A systematic review on the validity and reliability of tape measurement method in leg length discrepancy

Behshid Farahmand¹, Esmail Ebrahimi Takamjani², Hamid Reza Yazdi³, Hassan Saeedi⁴, Mohammad Kamali⁵, Masumeh Bagherzadeh Cham⁶

Received: 16 May 2018  Published: 26 May 2019

Abstract

Background: Leg length discrepancy measurement is an essential part of musculoskeletal clinical assessment, and tape measurement is a common clinical method. This study aimed to systematically review the results of the findings of studies on validity and reliability of the tape measurement method and the quality of reporting the literature on this topic.

Methods: A search was performed in PubMed, EBSCO, Science Direct, Web of Knowledge, Scopus, Embase, and Google Scholar using selected keywords from inception to December 2017. This systematic review was based on the PRISMA guideline. After a systematic selection process, the quality of the included studies was assessed independently by 2 reviewers using the Brink and Louw Scale for quality assessment.

Results: A total of 11 studies were finally considered for this systematic review. Two studies were about the validity of (a measurement tool) studies and 4 were reliability analysis only. Validity and reliability analyses were simultaneously applied to 5 studies. Also, 9 out of 11 studies were deemed to be of high quality based on Brink and Louw Scale. Studies showed high (ICC=0.9) levels of interrater and intrarater reliability. The validity of the methods ranged from low to very high depending on subjects.

Conclusion: Tape measurement method has acceptable reliability and validity in healthy people, but it does not have acceptable validity in measuring obese people and patients with musculoskeletal disorders. Thus, using a suitable method for LLD leg length discrepancy measurement seems to be necessary for obese and individuals with leg length discrepancy.

Keywords: Leg length discrepancy, Validity, Reliability

Introduction

Leg length discrepancy (LLD) is a common musculoskeletal disorder. Almost 70% of the general population suffers from LLD of up to 1 cm (1). LLD assessment is an important topic in evaluation of lower limb. Various leg length assessment procedures are commonly being used by several practitioners, such as: physical therapists, orthopedics, technical orthopedics, chiropractic, and podiatrists (2, 3). LLD assessment is a challenging task among researchers and clinicians. There is still controversy about the assessment of LLD measurements. Two general categories of methods have been used for LLD assessment: (1) imaging techniques, (2) clinical methods (4). Imaging techniques are the standard procedure for accurate LLD measurement, but they are costly, time-intensive, and expose patients to radiation (4). Therefore, clinical methods are more popular because of their availability, easy procedure, and usefulness in diagnosis.

Corresponding author: Dr Masumeh Bagherzadeh Cham, bagherzadehcham.m@iums.ac.ir

¹ Department of Orthotics & Prosthetics, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
² Department of Physiotherapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
³ Orthopedic Department, Firuzgar Hospital, Iran University of Medical Sciences, Tehran, Iran
⁴ Department of Orthotics & Prosthetics, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
⁵ Rehabilitation Management Department, Iran University of Medical Sciences, Tehran, Iran
⁶ Neuromusculoskeletal Research Center, Department of Physical Medicine and Rehabilitation, Iran University of Medical Sciences, Tehran, Iran

What is “already known” in this topic:
Tape measurement is a simple, affordable, and non-invasive method to evaluate limb length inequality in healthy people with leg length discrepancy more than 5 mm. The validity and reliability of this method decreased in obese people and in those with orthopedic disorder.

What this article adds:
New valid, reliable, non-invasive, and cost-effective method is needed to evaluate leg length discrepancy in obese people and in those with orthopedic or neuromuscular disorder.
Validity and reliability of tape measurement method

and low cost. Indirect and direct methods are used for clinical LLD assessment (4-6). In indirect methods, lift blocks of different thicknesses are put under the short leg and then the leveling of the pelvis is checked. In direct methods, tape measurement methods (TMM), a tape is used to measure the distance between the anterior superior iliac spine (ASIS) or anterior inferior iliac spine (AIIS) to the lateral or medial malleolus in supine position (4-8). Despite easy procedure, validity of this method is not known.

