

Measurement properties of instruments for assessment of cervical spine function in infants with torticollis: a systematic review.

AUTHOR(S)

Anthea Seager, Helen French, Dara Meldrum

CITATION

Seager, Anthea; French, Helen; Meldrum, Dara (2019): Measurement properties of instruments for assessment of cervical spine function in infants with torticollis: a systematic review.. figshare. Journal contribution. <https://hdl.handle.net/10779/rcsi.10798991.v1>

HANDLE

[10779/rcsi.10798991.v1](https://hdl.handle.net/10779/rcsi.10798991.v1)

LICENCE

CC BY-NC-SA 4.0

This work is made available under the above open licence by RCSI and has been printed from <https://repository.rcsi.com>. For more information please contact repository@rcsi.com

URL

https://repository.rcsi.com/articles/Measurement_properties_of_instruments_for_assessment_of_cervical_spine_function_in_infants_with_torticollis_a_systematic_review_/10798991/1

1 Title Page

2 Measurement properties of instruments for assessment of cervical spine function in infants with
3 torticollis: a systematic review.

4 Anthea Seager¹ anthea.seager@cuh.ie

5 Helen French² hfrench@rcsi.ie

6 Dara Meldrum³ meldrumd@tcd.ie

7

8 Corresponding author: Anthea Seager, Tel: 00353 1 8784563, Fax: 00353 1 8748355

9 Author roles: AS, HF and DM conceived the research idea and developed the protocol; AS did the
10 database searches; AS and HF screened studies for eligibility; AS performed data extraction; AS and
11 HF performed COSMIN assessment and DM arbitrated in the case of disagreement; AS drafted the
12 manuscript; All authors contributed to the manuscript and approved the final version for
13 publication.

1 Department of Physiotherapy, Temple Street Children's University Hospital, Dublin 1, Ireland
2 Department of Physiotherapy, Royal College of Surgeons in Ireland, St Stephen's Green, Dublin 2,
Ireland
3 Department of Physiotherapy, Dublin University, Trinity College, Dublin 2, Ireland

14 What is known?

- 15 • A thorough assessment of the infant presenting with torticollis is essential, in order to
16 correctly diagnose, rule out 'red flags' and manage appropriately
- 17 • Assessment tools need to have robust measurement properties in order to be of value for
18 clinical practice and research

19 What is new?

- 20 • This systematic review identified two valid and reliable tools for the assessment of cervical
21 spine function in infants with torticollis
- 22 • Further research is required to assess the measurement properties of tools already
23 described in the literature and to develop further tools for use in infants with torticollis

24 Abstract

25 The aim of this study was to systematically review the measurement properties of instruments
26 which assess cervical spine function in infants with torticollis. Electronic searches were performed in
27 MEDLINE, CINAHL, Embase, Web of Science and the Cochrane Library, combining three constructs
28 (“torticollis”, “cervical spine assessment” and “measurement properties”). Two reviewers
29 independently rated the methodological quality and the quality of measurement properties of
30 identified articles, using both the COSMIN (**CO**nsensus-based **S**tandards for the selection of health
31 status **M**easurement **I**Nstruments) checklist and quality criteria for measurement properties. Five
32 studies, using six instruments, met the inclusion criteria and were analysed. Included instruments
33 were the goniometer, electronic pendular goniometer, protractor, still photography, Muscle
34 Function Scale and a range of motion limitation scale. All studies assessed reliability and one study
35 also assessed content validity and hypothesis testing. The methodological quality of the studies
36 varied from poor to excellent according to the COSMIN checklist. Two instruments were found to
37 have good measurement properties from high-quality studies: still photography for the assessment
38 of habitual head tilt in supine and the Muscle Function Scale for the assessment of side-flexor
39 muscle function in lateral head righting.

40 **Conclusion** This systematic review identified two reliable tools for the assessment of cervical spine
41 function in infants with torticollis. Further research is required to assess the measurement
42 properties of tools already described in the literature and to develop further tools for use in this
43 population.

44 Abbreviations

45 CMT Congenital Muscular Torticollis

46 COSMIN **C**onsensus-based **S**tandards for the selection of health status **M**easurement **I**nstruments

47 CSp Cervical Spine

48 HRPRO Health-Related Patient-Reported Outcomes

49 MIC Minimal Important Change

50 PoT Postural Torticollis

51 ROM Range of Motion

52 Rot Rotation

53 SCM Sternocleidomastoid muscle

54 SDC Smallest Detectable Change

55 SEM Standard Error of Measurement

56 SF Side flexion

57 Introduction

58 Torticollis is a clinical sign of asymmetric neck function which may result from a variety of underlying
59 disorders. The main cause in infancy is muscular in nature [1-3] and can be divided into two clinical
60 subgroups of Congenital Muscular Torticollis or Postural Torticollis [4,5]. Congenital Muscular
61 Torticollis (CMT) results from fibrosis of the sternocleidomastoid (SCM) muscle, which may result in
62 a fibrotic lump or 'pseudo-tumour' [1]. Depending on diagnostic criteria, the incidence of CMT
63 ranges from 1.3% to 3.9% of new-borns [6,7]. Congenital Muscular Torticollis involves shortening or
64 excessive contraction of the SCM muscle with reduced cervical spine (CSp) ipsi-lateral rotation and
65 contra-lateral side flexion, and is associated with breech presentation, difficult labour and hip
66 dysplasia [8]. Postural Torticollis (PoT) results from asymmetry of the cervical musculature, without
67 any morphological changes, but secondary to asymmetric positioning and use. It is commonly
68 associated with plagiocephaly, which is cranial asymmetry due to external pressure [4,9-11]. Again,
69 incidence depends on diagnostic criteria, but has been reported as high as 16% of new-borns [12].

70 In both CMT and PoT, asymmetric neck function may be the most obvious clinical feature, but this
71 can have subsequent effects on other aspects of the infant's development, such as cranio-facial,
72 musculoskeletal and neuro-developmental effects [8,13-17]. Whilst muscular dysfunction is the
73 most common cause of torticollis in infancy, there are potentially up to 80 causes [18]. Therefore, a
74 thorough assessment of the infant presenting with torticollis is essential to correctly diagnose, rule
75 out 'red flags' and manage appropriately [19,20]. A number of interventions have been found to be
76 effective [20] but further evaluation is recommended, for which robust outcome measures are
77 indicated.

78 Although a variety of tools exist for the assessment of CSp function in adults [21-28] and in older
79 children [29-34], these are often not suitable for use in infants. Infants are more variable in
80 behaviour and movement, and less able to follow commands or tolerate equipment [20]. Many
81 studies investigating various aspects of torticollis have included CSp range of motion (ROM) as an

82 outcome measure but give limited detail about how the ROM was measured [35-42]. Others give
83 details of the assessment tool used, but measurement properties are not described [10,12,43-46] or
84 are described for an adult population [47].

85 Assessment techniques need to have robust measurement properties in order to be of value for
86 both research and clinical practice. The aim of this review was to systematically appraise the
87 evidence on measurement properties of assessment tools/techniques for the measurement of CSp
88 function in infants with torticollis.

89

90 [Methods](#)

91 This review is reported in accordance with the PRISMA (**P**referred **R**eporting **I**tems for **S**ystematic
92 **R**eviews and **M**eta-**A**nalysis) guidelines for the reporting of systematic reviews [48].

93 [Data Sources and Search Strategy](#)

94 A systematic search of the literature was performed by two independent reviewers (HF and AS).
95 Three groups of key words were used - one group relating to “torticollis”, the second relating to
96 “cervical spine assessment” and the third relating to “measurement properties”. See Appendix 1 for
97 the full list of search terms. The following electronic databases were searched from inception to
98 September 2016: PubMed, CINAHL (EBSCO), Embase, The Cochrane Library and Web of Science.
99 Google Scholar was searched for additional citations associated with final included studies. Authors
100 of relevant published abstracts were contacted to request a full-text version of the study and/or
101 study data. Reference lists of selected studies were hand searched for additional studies.

102 [Eligibility Criteria](#)

103 Studies were included if they were original studies, included infants aged 0-3 years, investigated one
104 or more measurement property of an assessment tool, the population of interest included torticollis
105 (diagnosed using recognised criteria) and were available full-text in a peer-reviewed publication.

106 Studies were excluded if they were case reports, were not written in the English language, involved
107 medical diagnostic tests, focussed solely on another aspect of torticollis (e.g. plagiocephaly) or
108 involved non-muscular causes of torticollis, such as neurological conditions or congenital bony
109 anomalies.

110 Data Extraction and Assessment of Methodological Quality

111 Data extraction and assessment of methodological quality were performed by two independent
112 reviewers (AS and HF). Any disagreements about study eligibility, data extracted, or the quality
113 scoring were discussed and, if possible, reconciled by mutual agreement. In the case of
114 disagreements, a third reviewer (DM) was available for arbitration.

