Taylor & Francis
Taylor & Francis Group

Journal of Manual & Manipulative Therapy

ISSN: 1066-9817 (Print) 2042-6186 (Online) Journal homepage: https://www.tandfonline.com/loi/yjmt20

# Measurement of range-of-motion in infants with indications of upper cervical dysfunction using the Flexion-Rotation-Test and Lateral-Flexion-Test: a blinded inter-rater reliability study in a clinical practice setting

Femke Driehuis, Noël L.W. Keijsers, Maria W.G. Nijhuis–van der Sanden, Rob A. De Bie, J. Bart Staal & Thomas J. Hoogeboom

**To cite this article:** Femke Driehuis, Noël L.W. Keijsers, Maria W.G. Nijhuis–van der Sanden, Rob A. De Bie, J. Bart Staal & Thomas J. Hoogeboom (2020): Measurement of range-of-motion in infants with indications of upper cervical dysfunction using the Flexion-Rotation-Test and Lateral-Flexion-Test: a blinded inter-rater reliability study in a clinical practice setting, Journal of Manual & Manipulative Therapy, DOI: 10.1080/10669817.2020.1746896

**To link to this article:** https://doi.org/10.1080/10669817.2020.1746896

JOURNAL OF MANUAL & MANIPULATIVE THERAPY
https://doi.org/10.1080/10669817.2020.1746896



# Measurement of range-of-motion in infants with indications of upper cervical dysfunction using the Flexion-Rotation-Test and Lateral-Flexion-Test: a blinded inter-rater reliability study in a clinical practice setting

Femke Driehuis [a], Noël L.W. Keijsers [b,c], Maria W.G. Nijhuis–van der Sanden [a], Rob A. De Bie [d], J. Bart Staal [a,e] and Thomas J. Hoogeboom [a]

[a]IQ Healthcare, Radboud Institute of Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; [b]Institute Research, Development & Education, Sint Maartenskliniek, Ubbergen, The Netherlands; [c]Department of Rehabilitation, Cognition and Behavior, Donders Institute for Brain, Radboud University Medical Center, Nijmegen, The Netherlands; [d]CAPHRI Research School, Department of Epidemiology, Maastricht University, Maastricht, The Netherlands; [e]Research Group Musculoskeletal Rehabilitation, HAN University of Applied Sciences, Nijmegen, The Netherlands

**ABSTRACT**

**Background**: In infants with indications of upper cervical dysfunction, the Flexion-Rotation-Test and Lateral-Flexion-Test are used to indicate reduced upper cervical range-of-motion (ROM). In infants, the inter-rater reliability of these tests is unknown.

**Objective**: To assess the inter-rater reliability of subjectively and objectively measured ROM by using the Flexion-Rotation-Test and Lateral-Flexion-Test.

**Methods**: 36 infants (<6 months) and three manual therapists participated in this cross-sectional observational study. Pairs of two manual therapists independently assessed infants' upper cervical ROM using the Flexion-Rotation-Test and Lateral-Flexion-Test, blinded for each other's outcomes. Two inertial motion sensors objectively measured cervical ROM. Inter-rater reliability was determined between each pair of manual therapists. For subjective outcomes, Cohen's kappa (κ) and the proportion of agreement (Pr*a)* were calculated. For objectively measured ROM, Bland Altman plots were conducted and Limits of Agreement and Intraclass Correlation Coefficients (ICC) were calculated.

**Results**: The inter-rater reliability of the Flexion-Rotation-Test and Lateral-Flexion-Test for subjective (κ: 0.077–0.727; Pr*a*: 0.46–0.86) and objective outcomes (ICC: 0.019–0.496) varied between pairs of manual therapists.

**Conclusion**: Assessed ROM largely depends on the performance of the assessment and its interpretation by manual therapists, leading to high variation in outcomes. Therefore, the Flexion-Rotation-Test and Lateral-Flexion-Test cannot be used solely as a reliable outcome measure in clinical practice and research context.

**KEYWORDS**
Manual therapy; Flexion-Rotation-Test; Lateral-Flexion-Test; diagnostic; reliability; infants; upper cervical dysfunction

## Introduction

In the current clinical practice, many children and infants are treated with manual therapy for various musculoskeletal and non-musculoskeletal conditions [1–4]. In the Netherlands, upper cervical dysfunction (UCD) is considered the primary treatment indication in infants [5]. Persistent UCD could induce the maintenance of postural asymmetry and lead to a reduced active and passive cervical range of motion (ROM), resulting in a fixed asymmetric position of the infant's head toward lateral flexion and contralateral rotation [6–8]. Infants with persistent positional preference and indications of UCD seem to have more signs of skull deformation, excessive crying, and restlessness [5,6,9].

In manual therapy practice, the clinical diagnosis of UCD is based on the assessment of upper cervical ROM using the Flexion-Rotation-Test (FRT) and the Lateral-Flexion-Test (LFT) [5]. These tests assess whether upper cervical passive mobility toward rotation in full flexion and lateral flexion is either normal or reduced. When at least one of these tests indicates reduced passive ROM, UCD is clinically diagnosed and, dependent on the direction of reduced mobility, treated with specific techniques [5]. To date, research has acknowledged good reliability of the FRT in adults [10–13] and children [14] while in infants only one study examined the intra-rater reliability; one rater examined infants with torticollis and found high intra-rater reliability (ICC: ≥0.77) [15]. Even though the validity and reliability of the FRT and LFT in infants are still largely unknown, manual therapists currently use these tests in their diagnostic clinical decision-making [5]. Therefore, there is a strong need to determine the reliability of the FRT and LFT in clinical practice. Our study aims to examine the inter-rater reliability of (1) subjectively reported outcomes by manual therapists on the FRT and LFT and related decision-making, and (2) objectively measured ROM by inertial motion sensors during the FRT and LFT, in

**CONTACT** Femke Driehuis femke.driehuis@radboudumc.nl IQ Healthcare, P.O. Box 9101, 114, Nijmegen 6500 HB, The Netherlands
Supplemental data for this article can be accessed here.

infants with indications of UCD. Additionally, we aimed to verify the subjectively reported outcomes with objectively measured ROM.

## Methods

The Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were used to report our study [16]. Our study was approved by the Medical Ethical Committee for Research Involving Human Subjects of the Radboud university medical center, Nijmegen (CMO, NL.58488.091.16).