There is some controversy about validity and reliability of these methods. Several literatures on TMM have shown conflicting results. In their study, Friberg et al (9) concluded that ASIS to medial malleolus measurement is an inaccurate and imprecise measure of LLD, with a mean LLD error of 8.6 mm compared to radiographs, and a 1.1 mm intraterster mean error. Authors of this study used a single measurement of ASIS to the medial malleolus. In agreement with this finding, Beattie et al (10) found validity estimates of LLD to have intra-class correlation coefficient (ICC) of 0.683, when utilizing a single measure using ASIS to medial malleolus measure. In addition, Gogia and Braatz (3) reported ICC of 0.99, with radiographs and intertester reliability of 0.98. Also, Hoyle et al (11) noted an intertester reliability ranging from 0.96 to 0.98 and an intraterster reliability ranging from 0.89 to 0.95 for ASIS to medial malleolus measurement. The widely different results have been reported on the validity and reliability of the method. Thus, the aim of this study was to perform a systematic review on the validity and reliability of TMM as a common clinical method for LLD measurement, to evaluate the quality of these studies, and to investigate the factors affecting the reliability and validity of this method.

Methods

Search strategy

To systematically search the literature published from the date of inception to December 2017, PubMed, EBSCO, Science Direct, Web of Knowledge, Scopus, and Embase databases were explored. The keywords were as follow: (“leg length discrepancy” OR “limb length discrepancy” OR “leg length inequality” OR “leg length” OR “limb length”) AND (“Validity” OR “reliability” OR “inter-tester” OR “intra-tester”) AND (“tape measure” OR “tape measurement” OR “clinical methods” OR “clinical assessment”). A word from each area was required to be in the text or the title-abstract-keyword of the study. An additional search of Google Scholar search engine was also performed. These searches were supplemented by hand searching the reference lists of the final articles found from the search.

Inclusion criteria of the studies:

a) English language
b) Full-text availability
c) Measuring the validity and/or reliability of the tape measurement method should have been the primary aim of the study
d) Pearson’s r, Cronbach α, and intraclass correlation coefficient should have been used for statistical analysis.

Studies were excluded if they examined the validity and/or reliability of tape measurement in total hip or knee disarticulation of patients and if they evaluated the accuracy and/ or precision of tape measurement.

Study selection

In this systematic review, all search procedures; selection, quality assessment, data extraction, and reading the articles were screened independently by 2 reviewers (BF & MB). In case of any difference of opinion between the reviewers, a third reviewer was asked to evaluate the article (MK). Initially, based on the inclusion criteria, studies were selected by reading the titles and abstracts. Then, full–text articles were explored to choose those that met the inclusion criteria.

Quality assessment

Ultimately, 11 articles were selected for final analysis. Two reviewers conducted an appraisal of the reporting quality of the 11 studies based on Kappa scores (12). Accordingly, a Kappa score of 0.92 or higher was regarded as acceptable.

The Brink and Louw Scale checklist was used for quality assessment (13). This 13-item scale has been developed by combining QUADAS (the Quality Assessment of Diagnostic Accuracy Studies) and QAREL (the Quality Appraisal of Diagnostic Reliability Studies) scales, as the selected studies could assess both reliability and validity of the tape measurement. Thus, this checklist can be more conveniently used as compared to either of QUADAS or QAREL independently (13, 14). The studies were considered as of high quality if their scores were higher than 60% (14, 15). Two reviewers independently assessed the quality of each study.

Data analysis

The intra-class correlation coefficient and Pearson's correlation coefficient were interpreted as follow: 0.00-0.29 as very low correlation, 0.30-0.49 as low correlation, 0.50-0.69 as moderate correlation, 0.70-0.89 as high correlation, and 0.90-1.00 as very high correlation (16).

Results

Study Selection

With a preliminary literature search, 496 abstracts were identified without excluding non-English articles. However, after applying the inclusion criteria, 11 articles were included in the review. Figure 1 presents the flow diagram of the selection process based on the PRISMA guidelines. A total of 11 studies were finally considered for this systematic review (Table 1). Two studies were subjected to validity studies only (10, 17), while 4 were subjected to reliability analysis only (11, 18-20). Validity and reliability analyses were applied to 5 studies simultaneously (3, 7, 8, 21, 22).