115 The following data were extracted using a standardised data extraction form: study demographics,
116 details of the tool (name, element of CSp function being assessed and protocol for use), details of
117 the subjects and raters, measurement property being assessed and statistical analysis.

118 The COSMIN (**C**onsensus-based **S**tandards for the selection of health status **M**easurement
119 **I**Nstruments) checklist [49,50] was used to rate the methodological quality of the study designs. This
120 tool was developed as a consensus-based checklist to evaluate the methodological quality of studies
121 on measurement properties and has demonstrated high inter-rater agreement [49]. It assesses 12
122 measurement properties: internal consistency; reliability; measurement error; content validity;
123 construct validity (i.e. structural validity, hypothesis testing and cross-cultural validity); criterion
124 validity; responsiveness; interpretability; item response theory and generalisability. Each
125 measurement property is evaluated in a separate section, employing a number of questions
126 regarding items of design and statistical methodology, in order to assess the extent to which the
127 study meets the standard for good methodological quality. Four response options (excellent, good,
128 fair and poor) are defined for each question, with an overall quality score obtained by taking the
129 lowest rating of any item in a section i.e. “worst score counts” [50]. The COSMIN checklist was
130 initially developed for use on Health-Related Patient-Reported Outcomes (HRPOs), but its authors

131 indicated that it is also appropriate for use with other forms of measurement tools, although some
132 of the scoring criteria (e.g. on sample size) should be used as guidelines and should be at the
133 discretion of the rater [49,50].

134 The reviewers also attributed a separate rating to the quantitative results for each study, following
135 the quality criteria for measurement properties reported by Terwee et al. [51]. These criteria,
136 derived from existing guidelines and consensus, contain nine sections which apply a rating of
137 positive (+), negative (-), indeterminate (?) or no information available (0) to the measurement
138 properties: content validity; internal consistency; criterion validity; construct validity; reproducibility;
139 longitudinal validity; responsiveness; floor/ceiling effect and interpretability.

140

141 Results

142 The selection process and search yield are shown in Figure 1. Following title and abstract screening,
143 15 articles remained. Clarification was sought from three corresponding authors. Two authors
144 [52,53] had made reference to unpublished data and they provided further information regarding
145 this data. The pilot study from one author [52] was included, as a full text version was provided as
146 part of a Master of Philosophy thesis [54] and from the other [53] was excluded as it was presented
147 as a poster format, with no full-text available [55]. The third author [29] was contacted for
148 clarification on the age of subjects, and this study was subsequently excluded as it was not clear
149 whether or not the subjects met the age criteria for inclusion.

150 Of the remaining 12 full text articles assessed for eligibility, eight were subsequently excluded. Four
151 assessed CSp function as an element of a wider condition [56-59] and four described methods of
152 assessment of torticollis but not assessment of their measurement properties [8,10,47,60].

153 Therefore, a final selection of five studies [54,61-64] was included in this review. These five studies
154 examined measurement properties of six assessment tools for infants with torticollis: electronic

155 pendular goniometer (digital inclinometer) [54]; standard goniometer [61]; large protractor [61];
156 Muscle Function Scale [62]; still photography [61,63] and ROM limitation scale [64]. Details are
157 described in Table 1 and in Figures 2 – 7.

158 Of these five studies, all investigated reliability and one also described content validity and
159 hypothesis testing [62]. Therefore, three of the potential 12 COSMIN sections were used to assess
160 methodological quality. Quantitative results of three properties were rated using quality criteria
161 developed by Terwee et al. [51] (see Table 2).

162

163 **Figure 1**

164

165 **Table 1**

166

167 **Table 2**

168

169 **Figure 2**

170

171 **Figure 3**

172

173 **Figure 4**

174

175 **Figure 5**

176

177 **Figure 6**

178

179 **Figure 7**

180

181

182 Rating of Methodological Quality

183 Reliability

184 All five studies assessed reliability (see Table 3). Sample sizes of infants ranged from 23 to 109. In
185 two studies the sample was divided into two groups, so the actual sample sizes in those studies were
186 12+12 [54] and 11+12 [61]. Numbers of raters ranged from one to nine. Sample size is one of the
187 criteria for which the COSMIN checklist may be used more as a guideline and at the discretion of the
188 review author, rather than as one of the criteria on which to determine the “worst score” [49,50]. In
189 addition, the tool does not differentiate between the sample size of the subjects and that of the
190 raters in a reliability study. Therefore, in some systematic reviews of physical measures using the
191 COSMIN checklist, the sample size score has been excluded [65,66] and a similar approach was taken
192 in this review.

193 Administration was considered to be independent in four of the studies [54,62-64] as the examiners
194 were unaware of the previous assessments or there was an adequate time interval between
195 assessments to minimize risk of recall. In Klackenberg et al. [61] it was considered by the review
196 authors that the administration was not independent for Group 1, as they were re-measured
197 without any time interval, and doubtful for Group 2, as there was only a one-hour time interval and
198 considering that the sample was small, there was a chance that the examiner would recall previous
199 measurements.

200 In all five studies it was either stated or presumed that patients were stable between test
201 administrations and test conditions were similar for both test administrations. The studies scored
202 ‘good’ [54,61] or ‘excellent’ [62,63] in relation to reliability statistics, however Murgia et al. [64]
203 scored as ‘fair’ due to use of Kappa statistics instead of the preferred weighted Kappa.

204 Other minor methodological flaws in the design or execution were identified by the review authors.
205 In the study by Chen [54] only the healthy side was assessed, which meant that a limited range was
206 not assessed. Klackenberg et al. [61] did not specify whether the measurements were taken to the

207 nearest 1°, 5° or 10°. However, the illustration of the protractor which had been fabricated (see
208 Figure 3) suggested that it was in 10° intervals (0-90° bilaterally). In two of the studies measuring
209 CSp rotation, the infants' heads were not over the edge of the surface, so true end of range may not
210 have been achieved [61,64]. Murgia et al. [64] used scales with wide intervals for ROM. Range of
211 motion is usually measured in degrees, with means of 110° CSp passive rotation and 70° passive
212 side-flexion reported in infants [67]. In Murgia's scale [64], CSp rotation was divided into four
213 measurement levels and side-flexion into three measurement levels (see Figures 4 and 5). In
214 addition, CMT was one of the exclusion criteria, so only cases of PoT would have been included.
215 There were no important flaws in the other studies [62,63].

216 One reliability checklist was completed for each element of the reliability study e.g. for inter-rater
217 reliability and for intra-rater reliability (available as Online Resource). The quality scores (using
218 "worst score counts") allocated to the five reliability studies, both including and excluding that for
219 sample size, are described in Table 3.

220 [Content Validity and Hypothesis Testing](#)

221 One study assessed content validity [62] in which a panel of experts discussed the *Muscle Function*
222 *Scale* and provided insight, constructive criticism and advice. Most of the items in the checklist were
223 rated as '*excellent*' but the overall score was '*poor*' due to lack of details regarding how this was
224 done (see Table 4). It was also discussed by the review authors that still photographs were used to
225 assess the infants, whereas the MFS is a dynamic measure, as the position needs to be held by the
226 infant for ≥5 seconds in order to obtain a score. Therefore, the infants' head position was being
227 analysed but not the length of time the position was being held.

228 One study examined hypothesis testing [62]. A number of the items on the hypothesis testing
229 checklist are similar to that of the reliability checklist [49]. Specific hypotheses to be tested should
230 have been formulated and include expected differences between groups or expected correlations
231 between instruments, as well as expected direction and magnitude of the correlations *a priori*

232 (before data collection). This was unclear in the study by Öhman et al. [62], therefore it was given a
233 score of 'poor' (see Table 4).

234 **Table 4**

235 Rating of Quantitative Results

236 All five studies reported ICC or Kappa Values >0.7, consistent with a positive reliability rating score,
237 except for the side-flexion component of the ROM limitation scale [64] which had a negative
238 reliability rating (see Table 3). Agreement was reported in one study [61] in which measurement
239 error was expressed as the standard error of measurement (SEM), but minimal important change
240 (MIC) and smallest detectable change (SDC) were not reported. However, the study authors had
241 made a case that $\leq 6^\circ$ repeatability was clinically acceptable. Therefore, in that study [61] the
242 goniometer and protractor were given positive reliability rating scores and photography a negative
243 reliability rating score. The one study which reported content validity [62] was found to have a
244 positive rating. See Table 3 for details.

245

246 Discussion

247 This systematic review identified six assessment tools measuring CSp function in infants with
248 torticollis, for which measurement properties had been evaluated: standard goniometer, electronic
249 pendular goniometer, large protractor, a ROM limitation scale, the *Muscle Function Scale* and still
250 photography.