### *Study design*

In a cross-sectional observational study, the inter-rater reliability of the FRT and LFT in infants with indications of the presence of UCD was determined. Three manual therapists participated in the study. Pairs of two manual therapists independently assessed upper cervical mobility by performing the FRT and LFT in each infant. Simultaneously, two light-weight inertial motion sensors with a sampling rate of 100 Hz (MTw, Xsens BV, Enschede, the Netherlands) were used to objectively measure ROM in three dimensions during the mobility assessment. A schedule (Figures A1 and A2) was used to ensure that both the order of manual therapists and measurements (tests) was counterbalanced. In total there were three pairs of manual therapists: therapists A–B, B–C, and A–C. Moreover, therapists were blinded for each other's outcomes on the FRT and LFT and for objectively measured ROM. The assessment was executed in the practices of participating therapists.

### *Study population*

Three expert manual therapists registered in the Dutch registry of pediatric manual therapists [17] were invited for study participation and gave written informed consent. These qualified manual therapists had 10 to 17 years of experience in the treatment of infants with UCD, treated at least four infants per month, and were able and willing to recruit parents and infants for the study and to travel between the participating practices during the study period. All three manual therapists worked independently of each other in private practices in the Netherlands.

Infants (<6 months) visiting these practices because of an indication of UCD (indicated by a referrer or the infant's parents) were eligible to participate in the study. Previous or ongoing treatment with pediatric physical therapy was allowed because this is usual care in the Netherlands. Infants who were previously treated by one of the participating manual therapists for the same treatment indication were excluded. Both of the infant's parents had to provide written informed consent for study participation.

### *Study procedure*

Manual therapists were instructed to inform parents about the study when they registered their infant for treatment. If interested in participation, parents received an extensive information letter and were contacted by the primary author (FD) to explain the study procedure, including informed consent. In each infant, mobility assessment was performed by one pair of manual therapists. The therapist working at the practice where the infant was registered was considered the primary therapist and was therefore always one of the assessing manual therapists. The recruitment period was between June and December 2017. Further information about the study procedure is given in the Appendix.

### *Mobility assessment*

The mobility assessment consisted of an intake and passive mobility assessment (Figure 1). First, while manual therapists were in another room, parents were requested by FD to complete a questionnaire regarding infants' demographics, complaints and symptoms, and pregnancy and delivery because of the potential relation with UCD [5,6,18]. Manual

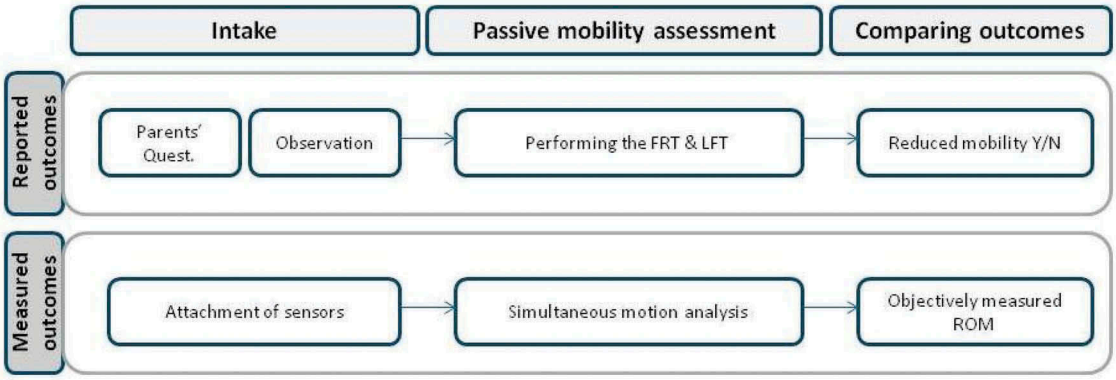


**Figure 1.** Mobility assessment procedure.

therapists were blinded for this questionnaire and parents were instructed to not share any details about their child with them. Meanwhile, the motion sensors were attached to the infant's forehead, and trunk by FD. Before the mobility assessment, signs of asymmetry of head, face, and trunk were observed by therapists. To minimize potential distress in the infant, only the first assessing manual therapist assessed active cervical mobility and verified whether there were no contra-indications preventing further study participation. The side-tilt-test was used to provoke active lateral flexion of the head and test whether the provoked response was comparable bilaterally. Active rotation was facilitated using sounds, toys, or the presence of the parents. Both tests are part of the normal clinical screening in infants [5,19,20].

For passive upper cervical mobility assessment, the FRT and LFT were performed in supine with low-amplitude and low-velocity. For the FRT, the infant's cervical spine was passively maximally flexed and carefully rotated. For the LFT, the infant's head was passively laterally flexed. Both tests were performed on both sides (see Table A1). If resistance was encountered or ROM limited before the expected end-point, reduced upper cervical mobility was indicated [21,22]. When at least one of these tests indicated reduced mobility, UCD and indication for further treatment were assumed [5]. On a standardized form (see Appendix), manual therapists reported on reduced mobility (yes/no), presence of UCD (yes/no), indication for further treatment (yes/no), and other observations, such as resistance or side effects. Manual therapists were blinded for each other's assessment performance and reported outcomes until both therapists completed the assessment. FD was always present during mobility assessment for study coordination and sensor registration. FD was not one of the participating manual therapists.

After study procedures were completed, FD shared the parents' questionnaire with the primary therapist to gain anamnestic information, and checked whether manual therapists disagreed on the reported 'presence of UCD' and 'indication for further treatment'. If so, discrepancy was discussed between therapists. Thereafter, the primary therapist informed parents on the assessment's findings and further possible steps in treatment, if indicated. These procedures fell outside the study's scope.