Characteristics

Total numbers of participants in all samples was 458. The healthy sample included 174 individuals and LLD sample included 284 (Table 2). Five studies had solely evaluated healthy individuals (3, 7, 11, 18, 19), 3 focused

http://mijri.iums.ac.ir
Records identified through database search (n = 493)
- Pubmed (n=29)
- Ebsco (n= 316)
- Science direct (n= 7)
- Web of Knowledge (n= 40)
- Scopus (n= 66)
- Embase (n= 29)
- Google Scholar (n= 6)

Additional records identified through other sources references searching (n = 3)

Records after duplicates removed (n = 412)

Records after title and abstract screened (n = 35)

Full-text articles assessed for eligibility (n = 11)

Studies included in qualitative synthesis (n = 11)

Validity (n = 2)
Reliability (n = 4)
Validity & Reliability (n = 5)

Full-text articles excluded, (n = 24) with reasons:
- Not full text (n = 1)
- Not validity or reliability (n = 7)
- New clinical methods (n = 3)
- No tape measurement (n = 12)
- Not English language (n = 1)

Fig. 1. PRISMA flow diagram

Among the studies that evaluated the validity of TMM, only Neelly et al (7) and Gogia et al (3) studies had used 2 examiners and in other studies only one examiner’s evaluation was compared to the reference method (8, 10, 17, 21, 22). Moreover, only Beattie et al (10) study compared the validity of TMM in healthy (ICC: 0.359- 0.786) and patient (ICC: 0.770- 0.852) groups (Table 1).

In studies that measured reliability, all studies applied expert examiners and intertester and intratester reliability were 0.477- 0.991 and 0.679- 0.990, respectively (3, 7, 8, 11, 18-22). Only in Duff et al (21) study expert and non-expert examiners evaluated LLD and found between expert examiners intertester reliability of 0.49, between expert and non-expert examiners reliability of 0.01- 0.69. Hoyle et al (11) study measured intertester and intratester reliability of TMM in right and left foot separately. Also, Terry et al (20) measured intertester and intratester reliability of TMM in 2 positions: ASIS to medial maleolus (ICC: 0.80) and to lateral maleolus (ICC: 0.83).

There were 4 radiography studies (3, 17, 21, 22), 1 study with CT Scan (7), and 3 with scanogram (8, 19, 20). In Beattie et al (10) study, both radiography and mini scanogram were used as reference method (Table 2).

In most studies, leg length was measured from ASIS to medial maleolus (3, 7, 8, 10, 11, 18, 19, 21). In Badii et al (22) study, leg length was measured from ASIS to lateral maleolus and in Terry et al (20) study, it was measured from ASIS to both medial and lateral maleolus. However, only in lamp et al (17) study, leg length was measured from ASIS to knee joint and from knee joint to medial maleolus (Table 2).
Validity and reliability of tape measurement method

<table>
<thead>
<tr>
<th>Table 1: Summary of studies and validity and reliability of the included studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Badii et al, 2014</td>
</tr>
<tr>
<td>Neelly et al, 2013</td>
</tr>
<tr>
<td>Jamalaadinn et al, 2011</td>
</tr>
<tr>
<td>Leard et al, 2009</td>
</tr>
<tr>
<td>Bretas et al, 2009</td>
</tr>
<tr>
<td>Terry et al, 2005</td>
</tr>
<tr>
<td>Duff et al, 2000</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Lampe et al, 1996</td>
</tr>
<tr>
<td>Hoyle et al, 1991</td>
</tr>
<tr>
<td>Beattie et al, 1990</td>
</tr>
<tr>
<td>Gogia et al,1986</td>
</tr>
</tbody>
</table>

T: test session, E: examiner, EX: expert examiner, NEX: non-expert examiner

<table>
<thead>
<tr>
<th>Table 2: Summary of the included studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Badii et al, 2014</td>
</tr>
<tr>
<td>Neelly et al, 2013</td>
</tr>
<tr>
<td>Jamalaadinn et al, 2011</td>
</tr>
<tr>
<td>Leard et al, 2009</td>
</tr>
<tr>
<td>Bretas et al, 2009</td>
</tr>
<tr>
<td>Terry et al, 2005</td>
</tr>
<tr>
<td>Duff et al, 2000</td>
</tr>
</tbody>
</table>

Quality of studies

The 2 reviewers had no disagreement about article selection (kappa score 1). However, they assessed the quality of the articles differently. Nine out of the 11 studies...
were deemed to be of high quality (score > 60%). One of the studies subjected to validity studies was found to be of high quality (10). Three out of 4 reliability studies were found to be of high quality (18-20). Also, all studies in which combined reliability and validity studies were performed were found to be of high quality (3, 7, 8, 21, 22). Details of the scoring process are presented in Table 3. The main areas of weakness were insufficient interrater and intrarater blinding, lack of variation in testing order, the time period between reference standard and index test, and withdrawal from the study.