251 Goniometer

252 A goniometer is a routinely used tool and is widely available in the clinical setting. Reliability was
253 rated as positive (ICC>0.7) in the one study [61] which used it, assessing intra-rater reliability of
254 passive CSp rotation, but methodological quality rated as poor, primarily due to the short time
255 interval between administrations.

256 Electronic pendular goniometer

257 An electronic pendular goniometer (digital inclinometer) is also available for clinical use, although
258 more expensive than a standard goniometer. Reliability was rated as positive ($ICC > 0.7$) in the one
259 study [54] which used it, assessing inter-rater and intra-rater reliability of passive CSp rotation, but
260 methodological quality rated as fair. Two examiners and one assistant were required for the
261 assessment procedure, which may not be clinically feasible.

262 Protractor

263 A mat with protractor markings, which was assessed in one study [61] appeared relatively simple to
264 develop. Alternative large joint protractors are available commercially (e.g. Baseline 12-1076
265 arthrodiagonal goniometer; Reedco arthrodiagonal protractor). Reliability was rated as positive ($ICC > 0.7$),
266 assessing intra-rater reliability of passive CSp side-flexion, but methodological quality rated as poor,
267 again primarily due to the short time interval between administrations.

268 Range of motion limitation scale

269 This scale [64] assessed passive CSp rotation/side-flexion and active CSp rotation. The validity of the
270 scale was not tested. Reliability was rated as positive ($Kappa > 0.7$) for rotation and negative
271 ($Kappa < 0.7$) for side-flexion, but methodological quality rated as fair. One of the limitations of this
272 scale identified by the review authors was the wide categories of ROM, which may limit the clinical
273 utility of the scale. However, it should be noted that it was developed as a simple method of
274 detection of limitation of CSp ROM in infants with plagiocephaly (i.e. PoT) and not for precise
275 measurements.

276 Muscle Function Scale

277 This scale [62] assessed CSp side-flexor muscle function in lateral head righting, by rating the
278 position of the head when the infant's body was moved from a vertical to a horizontal position.
279 Regarding content validity, very little information was given about the methodology of testing,

280 however a panel of experts provided insight, constructive criticism and advice, with subsequent
281 alterations having been made. The reliability of the scale was rated as positive ($Kappa > 0.7$) and the
282 methodological quality of the study as excellent. This scale is free to use clinically.

283 Still photography

284 Still photography (with measurements subsequently assessed using lines drawn on the photograph)
285 was used in two studies [61,63]. Rahlin and Sarmiento [63] used it to assess habitual head deviation
286 from midline (head tilt) in supine. Reliability was rated as positive ($ICC > 0.7$) and the methodological
287 quality of the study as excellent. Klackenberg et al. [61] used it to assess CSp passive rotation and
288 side-flexion, by a photograph taken at the end of range position. In comparison to measurements
289 taken simultaneously with a goniometer/protractor, its reliability was found to be positive ($ICC > 0.7$),
290 but agreement between the two tools was poor and the study quality rated as fair. Still photography
291 is a simple, practical assessment tool. In one study [63] it was performed by one examiner, whereas
292 in the other [61] two assistants were required in addition to the examiner.

293 Implications for Practice and Research

294 Torticollis is a relatively common condition of infancy presenting to healthcare professionals
295 [6,7,12]. However, this systematic review has revealed only two measurement tools which have
296 both good measurement properties and study methodological quality. These include still
297 photography for the assessment of habitual head tilt in supine and the *Muscle Function Scale* for the
298 assessment of side-flexor muscle function in lateral head righting. When using an assessment tool,
299 the measurement properties of that tool should be considered, particularly if the tool is being used
300 as a decision-making outcome measure. There has been some inconsistency in the literature [67-73]
301 where reliability values from Klackenburg et al. [61] and Cheng et al. [8,52] (i.e. the pilot study by
302 Chen [54]) are referred to in the measurement of CSp ROM, but the tool used in the subsequent
303 studies does not appear to be the same as the ones in which the measurement properties were
304 originally evaluated. In these subsequent studies [8,52,67-73] an arthrodiagonal protractor is used to

305 measure passive CSp rotation and side-flexion, whereas the original reliability studies were
306 performed on an electronic pendular goniometer [54] and a mat with protractor markings [61].
307 Based on the COSMIN checklist, both of these studies [54,61] were found to be of poor/fair quality,
308 which raises issues regarding the validity of outcomes assessed in those subsequent studies.

309 Sample sizes tended to be small, scoring as poor in two [54,61] and fair in one [63] of the studies.
310 The number of raters was not taken into consideration by the COSMIN tool. It had been decided to
311 exclude sample size in the final quality score allocated, which did not change the overall score for
312 two of the studies [61,64] but improved the scores for the other three studies [54,63,68]. A
313 challenge exists in performing reliability studies with infants, as repeated measures and a larger
314 number of raters may cause distress to the infants.

315 A number of studies were not included in the systematic review but indicate potential for further
316 research [8,12,47,52,57,58,60,64,65,66]. Some of these studies were excluded as CSp function was
317 assessed as a section of an overall assessment for another condition: active CSp rotation in supine
318 and prone were assessed on a six-point scale as part of an Idiopathic Infantile Asymmetry Scale [57]
319 and head tilt in upright was assessed on a four-point scale as part of a Severity Assessment of
320 Plagiocephaly Scale [58]. There is potential for the CSp sections to be analysed separately. The
321 following excluded studies described methods of assessment of torticollis, but not assessment of the
322 measurement properties: Emery [47] described use of a modified goniometer (with a spirit level)
323 but measurement properties were not analysed. However, the reliability of similar instruments had
324 previously been tested in adults [74] and could be tested in infants. Stellwagen & Hubbard [12] used
325 graph paper under the heads of infants to assess passive CSp side-flexion, which could potentially
326 have its measurement properties assessed. Visual estimation was used in a number of studies [64-
327 66], which has been tested for reliability in adults [24,27,75] but not in infants. Studies have
328 reported that over 90% of paediatric physiotherapists use visual estimation to assess cervical
329 function in infants with torticollis [76-78]. Therefore, the measurement properties of visual

330 estimation should be studied further. Cheng et al. [8,52,60] used an overall severity scoring system
331 for torticollis, consisting of six objective and subjective categories, but measurement properties
332 were not described. However, subsequent to this systematic review, a study has been published [81]
333 which examined the reliability of a seven-grade severity system, the Congenital Muscular Torticollis
334 Severity Classification System (CMT-SCS) and found good inter-rater (ICC 0.83) and intra-rater (ICC
335 0.81) reliability (n=24 infants, 145 raters).

336 An assessment tool must be clinically practical as well as having adequate measurement properties
337 in order for it to be used, especially in the case of assessing infants [20,79]. One of the studies which
338 had been excluded from this review as it was only in abstract form [55] reported excellent reliability
339 (ICC 0.89-0.97) of 2D video analysis in the measurement of CSp active ROM in infants, with rotation
340 but not side-flexion being comparable to 3D motion analysis (considered the gold standard).

341 However, a subsequent study [53] found the method to be clinically unfeasible in terms of time
342 requirements. In the assessment of infants, it may be practical to have one assistant, as a parent is
343 usually present, but in two of the studies analysed, more assistants were required for use of the
344 electronic pendular goniometer [54] and photography [61].

345 Of the 12 sections of the COSMIN checklist, reliability was the most common measurement property
346 evaluated, with content validity and hypothesis testing tested in just one study. Other measurement
347 properties such as responsiveness and interpretability should be evaluated in future studies, if such
348 tools are used for intervention studies and clinical decision making.

349

350 Strengths and Limitations of the Review

351 This systematic review adds to other authors' works, which have identified tools used in the
352 assessment of CSp function in infants with torticollis, but not systematically reviewed their
353 methodological quality [20,80]. A strength of this review was the use of the COSMIN checklist for the

354 assessment of methodological quality of studies assessing psychometric properties of measurement
355 tools. Two independent review authors extracted data and assessed methodological quality.

356 The inclusion/exclusion criteria may have limited the number of articles included in this review, such
357 as those not in the English language. The review was limited to full-text studies, so other relevant
358 studies may have been excluded. Tools tested on infants with other types of torticollis (e.g. ocular)
359 may be applicable to muscular torticollis but were excluded.

360 The COSMIN checklist was developed for HRPROs, therefore its validity for assessing the
361 methodological quality of these clinometric studies can be questioned. It also generates a floor
362 effect in terms of its “lowest score counts”, so that studies which may have scored well on many
363 items will result in a ‘poor’ or ‘fair’ overall score due to one inadequate item, even when sample size
364 was excluded.