### Measures

ROM was assessed by manual therapists (subjective, dichotomous outcomes) and measured by inertial motion sensors (objective outcomes) simultaneously. Each infant was assessed by two manual therapists, bilaterally, leading to a total of 72 measurements of both FRT and LFT. Information from the intake was used to describe the study population. Primary clinical outcome measures were the FRT and the LFT. The reported outcomes (reduced mobility yes/no) were used to determine agreement and inter-rater reliability between manual therapists. Moreover, based on the outcomes of these subjective tests, manual therapists decided on the presence of UCD (diagnosis) and treatment indication (yes/no). To assess ROM, two wireless sensors were placed on the infants' forehead and trunk (sternum) using soft bands (Image A.1). ROM was recorded in three dimensions: sagittal plane (e.g. flexion), frontal plane (e.g. lateral flexion), and transversal plane (e.g. rotation). The sensors were connected to a laptop, which simultaneously recorded and registered all 3D outcomes of the head relative to the trunk. The primary author (FD) checked the recording of the sensors during testing to ensure adequate detection of motion. Both manual therapists reported to FD when they started (start-point) and ended (end-point) a movement and to which side it was performed. This information was necessary for data verification and analysis to confirm the position of the head and trunk on time points, and to calculate ROM. Manual therapists were blinded for all sensor outcomes. The objective sensor data were used to determine the degree of agreement on measured ROM between two manual therapists (inter-rater reliability).

### Kinematic analysis

After all subjective measurements of mobility by manual therapists were completed, the sensor data were visually checked by FD for the reported time-points in MT Manager 4.6 (Xsens Technologies BV). Data were converted and analyzed in MATLAB (version 2017b, The MathWorks BV, Natick, USA) by the second author (NK) to allow additional analysis of objective ROM and degrees of angles.

First, degrees of mobility were defined based on reported time-points and specific motion analysis of that time-point. This resulted in the assessment of the ROM as measured on four time-points: starting position (start-point) before the execution of a test to the right side (T1), end position (end-point) of the movement to the right (T2), starting position before the execution of a test to the left side (T3), and end position of the movement to the left (T4). Second, for each infant in each measurement, the mean start-point was calculated by averaging T1 and T3, because of possible displacement of the sensor during the assessment due to the moving of the infant. Third, ROM to both sides was determined by calculating the difference between the mean start-point of a movement and the maximum ROM to a particular side around the time point of the end-point. For the Flexion-Rotation-Test, data were extracted from both the sagittal and transversal

plane. For the Lateral-Flexion-Test, only ROM measured in the frontal plane was extracted.

### Statistical analysis

Characteristics of the study population were analyzed using descriptive statistics. To exclude a potential order effect of measurements, objectively measured ROM was compared between the first and second measurement within an infant, for each test and each side, using a paired t-test. Because no order effect was found, data were grouped together per measurement. All analyses were performed using SPSS Statistics version 25 (SPSS, Chicago Iln., USA). Additional information about the kinematic analysis is presented in the Appendix.

### Subjective outcomes

To determine the inter-rater reliability of subjectively, therapist-reported outcomes on the FRT and LFT, and reported conclusions related to diagnosis and treatment indication, 2 × 2 tables were created to compare the outcomes between manual therapists. Then, the proportion of observed positive and negative agreement [23] between manual therapists and Cohen's kappa were calculated. The following indications for agreement were used: <0 ('poor'), 0–0.20 ('slight'), 0.21–0.40 ('fair'), 0.41–0.60 ('moderate'), 0.61–0.80 ('substantial'), >0.81 ('almost perfect') [24]. These analyses were performed between all three pairs of manual therapists.

### Objective outcomes

To determine the inter-rater reliability of objectively measured ROM, the absolute mean differences in ROM between manual therapists were calculated. Additionally, the intraclass correlation coefficient (ICC) and 95%-confidence interval were calculated per test by using a two-way random effect consistency model. These analyses were performed pair-wise between manual therapists (i.e. A vs B, B vs C, and A vs C). To examine if there were systematic differences in ROM between manual therapists, a one-sample t-test was performed, Bland Altman plots were created for both the FRT and LFT and mean differences of ROM between two manual therapists were plotted against the means in ROM of these two manual therapists, and limits of agreement (LOA) were calculated [25].

### Relationship subjective and objective outcomes

To determine the relationship between subjectively reported outcomes and objectively measured ROM, the data reported by manual therapists and the objectively measured ROM by the sensors were compared.

First, for each therapist, the mean ROM which was indicated as 'reduced' mobility and mean ROM indicated as 'not-reduced' mobility was calculated, for both the FRT and LFT. Per therapist, the differences in ROM between 'reduced' and 'not-reduced', including standard error of difference, were calculated using a paired t-test. Second, outcomes per manual therapist indicated as 'reduced' or 'not-reduced' mobility were plotted in figures to get more insight in differences between manual therapists, and between mobility reported as 'reduced' and 'not-reduced'.

## Results

During the recruitment period, 95 potentially eligible infants were registered at the three participating practices, of which 36 infants (38%) participated in the study. Reasons for exclusion were: infants' age, parents who did not want to participate or infants had received treatment before. Characteristics of the 36 included infants are shown in Table 1.

All infants were referred to (44%) or administered through direct access (56%) at the practice for manual therapy with indications of asymmetry and presence of UCD. The most reported complaints or symptoms by parents, besides asymmetry, were restlessness/anxiety (42%) and excessive crying (31%). The majority of parents (61%) reported more than one complaint or symptom. In most infants, multiple signs of asymmetry were observed by manual therapists, where positional preference of the head (61%), asymmetrical shape of the head (47%), and an asymmetric or hyperextended trunk (64%) were observed most frequently. Active lateral flexion and rotation of the head were frequently reported by manual therapists as reduced. The majority of parents reported complications in delivery (58%). No contra-indications for study participation were reported by parents nor by manual therapists.

### Inter-rater reliability

Passive mobility assessment was performed in all 36 infants. Due to distress during the assessment, the FRT could not be tested consistently in two infants leading to data availability of 34 infants and a total of 68 measurements. The LFT was performed in all 36 infants, leading to a total of 72 measurements. No side effects or harms during mobility assessment, besides crying, were reported.

