scores or values of other repeated measurement(s) in the same patients?	Yes (evidence provided)	Reasons to assume standard was met	Unclear	No (evidence provided)	
5 Did the professional(s) assign scores or determine values without knowledge of the scores or values of other repeated measurement(s) in the same patients?	Yes (evidence provided)	Reasons to assume standard was met	Unclear	No (evidence provided)	
6 Were there any other important flaws in the design or statistical methods of the study?	No		Minor methodological flaws	Yes	

1. Yes – the manuscript describes randomisation of order of assessment between two raters and also set up procedure for measurement (very good)
2. To assess responsiveness of the device the time of up to one hour was appropriate for initial analysis however assessment of responsiveness over several treatment sessions would have been more insightful. However the responsiveness of the device was not the primary question of this study as initial focus was in interrater reliability and agreement with the gold standard. (very good)
3. Yes – above (very good)
4. Yes – Page 7 Line 165 (doubtful)
5. Yes – Page 7 Line 165 (doubtful)
6. Some datasets relating to the modified finger goniometer (reference outcome measure) were non-normal therefore some analysis relating to ICCs and responsiveness could not be carried out however this was not the instrument of concern in this study (adequate)

GRRAS checklist for reporting of studies of reliability and agreement

Version based on Table I in: Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Robersts C, Shoukri M, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106

Section	Item #	Checklist item	Reported on page #
Title/Abstract	1	Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	Page 1
Introduction	2	Name and describe the diagnostic or measurement device of interest explicitly.	Page 6
	3	Specify the subject population of interest.	Page 4
	4	Specify the rater population of interest (if applicable).	Page 6
Methods	5	Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	Page 5
	6	Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	Page 5 and 6
	7	Describe the sampling method.	
	8	Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	Page 7 and 8
	9	State whether measurements/ratings were conducted independently.	Page 7
	10	Describe the statistical analysis.	Page 7-10
	11	State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	Page 6, Page 8
Results	12	Describe the sample characteristics of raters and subjects (e.g. training, experience).	Page 5 and 6
	13	Report estimates of reliability and agreement including measures of statistical uncertainty.	Page 7-10
	14	Discuss the practical relevance of results.	Page 10-13
Discussion	14	Discuss the practical relevance of results.	Page 10-13
Auxiliary material	15	Provide detailed results if possible (e.g. online).	See Figures and Tables