365

366 Conclusion

367 This systematic review has analysed five studies, in which measurement properties of six tools for
368 the assessment of CSp function in infants with torticollis were evaluated. The quality of the studies,
369 assessed using the COSMIN checklist, varied from poor to excellent. The only tools which were
370 found to have good measurement properties from a high-quality study were: still photography for
371 the assessment of habitual head tilt in supine and the *Muscle Function Scale* for the assessment of
372 side-flexor muscle function in lateral head righting. Further high-quality research is required to
373 investigate the measurement properties of assessment tools which have been described in the
374 literature, but not tested, and also to develop further tools for the assessment of CSp function in
375 this population.

376

377 Compliance with Ethical Standards

378 Funding: This study was funded by Temple Street Foundation (grant number 13.032).

379 Conflict of interest: All authors declare that they have no conflict of interest.

380 Ethical approval: This article does not contain any studies with human participants or animals
381 performed by any of the authors.

382

383

384 Conflict of interest

385 The authors declare no conflict of interest

386 References

- 387 [1] Cooperman D (1997) The differential diagnosis of torticollis in children. *Phys Occup Ther Paediatr*
388 17:1-11
- 389 [2] Nucci P, Kushner BJ, Seragino M (2005) A multi-disciplinary study of the ocular, orthopaedic and
390 neurological causes of abnormal head postures. *Am J Ophthalmol* 140:65-68
- 391 [3] Herman MJ (2006) Torticollis in infants and children: common and unusual causes. *Instr Course*
392 *Lect* 55:647-653
- 393 [4] van Vlimmeren LA, Helders P J, van Adrichem LN, Engelbert RH (2004) Diagnostic strategies for
394 the evaluation of asymmetry in infancy—a review. *Eur J Pediatr* 163:185-191
- 395 [5] Tatli B, Aydinli N, Çaliskan M, Ozmen M, Bilir F, Acar G (2006) Congenital muscular torticollis:
396 evaluation and classification. *Pediatr Neurol* 34:41-44
- 397 [6] Cheng JC, Au AW (1994) Infantile torticollis: a review of 624 cases. *J Ped Orthop* 14:802-808
- 398 [7] Chen MM, Chang HC, Hsieh CF, Yen MF, Chen TH (2005) Predictive model for congenital muscular
399 torticollis: analysis of 1021 infants with sonography. *Arch Phys Med Rehabil* 86:2199-2203
- 400 [8] Cheng J, Tang S, Chen TMK, Wong E (2000) The clinical presentation and outcome of treatment of
401 congenital muscular torticollis in infants - 1086 cases. *J Ped Surg* 35:1091-1096
- 402 [9] van Vlimmeren LA, Helders PJ, van Adrichem LN, Engelbert RH (2006) Torticollis and
403 plagiocephaly in infancy: therapeutic strategies. *Pediatr Rehabil* 9:40-46
- 404 [10] Aarnivala HE, Valkama AM, Pirttiniemi PM (2014) Cranial shape, size and cervical motion in
405 normal new-borns. *Early Hum Dev* 90:425-30
- 406 [11] de Chalain TM, Park S (2005) Torticollis associated with positional plagiocephaly: a growing
407 epidemic. *J Craniofac Surg* 16: 411-418
- 408 [12] Stellwagen L, Hubbard E (2008) Torticollis, facial asymmetry and plagiocephaly in normal new-
409 borns. *Arch Dis Child* 93:827-831
- 410 [13] Hylton N (1997) Infants with torticollis: the relationship between asymmetric head and neck
411 positioning and postural development. *Phys Occ Ther Ped* 17:91-117
- 412 [14] Yu C, Wong F, Lo L, Chen Y (2004) Cranio-facial deformities in patients with uncorrected
413 congenital muscular torticollis – an assessment from 3D imaging. *Plast Reconstr Surg* 113:24-33
- 414 [15] Tessner A, Mooney P, Pellard L (2010) A developmental perspective on congenital muscular
415 torticollis: a critical appraisal of the evidence. *Pediatr Phys Ther* 22:378-383
- 416 [16] Öhman AM (2009) Are infants with torticollis at risk of a delay in early motor milestones
417 compared with a control group of healthy infants? *Dev Med Child Neurol* 51:545-550
- 418 [17] Schertz M, Luk L, Green D (2013) Long term neurodevelopmental follow up of children with
419 congenital muscular torticollis. *J Child Neurol* 28:1215-1221
- 420 [18] Kiwak KJ (1984) Establishing an etiology for torticollis. *Postgrad Med* 75:126-134
- 421 [19] Nuysink J, van Haastert IC, Takken T, Helders, P (2008) Symptomatic asymmetry in the first six
422 months of life: differential diagnosis. *Eur J Pediatr* 167:613-619

- 423 [20] Kaplan S, Coulter C, Feters L (2013) Physical therapy management of congenital muscular
424 torticollis: an evidence-based clinical practice guideline. *Pediatr Phys Ther* 25:348-394
- 425 [21] Antonaci F, Ghirmai S, Bono G, Nappi G (2000) Current methods for cervical spine movement
426 evaluation: a review. *Clin Exp Rheumatol* 18:S-45
- 427 [22] Fortin C, Ehrmann Feldman D, Cheriet F, Labelle H (2011) Clinical methods for quantifying body
428 segment posture: a literature review. *Disabil Rehabil* 33:367-383
- 429 [23] Schaufele MK, Boden SD (2003) Physical function measurements in neck pain. *Phys Med Rehabil*
430 *Clin N Am* 14:569-588
- 431 [24] Youdas, JW, Carey JR, Garrett TR (1991) Reliability of measurements of cervical spine range of
432 motion—comparison of three methods. *Phys Ther* 71:98-104
- 433 [25] Jordan K (2000) Assessment of published reliability studies for cervical spine range-of-motion
434 measurement tools. *J Manip Physiol Ther* 23:180-95
- 435 [26] de Koning CH, van den Heuvel SP, Staal JB, Smits-Engelsman BC, Hendriks EJ (2008) Clinimetric
436 evaluation of active range of motion measures in patients with non-specific neck pain: a systematic
437 review. *Eur Spine J* 17: 905-21
- 438 [27] Whitcroft KL, Massouh L, Amirfeyz R, Bannister G (2010) Comparison of methods of measuring
439 active cervical range of motion. *Spine* 35: E976-80
- 440 [28] Tyson S (2003) A systematic review of methods to measure posture. *Phys Ther Rev* 8:45-50
- 441 [29] Hald ES, Hertle RW, Yang D (2011) Application of a digital head-posture measuring system in
442 children. *Am J Ophthalmol* 151:66-70
- 443 [30] Arbogast KB, Gholve PA, Friedman JE, Maltese MR, Tomasello MF, Dormans JP (2007) Normal
444 cervical spine range of motion in children 3–12 years old. *Spine* 32: E309-E315
- 445 [31] Lewandowski J, Szulc P (2003) The range of motion of the cervical spine in children aged from 3
446 to 7 years—an electrogoniometric study. *Folia Morphol* 62:459-461
- 447 [32] Lynch-Caris T, Majeske KD, Brelin-Fornari J, Nashi S (2008) Establishing reference values for
448 cervical spine range of motion in pre-pubescent children. *J Biomech* 41:2714-2719
- 449 [33] Feipel V, Rondelet B, Le Pallec JP, Rooze M (1999) Normal global motion of the cervical spine: an
450 electrogoniometric study. *Clin Biomech* 14:462-470
- 451 [34] Sforza C, Grassi G, Fragnito N, Turci M, Ferrario VF (2002) Three-dimensional analysis of active
452 head and cervical spine range of motion: effect of age in healthy male subjects. *Clin Biomech* 17:611-
453 614
- 454 [35] Rubio AS, Griffet JR, Caci H, Bérard E, El Hayek T, Boutté P (2009) The moulded baby syndrome:
455 incidence and risk factors regarding 1,001 neonates. *Eur Journal Ped* 168:605
- 456 [36] Petronic I, Brdar R, Cirovic D, Nikolic D, Lukac M, Janic D, Pavicevic P, Golubovic Z, Knezevic T
457 (2010) Congenital muscular torticollis in children: distribution, treatment duration and outcome. *Eur*
458 *J Phys Rehabil Med* 46:153-8