### Reported outcomes

Inter-rater reliability and agreement between pairs of manual therapists are presented in Table 2. For the FRT, inter-rater reliability between pairs of manual therapists ranged from slight to substantial (κ = 0.195–0.657). The proportion of agreement between pairs of manual therapists ranged from 0.57 to 0.86. For the LFT, inter-rater reliability between pairs of manual therapists ranged from poor to substantial (κ = −0.077–0.727). The proportion of agreement between pairs of manual therapists ranged from 0.46 to 0.86. Inter-rater reliability on

**Table 1.** Characteristics and posture observations of included infants (n = 36).

| | |
|---|---|
| **Demographics** | |
| Age infant (weeks, mean, SD) | 10.4 (6.5) |
| Sex (boys, number, %) | 18 (50%) |
| Birth weight (grams, mean, SD) | 3510 (520) |
| Duration pregnancy (weeks, mean, SD) | 39.2 (1.6) |
| **Pregnancy and delivery** | |
| Position *in utero* | |
| – Normal | 26 (72%) |
| – Occipital position | 1 (3%) |
| – Breech position | 4 (11%) |
| – Stargazing | 1 (3%) |
| – Unknown | 4 (11%) |
| Delivery | |
| – Spontaneous/natural | 22 (61%) |
| – Induced labor | 14 (39%) |
| Complicated delivery (epidural, cesarean, vacuum delivery and/or pulling the infant's head) | 21 (58%) |
| Complications during birth (hemorrhage, umbilican cord around infant's neck, meconium in amniotic fluid) | 15 (42%) |
| **Complaints and/or symptoms* reported by parents** | **Number (%)** |
| Excessive crying | 11 (31%) |
| Restlessness/anxiety | 15 (42%) |
| Grabbing the head or ears | 9 (25%) |
| Sleeping problems | 10 (28%) |
| Reflux | 10 (28%) |
| Disliking prone position | 9 (25%) |
| Breastfeeding problems | 6 (17%) |
| Problems with defecation | 4 (11%) |
| **Observation of posture and mobility reported by manual therapists** | **Number (%)** |
| Position of the head *(missing n = 2 (6%))* | |
| – Normal | 12 (33%) |
| – Positional preference | 22 (61%) |
| Shape of the head *(missing n = 3 (8%))* | |
| – Normal | 9 (25%) |
| – Asymmetry (plagiocephaly) | 17 (47%) |
| – Flattening (brachycephaly) | 6 (17%) |
| – Pointy head (dorsal) | 1 (3%) |
| Position of the trunk* *(missing n = 3 (8%))* | |
| – Normal | 10 (28%) |
| – Hyperextension | 9 (25%) |
| – Asymmetry | 14 (39%) |
| Reduced active lateral flexion of the head (positive side-tilt test) | |
| – To the right side | 11 (31%) |
| – To the left side | 5 (14%) |
| – To both sides | 2 (5%) |
| – Not adequate to test** | 18 (50%) |
| Reduced active rotation of the head | |
| – To the right side | 9 (25%) |
| – To the left side | 12 (33%) |
| – To both sides | 6 (17%) |
| – Not adequate to test** | 7 (19%) |

*multiple answers were allowed, **the side-tilt-test could not be adequately performed because of the age of infants (<3 months). Active rotation could not be tested because of too much crying before the test was performed, infants could not be provoked in following and turning their head or were asleep.

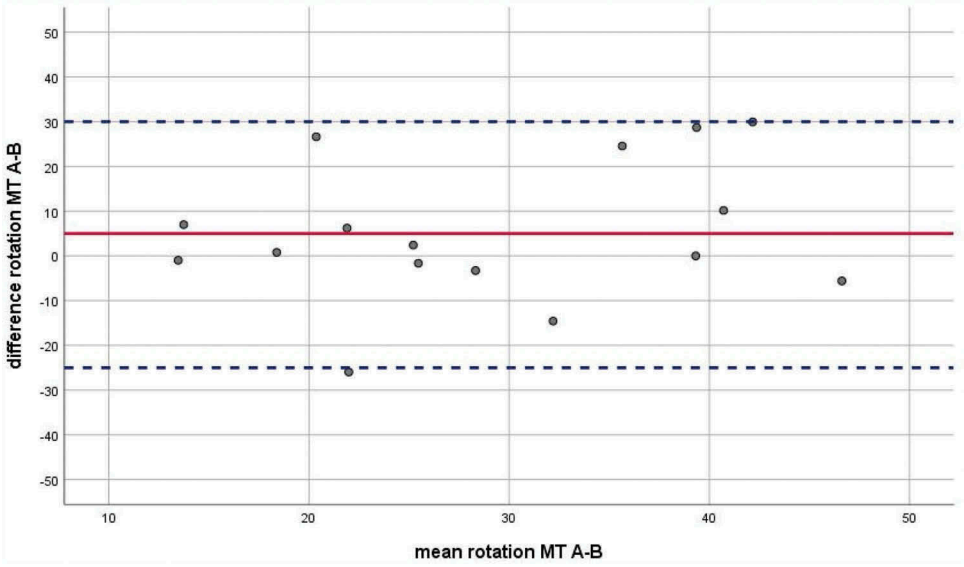


**Figure 2.** Example of Bland Altman plot of mean differences between manual therapists in degrees and limits of agreement.

**Table 2.** Inter-rater reliability and agreement of subjectively reported outcomes between pairs of manual therapists.

| | Manual Therapist (MT) A | | Manual Therapist (MT) B | | | MT A-B | MT B-C | MT A-C |
|---|---|---|---|---|---|---|---|---|
| Flexion Rotation Test | | | | | | | | |
| Manual Therapist (MT) B | *Reduced* | *Not-reduced* | | | Number of measurements | 22 | 24 | 22* |
| *Reduced* | 11 | 2 | | | Kappa | 0.624 | 0.657 | 0.195 |
| *Not-reduced* | 2 | 7 | | | 95% CI (kappa) | 0.292; 0.956 | 0.351; 0.963 | 0; 0.588 |
| Manual Therapist (MT) C* | *Reduced* | *Not-reduced* | *Reduced* | *Not-reduced* | Proportion observed agreement | 0.82 | 0.83 | 0.59 |
| *Reduced* | 7 | 3 | 12 | 2 | Proportion positive agreement | 0.86 | 0.86 | 0.61 |
| *Not-reduced* | 6 | 6 | 2 | 8 | Proportion negative agreement | 0.78 | 0.85 | 0.57 |
| Lateral Flexion Test | | | | | | | | |
| Manual Therapist (MT) B | *Reduced* | *Not-reduced* | | | Number of measurements | 22 | 24 | 26 |
| *Reduced* | 9 | 1 | | | Kappa | 0.727 | 0.510 | −0.077 |
| *Not-reduced* | 2 | 10 | | | 95% CI (kappa) | 0.442; 1 | 0.189; 0.832 | |
| Manual Therapist (MT) C | *Reduced* | *Not-reduced* | *Reduced* | *Not-reduced* | Proportion observed agreement | 0.86 | 0.75 | 0.46 |
| *Reduced* | 7 | 6 | 10 | 1 | Proportion positive agreement | 0.86 | 0.77 | 0.50 |
| *Not-reduced* | 8 | 5 | 5 | 8 | Proportion negative agreement | 0.87 | 0.73 | 0.53 |