- 459 [37] Captier G, Dessauge D, Picot MC, Bigorre M, Gossard C, El Ammar J, Leboucq N (2011)
 460 Classification and pathogenic models of unintentional postural cranial deformities in infants:
 461 plagiocephalies and brachycephalies. *J Craniofac Surg* 22:33-41
- 462 [38] Lee YT, Yoon K, Kim YB, Chung PW, Hwang JH, Park YS, Han BH (2011) Clinical features and
 463 outcome of physiotherapy in early presenting congenital muscular torticollis with severe fibrosis on
 464 ultrasonography: a prospective study. *J Pediatr Surg* 46:1526-1531
- 465 [39] Lee JY, Koh SE, Lee IS, Jung H, Lee J, Kang JI, Bang H (2013) The cervical range of motion as a
 466 factor affecting outcome in patients with congenital muscular torticollis. *Ann Rehab Med* 37:183-90
- 467 [40] Jung AY, Kang EY, Lee SH, Nam DH, Cheon JH, Kim HJ (2015) Factors that affect the rehabilitation
 468 duration in patients with congenital muscular torticollis. *Ann Rehab Med* 39:18-24
- 469 [41] Oleszek JL, Chang N, Apkon SD, Wilson PE (2005) Botulinum toxin type a in the treatment of
 470 children with congenital muscular torticollis. *Am J Phys Med Rehabil* 84:813-816
- 471 [42] Lal S, Abbasi AS, Jamro S (2011) Response of primary torticollis to physiotherapy. *J Surg Pak*
 472 (Intl) 16:153-156
- 473 [43] Aarnivala H, Vuollo V, Harila V, Heikkinen T, Pirttiniemi P, Holmström L, Valkama AM (2016) The
 474 course of positional cranial deformation from 3 to 12 months of age and associated risk factors: a
 475 follow-up with 3D imaging. *Eur J Pediatr* 175:1893-1903
- 476 [44] Laughlin J, Luerssen TG, Dias MS (2011) Prevention and management of positional skull
 477 deformities in infants. *Pediatrics* 128:1236-1241
- 478 [45] Rogers GF, Oh AK, Mulliken JB (2009) The role of congenital muscular torticollis in the
 479 development of deformational plagiocephaly. *Plast Reconstr Surg* 123:643-52
- 480 [46] Persing J, James H, Swanson J, Kattwinkel J (2003) Prevention and management of positional
 481 skull deformities in infants. *Pediatrics* 112:199-202
- 482 [47] Emery C (1994) The determinants of treatment duration for congenital muscular torticollis.
 483 *Phys Ther* 74:921-928
- 484 [48] Moher D, Liberati A, Tetzlaff J, Altman DG & PRISMA Group (2009) Preferred reporting items for
 485 systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 6:e1000097
- 486 [49] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol D L, De Vet HC (2010) The
 487 COSMIN checklist for assessing the methodological quality of studies on measurement properties of
 488 health status measurement instruments: an international Delphi study. *Qual Life Res* 19:539-549
- 489 [50] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC (2012) Rating the
 490 methodological quality in systematic reviews of studies on measurement properties: a scoring
 491 system for the COSMIN checklist. *Qual Life Res* 21:651-657
- 492 [51] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, de Vet HC (2007) Quality
 493 criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*
 494 60:34-42
- 495 [52] Cheng JC, Wong MW, Tang SP, Chen TMK, Shum SL, Wong EM (2001) Clinical determinants of
 496 the outcome of manual stretching in the treatment of congenital muscular torticollis in infants: a
 497 prospective study of 821 cases. *J Bone Joint Surg Am* 83:679-687

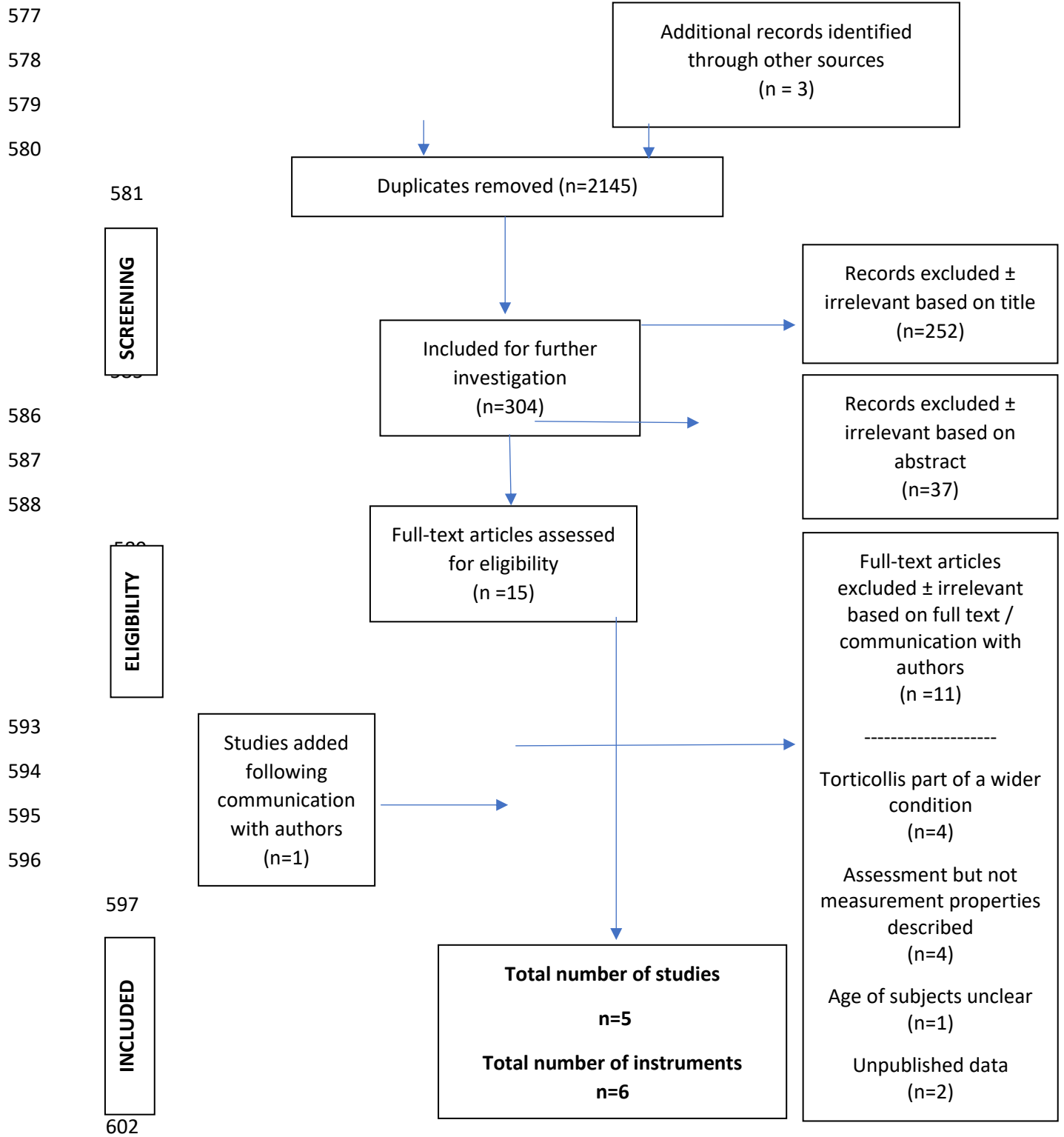
- 498 [53] Christensen E, Castle KB, Hussey E (2015) Clinical Feasibility of 2-Dimensional Video Analysis of
499 Active Cervical Motion in Congenital Muscular Torticollis. *Pediatr Phys Ther* 27:276-283
- 500 [54] Chen TMK (1995) Infantile muscular torticollis: Efficacy of conservative treatment with active
501 programme and manual stretching. Chapter 4: Pilot study on the measurement of passive rotation
502 of the neck. Master of Philosophy Dissertation, Chinese University of Hong Kong (personal
503 communication)
- 504 [55] Barske E, Draeger D, Keck K (2014) Comparison of 2-dimensional and 3-dimensional motion
505 analysis to measure infant cervical range of motion. Wisconsin Physical Therapy Association,
506 Wisconsin, USA
- 507 [56] Nuysink J, van Haastert IC, Takken T, Helders PJ (2011) Symptomatic asymmetry in very young
508 infants: a Delphi study on the development of a screening instrument. *Physiother Theory
509 Pract* 27:194-212
- 510 [57] Philippi H, Faldum A, Bergmann H, Jung T, Pabst B, Schleupen A (2004) Idiopathic infantile
511 asymmetry, proposal of a measurement scale. *Early Human Dev* 80:79-90
- 512 [58] Öhman A (2012) The inter-rater and intra-rater reliability of a modified “severity scale for
513 assessment of plagiocephaly” among physical therapists. *Physiother Theory Pract* 28:402-406
- 514 [59] Leung AY, Mandrusiak A, Watter P, Gavranich J, Johnston LM (2016) Clinical assessment of head
515 orientation profile development and its relationship with positional plagiocephaly in healthy term
516 infants—a prospective study. *Early Hum Dev* 96:31-38
- 517 [60] Cheng JC, Tang SP, Chen TMK (1999) Sternocleidomastoid pseudotumor and congenital
518 muscular torticollis in infants: a prospective study of 510 cases. *J Pediatr* 134:712-716
- 519 [61] Klackenberg E, Elfving B, Haglund-Akerlind Y, Brogren Carlberg E (2005) Intra-rater reliability in
520 measuring range of motion in infants with congenital muscular torticollis. *Adv Physio* 7:84-91. doi:
521 10.1080/14038190510010331
- 522 [62] Öhman AM, Nilsson S, Beckung E (2009) Validity and reliability of muscle function scale, aimed
523 to assess the lateral flexors of the neck in infants. *Physiother Theory Pract* 25:129-137. doi:
524 10.1080/09593980802686904
- 525 [63] Rahlin M, Sarmiento B (2010) Reliability of still photography measuring habitual head deviation
526 from midline in infants with congenital muscular torticollis. *Pediatr Phys Ther* 22:399-406. doi:
527 10.1097/PEP.0b013e3181f9d72d
- 528 [64] Murgia M, Venditto T, Paoloni M, Hodo B, Alcuri R, Bernetti A, Mangone M (2016) Assessing the
529 cervical range of motion in Infants with positional plagiocephaly. *J Craniofac Surg* 27:1060-1064. doi:
530 10.1097/SCS.0000000000002644
- 531 [65] Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL (2012). Measurement properties
532 of performance-based measures to assess physical function in hip and knee osteoarthritis: a
533 systematic review. *Osteoarthr Cartilage* 20:1548-1562
- 534 [66] Bartels B, De Groot JF, Terwee CB (2013) The six-minute walk test in chronic paediatric
535 conditions: a systematic review of measurement properties. *Phys Ther* 93:529-541
- 536 [67] Öhman AM, Beckung E (2008) Reference values for range of motion and muscle function of the
537 neck in infants. *Pediatr Phys Ther* 20:53-57

- 538 [68] Öhman AM, Mårdrbrink EL, Stensby J, Beckung E (2011) Evaluation of treatment strategies for
539 muscle function in infants with congenital muscular torticollis. *Physiother Theory Pract* 27:463-470
- 540 [69] Öhman AM, Nilsson S, Beckung E (2010) Stretching treatment for infants with congenital
541 muscular torticollis: physiotherapist or parents? A randomised pilot study. *Phys Med Rehabil* 2:1073-
542 1079
- 543 [70] Kwon DR, Park Y (2014) Efficacy of microcurrent therapy in infants with congenital muscular
544 torticollis involving the entire sternocleidomastoid muscle: a randomized placebo-controlled
545 trial. *Clin Rehabil* 28:983-991
- 546 [71] Hautopp L, Wester S, Bang B, Buus L, Grindsted J, Christensen K, Vinther A (2014) Benefit of
547 physiotherapeutic treatment in children with torticollis. *Dan Med J* 61: A4970
- 548 [72] Giray E, Karadag-Saygi E, Mansiz-Kaplan B, Tokgoz D, Bayindir O, Kayhan O (2017) A randomised,
549 single-blinded pilot study evaluating the effects of kinesiology taping and the tape application
550 techniques in addition to therapeutic exercises in the treatment of congenital muscular torticollis.
551 *Clin Rehabil* 31:1098-106
- 552 [73] He L, Yan X, Li J, Guan B, Ma L, Chen Y, Xu K (2017) Comparison of two dosages of stretching
553 treatment in infants with congenital muscular torticollis: a randomised trial. *Am J Phys Med Rehabil*
554 96:333-340.
- 555 [74] Zachman ZJ, Traina AD, Keating JJ, Bolles ST, Braun-Porter L (1989) Inter-examiner reliability and
556 concurrent validity of two instruments for the measurement of cervical ranges of motion. *J Manip
557 Physiol Ther* 12:205-210
- 558 [75] Viikari-Juntura E (1987) Inter-examiner reliability of observations in physical examinations of the
559 neck. *Phys Ther* 67:1526-1532
- 560 [76] Luxford B, Hale L, Piggott J (2009) The physiotherapy management of infants with congenital
561 muscular torticollis: a survey of current practice in New Zealand. *N Z J Physiother* 37:127-135
- 562 [77] Öhman AM, Mårdrbrink EL, Orefelt C, Seager A, Tell L, Klackenber E (2013) The physical therapy
563 assessment and management of infants with congenital muscular torticollis: a survey and suggested
564 assessment protocol for congenital muscular torticollis. *J Nov Physiother* 3:165
- 565 [78] Seager A (2017) Assessment and management of torticollis in infancy: A survey of current
566 physiotherapy practice. *Physiother Prac Res* 38:59-66
- 567 [79] Fradette J, Gagnon I, Kennedy E, Snider L, Majnemer A (2011) Clinical decision-making regarding
568 intervention needs of infants with torticollis. *Pediatr Phys Ther* 23:249-256
- 569 [80] Christensen C, Landsettle A, Antoszewski S, Ballard B, Carey H, Pax Lowes L (2013) Conservative
570 management of congenital muscular torticollis: an evidence-based algorithm and preliminary
571 treatment parameter recommendations. *Phys Occup Ther Pediatr* 33:453-466
- 572 [81] Oledzka MM, Kaplan SL, Sweeney JK, Coulter C, Evans-Rogers DL (2018) Interrater and intrarater
573 reliability of the congenital muscular torticollis severity classification system. *Pediatr Phys Ther*
574 30:176-82

IDENTIFICATION

Records identified through database
searching (n =2442)

PubMed=93
CINAHL=528
Web of Science=1684
Embase=137



603 **Fig. 1** Flowchart of study selection process and search yield

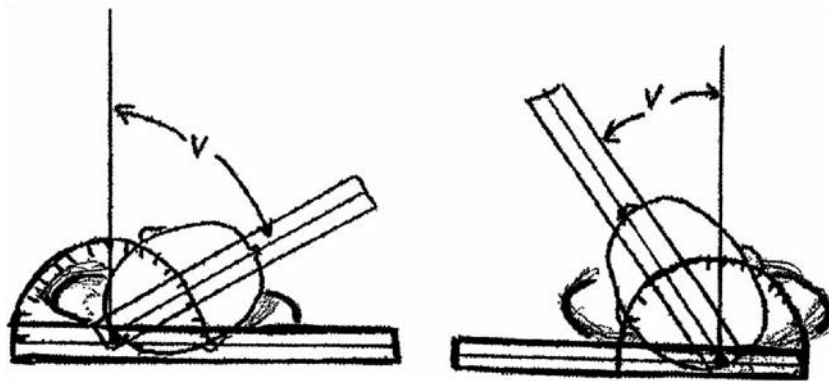
604

605

606

607

608



609

610

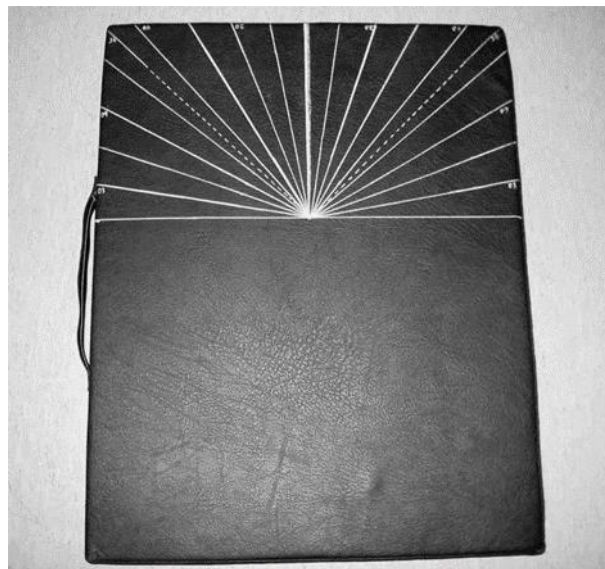
611

612

Fig. 2 Cervical rotation measured with goniometer placed above the infant's head. Reproduced from Klackenberg et al. [61] with permission from the publisher (Taylor & Francis Ltd., <http://www.tandfonline.com>)

613

614



615

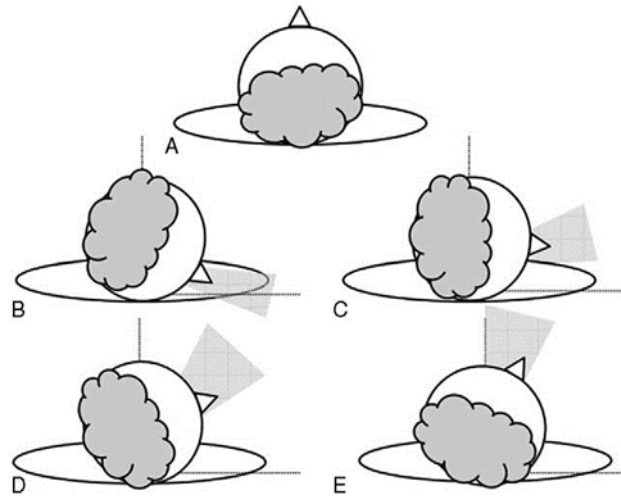
616

617

618

Fig. 3 Protractor used to measure cervical side flexion. Reproduced from Klackenberg et al. [61] with permission from the publisher (Taylor & Francis Ltd., <http://www.tandfonline.com>)