*MT A and MT C did not perform the FRT in two infants because of too much crying.
MTs A, B & C tested pair-wise 36 infants: two tests to both sides for each infant. MT A and MT B tested 11 infants (22 measurements); MT B and MT C 12 infants (24 measurements) and MT A and MT C 13 infants (26 measurements).

reported diagnosis and treatment indication was high; agreement was found in, respectively, 34 (94%) and 35 (97%) infants. If passive mobility was reduced to at least one direction or side, diagnosis of UCD and an indication for further treatment were reported by manual therapists.

### *Objectively measured ROM*

The inter-rater reliability of objectively measured ROM toward both flexion-rotation and lateral flexion varied between poor and moderate (Table 3). Measurements between manual therapists showed large variation and LOA were wide (Figure 2). Absolute mean differences within pairs of manual therapists were minor, while the range in mean differences was wide (Table 3). No systematic differences between manual therapists in measured ROM were found. All Bland Altman plots are shown in Figure A3.

In this Bland Altman plot, the mean differences in ROM between manual therapist A and B are plotted against the means in ROM of these two manual therapists. The red line indicates the mean difference of objectively measured ROM (in degrees) between manual therapists A and B. The blue dashed lines indicate the upper and lower limits of agreement (LOA). The small mean difference indicates no systematic difference in measured ROM between manual therapists. The wide LOA indicates large discrepancies in ROM between manual therapists.

### *Relationship subjectively reported outcomes and objectively measured ROM*

The mean ROM was significantly smaller in measurements indicated as 'reduced' mobility by manual therapists than in measurements that were indicated as 'not-reduced' mobility (Table 4). As shown in Figure 3 there is an overlap in outcomes indicated as 'reduced' and 'not-reduced' mobility.

## Discussion

Our study is the first to assess the inter-rater reliability of the FRT and LFT in infants in a clinical practice setting. The inter-rater reliability of the FRT and LFT on reported outcomes by manual therapists varied between poor and substantial among pairs of manual therapists. The inter-rater reliability on objectively measured ROM varied between poor and moderate among pairs of manual therapists. The assessed ROM varied widely within and between infants. Furthermore, we verified that ROM was statistically significantly smaller toward the side reported to be reduced in mobility by manual therapists, as compared to the not-reduced mobility side. This suggests that in infants with indications of UCD passive upper cervical mobility restrictions are present but variably measured.

**Table 4.** Objectively measured mean ROM reported as 'reduced' or 'not-reduced' mobility per manual therapist.

| | Manual therapist A | Manual therapist B | Manual therapist C |
|---|---|---|---|
| Flexion-Rotation-Test (FRT) | | | |
| Mean ROM reduced mobility (SD) | $29^0$ ($13^0$) | $23^0$ ($12^0$) | $27^0$ ($9^0$) |
| Mean ROM not-reduced mobility (SD) | $44^0$ ($12^0$) | $33^0$ ($11^0$) | $36^0$ ($12^0$) |
| Difference in ROM* (SED) | $15^0$ ($4^0$) | $10^0$ ($4^0$) | $9^0$ ($3^0$) |
| 95% Confidence interval of difference | [$7^0$; $24^0$] | [$3^0$; $19^0$] | [$3^0$; $16^0$] |
| p-value | <0.0001 | 0.012 | 0.007 |
| Lateral-Flexion-Test (LFT) | | | |
| Mean ROM reduced mobility (SD) | $30^0$ ($15^0$) | $24^0$ ($9^0$) | $21^0$ ($10^0$) |
| Mean ROM not-reduced mobility (SD) | $38^0$ ($8^0$) | $33^0$ ($9^0$) | $36^0$ ($8^0$) |
| Difference in ROM* (SED) | $8^0$ ($4^0$) | $9^0$ ($3^0$) | $15^0$ ($3^0$) |
| 95% Confidence interval of difference | [$0^0$; $15^0$] | [$3^0$; $15^0$] | [$9^0$; $22^0$] |
| p-value | 0.04 | 0.002 | <0.0001 |

ROM: range of motion, SD: standard deviation, SED: standard error of difference.
*Difference between mean ROM reported as reduced and mean ROM reported as not-reduced mobility.

Previous research on cervical mobility in infants with torticollis demonstrated a measurement error between raters of 5–$10^0$ [22]. In our study, in every test and in every pair of manual therapists, LOA were larger than the measurement error of $10^0$ (see Table 3), indicating a substantial discrepancy between manual therapists. Although manual therapists were instructed to move the infant's head toward the end-point in ROM, the large variation and disagreement between manual therapists within infants could indicate that the absolute end-point in ROM was not always reached. Furthermore, as shown in Figure 3, the degrees in ROM used as a potential cutoff point to conclude on either normal or reduced mobility, differed between infants within manual therapists. Manual therapists emphasized that they do not rely solely on the ROM to indicate reduced mobility, but also on the perceived feeling at the end-point, the infant's reaction and bilateral differences.

Agreement on reported outcomes of the FRT and LFT between manual therapists A and C was much lower and differences in ROM were larger compared to other pairs of manual therapists. Further analysis of the subgroup of this particular pair showed that the mean age of infants was significantly lower (8.3 weeks) than infants assessed by the other pairs (12.6 and 10.7 weeks). Moreover, during mobility assessment manual therapists reported observations that may have limited the assessment in 12 infants, 7 (58%) of them were assessed by manual therapists A and C (Table A2). Given this, these infants seemed to be more resistant, which could have led to increased muscle tension and therefore inadequate mobility assessment. We suggest that lower age and stronger reactions and resistance on the assessment by infants make it harder for manual therapists to (1) perform mobility assessment, and (2) interpret the test outcomes and draw conclusions; similarity of assessment is conditional to reach an agreement.