619



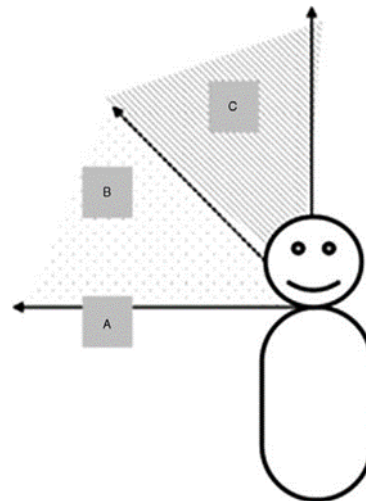
620

621

622 **Fig. 4** Range of Motion Limitation Scale - active and passive cervical rotation (A) Neutral position
 623 (B) Full ROM (C) Mild limitation (D) Moderate limitation (E) Severe limitation. Reproduced from
 624 Murgia et al. [64] with permission from the publisher (Wolters Kluwer Health Inc.,
 625 <https://wolterskluwer.com>)

626

627



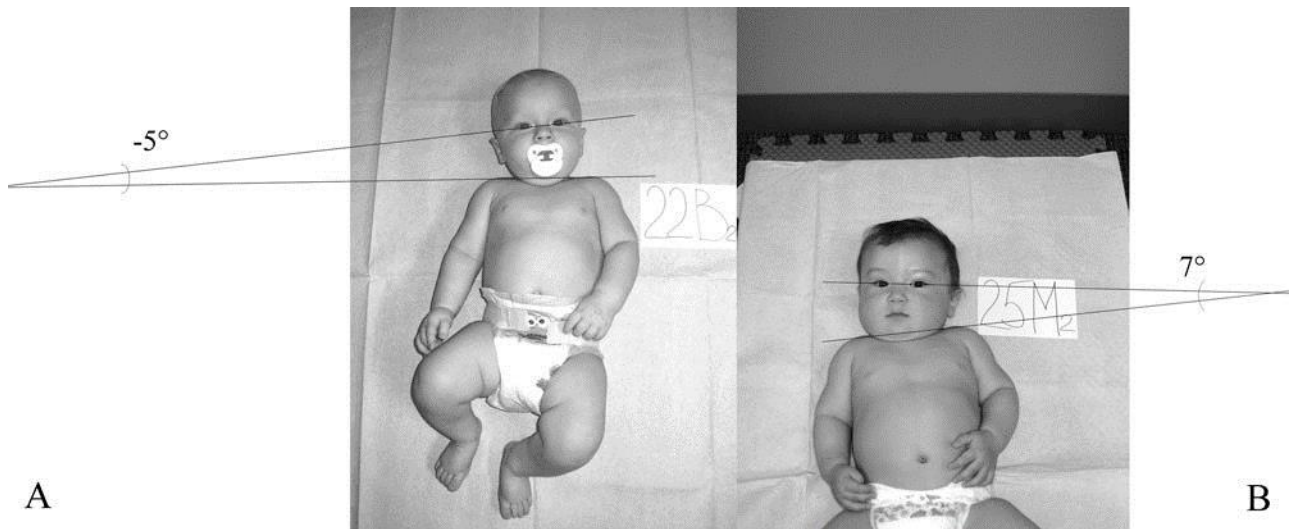
628

629 **Fig. 5** Range of Motion Limitation Scale - passive cervical side flexion (A) Full range of motion (B)
 630 Moderate limitation (C) Severe limitation. Reproduced from Murgia et al. [64] with permission from
 631 the publisher (Wolters Kluwer Health Inc., <https://wolterskluwer.com>)

632

633

634



635

636

637

638

639

640

641

642

643

644

Fig. 6 Still photography measurement procedure, with lines drawn through the eyes and acromion processes using a ruler and a pencil (A: Participant number 22, habitual head-tilt angle measured at -5° , negative value indicates head tilt to the right; B: Participant number 25, habitual head-tilt angle measured at 7° , positive value indicates head tilt to the left). Reproduced from Rahlin & Sarmiento [63] with permission from the publisher (Wolters Kluwer Health Inc., <https://wolterskluwer.com>)



645

646

5. $>75^\circ$ (Head more than 75° above the horizontal line).



647

648

4. $>45^\circ - <75^\circ$ (Head more than 45° and less than 75° above the horizontal line).



649

650

3. $>15^\circ - <45^\circ$ (Head more than 15° and less than 45° above the horizontal line).



651

652

2. $>0^\circ - <15^\circ$ (Head more than 0° and less than 15° over the horizontal line).



653

654

1. 0° (Head on the horizontal line at 0°).



655

656

657

0. $<0^\circ$ (Head below horizontal line less than 0°).

658

659 **Fig. 7** Muscle Function Scale (MFS) for Infants (Version II). Reproduced from Öhman et al. [62]

660 with permission from the publisher (Taylor & Francis Ltd., <http://www.tandfonline.com>)

661

662

Table 1: Characteristics of Included Studies

Study	Tool(s)	Measurement Property	Sample size	Diagnosis (Age)	Observers	Measure (protocol)	Time interval	Results
Chen [54]	Electronic pendular goniometer	Reliability -Inter-rater -Intra-rater	Group I: n=12 (inter-rater)	CMT (0-12 mo)	2 PTs (+assistant)	Passive CSp rot (infant supine, HOEOS, body stabilised, head aligned midline & position recorded, movement performed & EOR recorded, used landmarks of interorbital line to 7 th spinal process / posterior edge of ears, only non-affected side tested)	1 hour (inter-rater)	ICC 0.87-0.91 (Inter-rater)
			Group II: n=12 (intra-rater)				2-3 weeks (intra-rater)	ICC 0.71-0.79 (Intra-rater)
Klackenberg et al. [61]	Goniometer and adapted protractor (made from rectangular surface, 75x60x9 cm, with lines marked out in 10° intervals) <i>See Fig. 2&3</i>	Reliability -Intra-rater	Group I: n=11	CMT (1-5 mo)	1 PT (+assistant)	Passive CSp rot (infant supine, body stabilised, head/body aligned, goniometer above infant's head aligned with nose, rot performed, moveable arm of goniometer realigned with nose & angle recorded)	Group I: immediately	Group I: ICC 0.97-0.99 SEM 2-5°
			Group II: n=12				Group II: 1 hour	Group II ICC 0.77-0.95 SEM 4-6°
						Passive CSp SF (infant supine on protractor, body		

						stabilised, head/body aligned, SF performed & angle recorded – ? landmarks used)		
	Still photography (plus goniometer)	-Photographs versus the goniometer/protractor readings taken simultaneously	Groups combined n=23		1 PT (+2 assistants)	Passive CSp rot Passive CSp rot (Digital photographs taken in standardised position at EOR of above, markers on os frontale in line with nose & tip of nose, goniometer/protractor edited out of photograph, angles measured with goniometer on the photograph).		ICC 0.74-0.90 SEM 6-10° Photograph readings 2-3° lower than goniometer/protractor readings performed on infant.
Murgia et al. [64]	ROM limitation scale‡ <i>See Fig. 4&5</i>	Reliability -Inter-rater -Intra-rater	n=109	Plagiocephaly/ being assessed for PoT (0-18 mo)	2 Physicians	Active and passive CSp rot, passive CSp SF (infant supine, body stabilised, infant stimulated to rot CSp actively & position scored, gentle over-pressure given & position scored, passive SF performed & position scored)	1-24 hours (inter-rater) 48 hours (intra-rater)	Kappa 0.80 active rot; 0.83 passive rot; 0.49 passive SF (Inter-rater) Kappa 0.72 active rot; 0.73 passive rot; 0.41 passive SF (Intra-rater)

Rahlin & Sarmiento [63]	Still photography (plus protractor) <i>See Fig. 6</i>	Reliability -Inter-rater -Intra-rater	n=30 (2 photographs of each subject taken by 2 investigators = 120 photographs)	CMT (4-16 mo)	2 PTs	Habitual CSp SF (infant supine, visual stimulus at midline, photographs taken from above, lines drawn through lateral corners of eyes/ acromium processes & extended until intersect, then angle measured with protractor)	Weekly intervals x 4	ICC 0.72-0.99; Pearson r 0.72-0.99 (Inter-rater) ICC 0.79-0.84; Pearson r 0.80-0.85 (Intra-rater)
Öhman et al. [62]	Muscle Function Scale [†] Version I* Version II** <i>See Fig. 7</i>	Reliability -Inter-rater -Intra-rater Content validity	n=68 photographs	CMT (infants, age not stated)	2 groups of 7 PTs + 2 PTstudents Panel of 5 experts	Active CSp SF (observers shown the photographs of infants with their bodies supported in a horizontal position & their heads in different positions of lateral head righting. One group scored the infants using Version I and the other group using Version II of the scale). MFS (Versions I & II) discussed and panel provided insight, constructive criticism & advice.	Minimum 1 week (photographs in a different order)	Kappa 0.96; ICC 0.92-0.96 (Inter-rater Version I) Kappa 0.96; ICC 0.90-0.95 (Inter-rater Version II) Kappa 0.96-0.99; ICC 0.94-0.98 (Intra-rater Version I) Kappa 0.96-0.99; ICC 0.93-0.97 (Intra-rater Version II) Panel of experts reported validity, recommending a change in scoring system from 5 to 6 levels and

†Muscle Function Scale = an ordinal visual scale that interprets lateral head righting through a categorical scale of 0-5 (previously 0-4).