Therefore, the assessed ROM is largely dependent on the performance of the assessment and its interpretation, and the resistance of the infant. Hence, mobility assessment in these infants is difficult and needs special expertise. In line with our observations, recently published studies also highlight the challenge in reliability studies because repeated measures could result in distress in infants [26], and because of the experienced difficulty for therapists to interpret outcomes in infants [27]. Therapists could use visual inspection to assess cervical mobility in infants instead of using measurement instruments [28], but show no consistency and clarity in measurement and interpretation of outcomes regarding ROM [27]. However, reliable measurement instruments to assess cervical ROM

**Table 3.** Objectively measured ROM toward flexion-rotation and lateral flexion, between pairs of manual therapists.

| | Manual therapists A-B | Manual therapists B-C | Manual therapists A-C |
|---|---|---|---|
| | | Flexion Rotation Test | |
| Number of reliable measurements* | 17 | 14 | 20 |
| Mean difference in ROM (SD) | $11^0$ ($11^0$) | $9^0$ ($7^0$) | $13^0$ ($11^0$) |
| Range mean differences in ROM | $0^0$ – $30^0$ | $0^0$ – $27^0$ | $1^0$ – $45^0$ |
| Limits of agreement (lower; upper) | $-25^0$; $35^0$ | $-25^0$; $21^0$ | $-25^0$; $37^0$ |
| ICC | 0.380 | 0.496 | 0.086 |
| 95% confidence interval of ICC | [−0.078; 0.715] | [−0.026; 0.805] | [−0.305; 0.480] |
| | | Lateral Flexion Test | |
| Number of reliable measurements | 20 | 13 | 20 |
| Mean difference in ROM (SD) | $9^0$ ($6^0$) | $11^0$ ($10^0$) | $17^0$ ($10^0$) |
| Range mean difference in ROM | $1^0$ – $22^0$ | $1^0$ – $38^0$ | $0^0$ – $33^0$ |
| Limits of agreement (lower; upper) | $-14^0$; $23^0$ | $-22^0$; $32^0$ | $-31^0$; $44^0$ |
| ICC | 0.479 | −0.043 | −0.019 |
| 95% confidence interval of ICC | [0.085; 0.750] | [−0.416; 0.403] | [−0.416; 0.403] |

*Due to not reliable registered data with the inertial sensors, the number of reliable measurements differs from the number of total measurements performed by pairs of manual therapists.
ROM: range of motion, SD: Standard deviation, ICC: intraclass correlation coefficient.

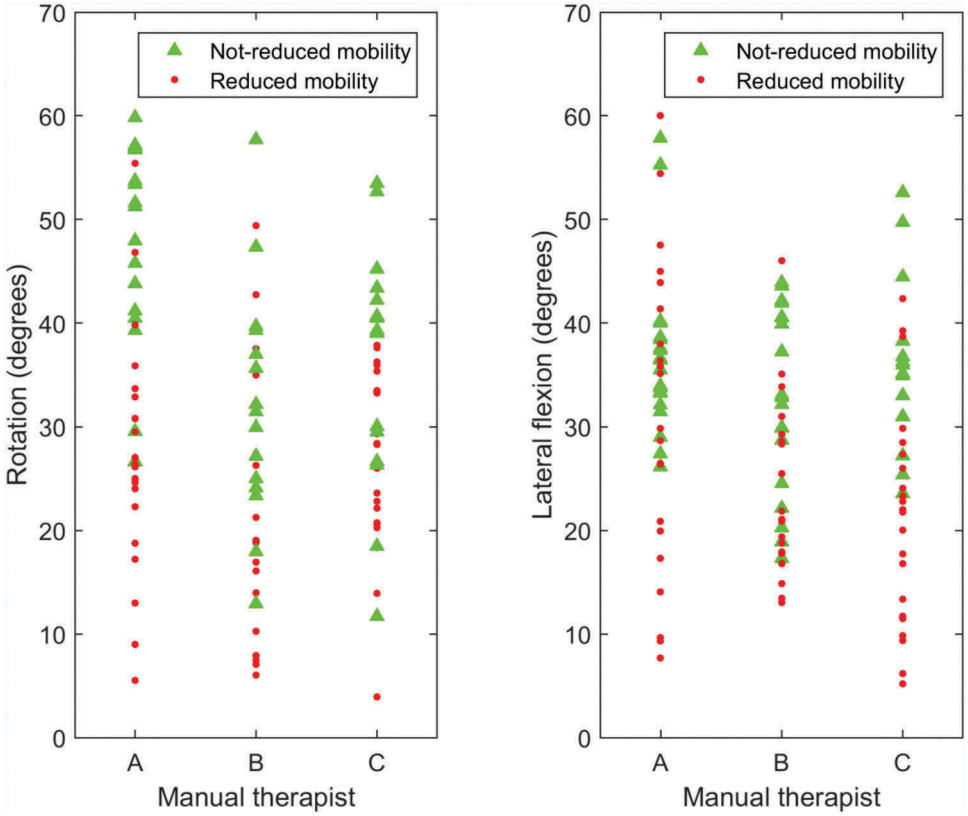


**Figure 3.** Range of motion indicated as 'reduced' and 'not-reduced' mobility, per manual therapist.

in infants are limited [26]. A previous study showed that the intra-rater reliability of the FRT and LFT in infants with torticollis was high (ICC 0.77 and 0.99, respectively) [15]. This could indicate that therapists have their own way of performing a test and could do this reliably by themselves. But when comparing the performance with another therapist, it becomes more difficult. On the other hand, infantile torticollis is a condition of the Sternocleidomastoid muscle leading to reduced active mobility, whereas infants with UCD have reduced passive mobility. In addition, studies assessing the validity of the FRT and LFT in the pediatric population are lacking. Validity of the FRT has only been indicated in adults; Takasaki et al. showed that the FRT predominantly and validly assesses upper cervical rotation in adults [13]. Whether these tests also validly assess upper cervical ROM in infants is however still unknown.