*Version I: described mainly in words; **Version II: described mainly in degrees

‡ ROM limitation scale: an ordinal visual scale that categorises CSp Rot into full (chin beyond shoulder), mild limitation (chin reaches shoulder), moderate limitation (chin goes beyond mid-clavicle) or severe limitation (chin does not go beyond mid-clavicle); and categorises CSp SF into full (ear to shoulder), moderate limitation (ear goes beyond 45°) or severe limitation (ear does not go beyond 45°).

CMT, Congenital Muscular Torticollis; CSp, cervical spine; EOR, end of range; HOEOS, head over edge of surface; ICC, intra-class coefficient; mo, months; PT, Physiotherapist; PoT, Postural Torticollis; Rot, rotation; SEM, Standard Error of Measurement; SF, side-flexion

Table 2: Quality criteria for rating the results of measurement properties (relevant sections taken from Terwee et al. 2007)

Property	Quality Criteria	Rating
Reliability	ICC or Weighted Kappa ≥ 0.70	+
	ICC or Weighted Kappa < 0.70	-
	Doubtful design or method	?
	No information found on reliability	0
Measurement Error	MIC $<$ SDC or MIC outside LOA or convincing arguments that agreement is acceptable	+
	MIC \geq SDC or MIC equals or inside LOA	-
	MIC not defined and no convincing arguments that agreement is acceptable	?
	No information found on agreement	0
Content Validity	A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection	+
	No information found on target population involvement	-
	A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method	?
	No information found on target population involvement	0

+, positive rating; -, negative rating; ?, indeterminate rating; 0, no information available

ICC, intra-class coefficient; LOA, limits of agreement; MIC, minimal important change; SDC, smallest detectable change; SEM, Standard Error of Measurement.

Table 3: Summary of results of the reliability-related studies (n=5)

Clinical Tool (Assessing)	N	Measurement Property Evaluated	Results (quality criteria rating)				COSMIN Worst Score	Worst Score Item(s)	COSMIN Worst Score (excl. sample size)
			ICC	Pearson r	SEM	Kappa			
Goniometer [61] (Passive rot in supine)	11*	Intra-rater reliability	0.98-0.99 (+)		2-3° (+)		Poor	Sample size	Poor
	12**		0.77-0.79 (+)		5-6° (+)		Independence of administration Time interval between administrations		
Electronic pendular goniometer [54] (Passive rot in supine)	12	Inter-rater reliability	0.87-0.91 (+)				Poor	Sample size	Fair
	12	Intra-rater reliability	0.71-0.79 (+)						
Protractor [61] (Passive SF in supine)	11*	Intra-rater reliability	0.97-0.98 (+)		5° (+)		Poor	Sample size	Poor
	12**		0.94-0.95 (+)		4-6° (+)		Independence of administration Time interval between administrations		
Muscle Function Scale† [62] (Active SF in horizontal)	68	Inter-rater reliability					Good	Sample size	Exc
		Version I	0.92-0.96 (+)		0.96 (+)				
		Version II	0.90-0.95 (+)		0.96 (+)				

		Intra-rater reliability						
		Version I	0.94-0.98 (+)			0.96-0.99 (+)		
		Version II	0.93-0.97 (+)			0.96-0.99 (+)		
Still photography + protractor [63] <i>(Postural SF in supine)</i>	40	Inter-rater reliability	0.72-0.99 (+)	0.72-0.99 (+)		Fair	Sample size	Exc
		Intra-rater reliability	0.79-0.84 (+)	0.80-0.85 (+)				
Still photography + protractor [61] <i>(Passive SF in supine)</i>	23	Intra-rater reliability	0.79-0.88 (+)		8-10° (-)	Poor	Sample size	Fair
Still photography + goniometer [61] <i>(Passive rot in supine)</i>			0.74-0.90 (+)		6-8° (-)			
ROM Scale‡ [64] <i>(Active rot, Passive rot, Passive SF in supine)</i>	109	Inter-rater reliability				0.80 A rot. (+) 0.83 P rot. (+) 0.49 P SF. (-)	Fair	Handling of missing items
		Intra-rater reliability				0.72 A rot. (+) 0.73 P rot. (+) 0.41 P SF. (-)		Flaw in design Kappa not weighted

* Group I: remeasured immediately ** Group II: remeasured after one hour +: positive rating -: negative rating

†Muscle Function Scale = an ordinal visual scale that interprets lateral head righting through a categorical scale of 0-5

‡ ROM limitation scale: categorised rot into full, mild limitation, moderate limitation or severe limitation; categorised SF into full, moderate limitation or severe limitation.

A, active; Exc, Excellent; excl, excluding; P, passive; rot, CSp rotation; SF, CSp side-flexion

Table 4: Summary of results for validity related and hypothesis testing study (n=1)

Clinical tool (Assessing)	N	Measurement property evaluated	Results	Study Quality Rating	COSMIN worst score count	COSMIN worst Score item
Muscle Function Scale† [62] (Active SF in horizontal)	68	Content validity	A panel of experts, consisting of five paediatric physiotherapists experienced in the assessment of muscle function in children (mean 23 years, range 12-33 years professional experience), discussed the scale and provided insight, constructive criticism and advice. This included changing the scale from five to six ordered scores and testing both the original version and a revised version (see Table 2 for details)	+	Poor	Flaw in design (lack of detail given)
	Hypothesis testing	Poor			Lack of formation of <i>a priori</i> hypotheses Constructs and measurement properties of comparator instruments not described Flaw in design (lack of detail given)	

†Muscle Function Scale = an ordinal visual scale that interprets lateral head righting through a categorical scale 0-5

Appendix 1: Search strategy for systematic review using three groups of keywords, “torticollis”, “cervical spine assessment” and “measurement properties”.

TORTICOLLIS		CERVICAL Sp ASSESSMENT		MEASUREMENT PROPERTIES
<p>Neck (MeSH)</p> <p>OR</p> <p>Neck Muscles (MeSH)</p> <p>OR</p> <p>Cervical Vertebrae (MeSH)</p> <p>OR</p> <p>Neck</p> <p>OR</p> <p>Cervical</p> <p>OR</p> <p>Cervical spine</p> <p>OR</p> <p>Torticollis (MeSH)</p> <p>OR</p> <p>Torticollis</p> <p>OR</p> <p>Congenital Muscular Torticollis</p> <p>OR</p> <p>Fibromatosis Colli</p> <p>OR</p> <p>Wryneck</p> <p>OR</p> <p>Plagiocephaly, Nonsynostotic (MeSH)</p> <p>OR</p> <p>Positional plagiocephaly</p> <p>OR</p> <p>Deformational plagiocephaly</p> <p>OR</p> <p>Non-synostotic plagiocephaly</p>	<p>AND</p>	<p>Range of motion, articular (MeSH)</p> <p>OR</p> <p>Joint Flexibility</p> <p>OR</p> <p>Joint Range of Motion</p> <p>OR</p> <p>Passive Range of Motion</p> <p>OR</p> <p>Active Range of Motion</p> <p>OR</p> <p>Movement</p> <p>OR</p> <p>Muscle Strength (MeSH)</p> <p>OR</p> <p>Muscle Strength</p> <p>OR</p> <p>Posture (MeSH)</p> <p>OR</p> <p>Posture</p> <p>OR</p> <p>Symptom assessment (MeSH)</p>	<p>AND</p>	<p>Reproducibility of Results (MeSH)</p> <p>OR</p> <p>Validation Studies (MeSH)</p> <p>OR</p> <p>Outcome Assessment, Health Care (MeSH)</p> <p>OR</p> <p>Valid*</p> <p>OR</p> <p>Reliab*</p> <p>OR</p> <p>Repeatab*</p> <p>OR</p> <p>Reproducibility</p> <p>OR</p> <p>Psychometric*</p> <p>OR</p> <p>Clinimetric*</p>