In our study, the range of ROM-outcomes reported as 'reduced' and 'not-reduced' mobility was wide and showed overlap. Possibly, individual cutoff points vary between manual therapists and manual therapists interpret ROM on a different moment in the movement. During the assessment, manual therapists were instructed to move the infants' head back to the start-point in between the measurements to the right and left side. In the sensor data analysis, we concluded that this start-point differed between manual therapists within an infant. Moreover, the calculations of ROM were based on this start-point, and objective measurements are therefore related to the position of the infant's head at the start of the movement. Hence, differences in start-points between manual therapists and not returning to the start-point could have influenced the measured ROM and therefore the mobility outcomes. This means that the interpretation of outcomes and the process of decision-making based on these tests are still unclear.

In contrast, agreement on the diagnosis of UCD and treatment indication was high between manual therapists. However, the presence of indications of UCD was an inclusion criteria for our study participants and could therefore be influenced by selection bias. In addition, the participating manual therapists in our study reported difficulties in clinical reasoning and getting the total picture of the infant because they were limited to execute a small number of tests and were not informed on infant's characteristics and parent-reported symptoms prior to the assessment. In clinical practice, manual therapists do have this information and pay more attention to the development and neuromuscular functions [19]. These reports indicate that only performing the FRT and LFT is not enough for manual therapists to interpret the outcomes and make clinical decisions. Hence, background information and more insight into the infants' neuromuscular functions are needed to optimize the value of performing the FRT and LFT and its interpretation.

### *Strengths and limitations*

Strengths of this study were the clinical practice setting, the use of motion sensors to assess ROM objectively to support the reported outcomes of the FRT and LFT, and blinding of manual therapists for each other's reported outcomes during mobility assessment and blinding for the motion analysis outcomes. In contrast to two-dimensional measures used in previous studies [26], we assessed mobility in three dimensions. Due to the use of two sensors we were able to subtract movements made in the infants' trunk and solely measure the cervical ROM. At the same time, a potential limitation was the possible measurement error due to the movement of the sensors if infants were restless, crying of moving during the assessment. We did not make video recordings, which limited us to draw conclusions on the execution of tests by manual therapists. Another important limitation was that both the subjective and objective outcomes were based on the same assessment, of the same manual therapist. This could have resulted in work-up bias. Moreover, because the Medical Ethical Committee did not approve to also include infants without UCD, all included infants in our study had indications of UCD (selection-bias). Furthermore, in the Netherlands, the use of imaging techniques, such as MRI, in infants without a life-threatening indication is forbidden. This prevented us from further validation of the FRT and LFT in infants.

## Conclusion

Inter-rater reliability of the FRT and LFT in infants with indications of UCD varied between poor and substantial and agreement on decision-making between manual therapists was high. Assessed ROM largely depends on the performance of the assessment and its interpretation by manual therapists, leading to high variation between therapists. Because of this high variation, the FRT and LFT cannot reliably assess reduced upper cervical mobility in infants with indications of UCD. Therefore, these tests should not be used solely as an outcome measure in clinical practice and in the research context.

## Acknowledgments

We gratefully thank the parents and infants for their willingness to participate in this study and the participating manual therapists E. Saedt, K. Giesen and S. Roode for their dedication and active participation in this study. We thank the Dutch Association for Manual Therapy (NVMT) for their support and funding. We thank E. Saedt and B. van der Woude for their continuous dedication to this study. At last, we thank A.T.M. van de Water for reviewing our manuscript thoroughly.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Dutch Association for Manual Therapy (NVMT).

## Notes on contributors

***Femke Driehuis*** is a physiotherapist and health scientist. She is currently completing her PhD in manual therapy in infants at the Radboud university medical center, Research Institute for Health Sciences, Scientific center for Quality in healthcare, Nijmegen. Besides finishing her PhD she works as policy advisor at the Guideline Department at the Royal Dutch Society for Physical Therapy (KNGF). Here she focuses on evidence based practice physiotherapy, guidelines and professional expertise and competences.

***Noël L.W. Keijsers*** After his PhD and since 2005, Noël Keijsers works a human movement scientist at the Sint Maartenskliniek, Institute Research, Development & Education. Since 2017 he is the project leader of the RevaResearch team and focuses on biomechanics during walking of patients from Rehabilitation and Orthopedics Departments.

***Maria W.G. Nijhuis–van der Sanden*** worked as Pediatric Physical Therapist between 1974 and 2011. In 2003 she got her PhD and between 2009 and 2019 she worked as professor in allied health sciences (physiotherapy, occupational therapy, speech therapy) at the Radboud university medical center, Research Institute for Health Sciences, Scientific center for Quality in healthcare and the department of Rehabilitation. Qualitative and quantitative research methodologies are used to design, implement and monitor efficient and effective allied health care interventions. Knowledge on the underlying mechanisms of motor learning is used to increase the effect of remedial therapy within physiotherapy, speech therapy and occupational therapy. Important themes are: the patient as active participant in the rehabilitation process, behavioral change of professionals to use evidence based reasoning and to cope with patient preferences in the same time, and development of reliable outcome measurements.

***Rob A. De Bie*** is a professor of Physiotherapy Research at Maastricht University. His expertise is in systematic review methodologies and professional guidelines and is known for his contributions to the Cochrane Back Review Group, which coordinates international literature reviews of primary and secondary prevention and treatment of neck and back pain and other spinal disorders. He has also provided support to the OTseeker and PEDro projects – databases that contain abstracts of systematic reviews and randomized controlled trials relevant to occupational and physical therapy.His research focusses on musculoskeletal disorders and relevant co-morbidities, while he teaches mainly clinical epidemiology at the Faculty of Health, Medicine and Life sciences, Maastricht University. He is coordinator for academic skills in the medicine curriculum.

***J. Bart Staal*** is a graduation as human movement scientist in 1996, he worked as a physiotherapist both in Germany and the Netherlands. In 1998, he started working as a researcher

at the Institute for Research in Extramural Medicine (EMGO-Institute) of the Vrije Universiteit in Amsterdam, where he conducted a PhD study on the effectiveness of a graded activity intervention for workers who were sick-listed due to low back pain. In 2003 he obtained his PhD-degree with a thesis entitled: 'low back pain, graded activity and return to work'. Beside his research activities, he also attended statistical and methodological courses in the Netherlands and USA. Subsequently, he started working as a lecturer and senior researcher at the Department of Epidemiology and the Center of Evidence Based Physiotherapy at Maastricht University. He taught evidence based medicine, statistics and implementation research in several bachelor and master courses both to medical and health sciences students. Currently he works as assistant professor at Radboud university medical center and HAN university of Applied Sciences, Research group Musculoskeletal Rehabilitation. His research themes include: musculoskeletal disorders, guideline development, cardiac rehabilitation, cancer rehabilitation, sports injuries, and physiotherapy in general.

***Thomas J. Hoogeboom*** Since dr. Hoogeboom finished his in PhD (2012), he has worked at the University of Colorado (2012), Maastricht University Medical Center+/TNO (2012-2015) and Radboudumc (2015-present). As a physiotherapist and biomedical scientist, he is working with his team to study the merit of physiotherapy for people with chronic conditions. He and his team proceed from the assumption that to do clinically-relevant and -impactful research, researchers need to embrace the complexity of daily (physiotherapy) practice and need to study personalized, rather than protocolized, interventions.

## ORCID

Femke Driehuis http://orcid.org/0000-0001-6774-5646
Noël L.W. Keijsers http://orcid.org/0000-0001-7835-7112
Maria W.G. Nijhuis–van der Sanden http://orcid.org/0000-0003-2637-6877
Rob A. De Bie http://orcid.org/0000-0001-5882-9303
J. Bart Staal http://orcid.org/0000-0002-0083-6380
Thomas J. Hoogeboom http://orcid.org/0000-0003-2103-419X

## Ethical approval

Approved by the Medical Ethical Committee for Research Involving Human Subjects of the Radboudumc Nijmegen (CMO, NL.58488.091.16).

## References

[1] Gotlib A, Rupert R. Chiropractic manipulation in pediatric health conditions–an updated systematic review. Chiropr Osteopat. 2008;16:11.
[2] Gleberzon BJ, Arts J, Mei A, et al. The use of spinal manipulative therapy for pediatric health conditions: a systematic review of the literature [review]. J Can Chiropr Assoc. 2012;56(2):Onlineaccess only p 128–41.
[3] Ferrance RJ, Miller J. Chiropractic diagnosis and management of non-musculoskeletal conditions in children and adolescents. Chiropr Osteopat. 2010;18:14.
[4] Allen-Unhammer A, Wilson FJ, Hestbaek L. Children and adolescents presenting to chiropractors in Norway: National Health Insurance data and a detailed survey. Chiropr Man Therap. 2016;24:29.
[5] Saedt E, Driehuis F, Hoogeboom TJ, et al. Nijhuis-van der Sanden MWG. Common manual therapy practices in the Netherlands for infants with upper cervical dysfunction: a prospective Cohort study. J Manip Physiol Ther. 2018;41(1):52–61.
[6] Biedermann H. Manual therapy in children: proposals for an etiologic model. J Manip Physiol Ther. 2005;28(3):2–16.
[7] Slate RK, Posnick JC, Armstrong DC, et al. Cervical spine subluxation associated with congenital muscular torticollis and craniofacial asymmetry. Plast Reconstr Surg. 1993;91(7):1187–1195.
[8] Sacher R, Seifert I, Martin M, et al. Empfehlungen zur manuellen Untersuchung und Behandlung von Säuglingen mit funktionellen Störungen der sensomotorischen Organisation und ohne Vorliegen systemischer Grunderkrankungen. Man Med. 2012;50:80–81.
[9] Biedermann H. Das KISS-Syndrom der Neugeborenen und Kleinkinder. Manuel Med. 1993;31:97–107.
[10] Hall TM, Robinson KW, Fujinawa O, et al. Intertester reliability and diagnostic validity of the cervical flexion-rotation test. J Manip Physiol Ther. 2008;31(4):293–300.
[11] Schafer A, Ludtke K, Breuel F, et al. Validity of eyeball estimation for range of motion during the cervical flexion rotation test compared to an ultrasound-based movement analysis system. Physiother Theor Pr. 2018;34(8):622–628.
[12] Ogince M, Hall T, Robinson K, et al. The diagnostic validity of the cervical flexion-rotation test in C1/2--related cervicogenic headache. Manual Ther. 2007;12(3):256–262.
[13] Takasaki H, Hall T, Oshiro S, et al. Normal kinematics of the upper cervical spine during the Flexion-Rotation Test: *in vivo* measurements using magnetic resonance imaging. Manual Ther. 2011;16(2):167–171.
[14] Budelmann K, von Piekartz H, Hall T. Is there a difference in head posture and cervical spine movement in children with and without pediatric headache? Eur J Pediatr. 2013;172(10):1349–1356.
[15] Klackenberg EP, Elfving B, Haglund-Akerlind Y, et al. Intra-rater reliability in measuring range of motion in infants with congenital muscular torticollis. Adv Physiother. 2005;7(2):84–91.
[16] Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96–106.
[17] Dutch register for specialized child manual therapists [Internet]. 2019.
[18] Koch LE, Graumann-Brunt S. Kopfgelenk-induzierte Symmetriestörungen und deren Folgenpathologien. Manuel Med. 1999;37:67–78.
[19] Coenen W. Bewegungsstörungen im Säuglingsalter. Manuel Med. 2011;49:171–188.
[20] Sables-Baus S, Robinson MV. Pediatric neurologic exam. Int Emerg Nurs. 2011;19(4):199–205.
[21] Hall T, Robinson K. The flexion-rotation test and active cervical mobility–a comparative measurement study in cervicogenic headache. Manual Ther. 2004;9(4):197–202.
[22] Ohman AM, Beckung ER. Reference values for range of motion and muscle function of the neck in infants. Pediatr Phys Ther. 2008;20(1):53–58.
[23] de Vet HC, Mokkink LB, Terwee CB, et al. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.

[24] Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33(2):363–374.
[25] Bland JM, Altman DG. Statistical method for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–310.
[26] Seager A, French H, Meldrum D. Measurement properties of instruments for assessment of cervical spine function in infants with torticollis: a systematic review. Eur J Pediatr. 2019;178(5):657–671.
[27] Oledzka MM, Kaplan SL, Sweeney JK, et al. Interrater and intrarater reliability of the congenital muscular torticollis severity classification system. Pediatr Phys Ther. 2018;30(3):176–182.
[28] Seager A. Assessment and management of torticollis in infancy: a survey of current physiotherapy practice. Physiother Pract Res. 2017;38:59–66.