Table 3. Results of quality assessment of studies using Brink and Louw Scale

<table>
<thead>
<tr>
<th>Brink and Louw Scale</th>
<th>Badli et al., 2014</th>
<th>Needly et al., 2013</th>
<th>Jamaldin et al., 2011</th>
<th>Leard et al., 2009</th>
<th>Beattie et al., 2009</th>
<th>Terry et al., 2005</th>
<th>Daff et al., 2000</th>
<th>Lampe et al., 1994</th>
<th>Hoyle et al., 1991</th>
<th>Beattie et al., 1990</th>
<th>Gogia et al., 1986</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the sample of subjects’ representative?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Was the sample of raters’ representative?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Was the reference standard explained?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Not</td>
<td>Not</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Were raters blinded to the findings of other raters?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Not</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Were raters blinded to their own prior findings?</td>
<td>+</td>
<td>+</td>
<td>Not</td>
<td>+</td>
<td>Un</td>
<td>Un</td>
<td>+</td>
<td>Not</td>
<td>Un</td>
<td>Not</td>
<td>Not</td>
</tr>
<tr>
<td>Was the order of examination varied?</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>+</td>
<td>+</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
</tr>
<tr>
<td>Is the time period between reference standard and index test short enough?</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>Not</td>
<td>Un</td>
<td>+</td>
<td>Un</td>
<td>Un</td>
<td>Not</td>
<td>Un</td>
<td>+</td>
</tr>
<tr>
<td>Was the time interval between repeated measures appropriate?</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Un</td>
<td>Un</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Was the reference standard independent to the index test?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Not</td>
<td>Not</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Not</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Was the execution of the index test described in sufficient detail?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Was the execution of the reference standard described in sufficient detail?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Not</td>
<td>Not</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>Not</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Were withdrawals from the study explained?</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>Un</td>
<td>+</td>
<td>Un</td>
<td>+</td>
</tr>
<tr>
<td>Were the statistical methods appropriate?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Not = Not applicable, Un = Unclear, (+) = Yes, (-) = No.
Validity and reliability of tape measurement method

Discussion

The aim of this study was to review the validity and reliability of TMM for LLD measurement. Results of this study showed that TMM is a valid and reliable tool for LLD measurement in healthy people with no excess weight or musculoskeletal disorder. Moreover, an expert examiner measured LLD from ASIS to medial or lateral malleolus twice or more, and the mean measurements were considered. Even though radiography is the standard procedure for LLD measurement, it is too expensive, exposes the subject to radiation, and requires special equipment. Therefore, based on the results of this study, it seems that TMM serves as a safe method for LLD in clinics and research centers.

TMM refers to direct measurement of LLD using bony landmarks (4, 5). In this method, the person is in supine position with his/her lower limb being in anatomical position (extended hip and knee with the ankles in neutral position) (3, 10). Different studies have checked different distances (eg, ASIS to medial malleolus, ASIS to lateral malleolus, ASIS to medial knee joint, medial knee joint to medial malleolus, umbilicus to medial malleolus, or sternum to malleolus) (17, 20, 22, 23).

The results of the present study showed that the distance between ASIS to the medial malleolus provides the highest validity, while the distance between ASIS to the lateral malleolus merely provides acceptable validity (17, 20). Other measurement methods provided significantly lower validities. Results of a study by Terry et al (20) demonstrated that the values of ICC for intratester reliability of the ASIS to the medial malleolus and lateral malleolus was 0.78 and 0.88, respectively, and those for interrater reliability of the same distances was 0.80 and 0.83, respectively.

In these studies, validity and reliability of TMM were investigated in healthy individuals and patients. The patients had a history of lower limb fractures, Blount disease, Tibia or Femur Hemimelia, hemihypertrophy, osteomyelitis, osteoarthritis, low back pain, tibia dysplasia, obesity, or other deformities. The results also revealed that validity and reliability of the TMM for LLD measurement were acceptable in healthy participants (0.97-0.98) (7) compared to the patients (ICC=0.33-0.39) (19). However, Beattie et al (10) found relatively different results (Table 1). Significantly lower validity and reliability were observed in obese participants (ICC=0.22) (19).

In the meantime, TMM uses bony landmarks for LLD measurement, making it difficult to detect accurate location of the ASISin obesity. Asymmetry in other segments (thighs, knees, ankles) due to swelling, muscle atrophy, contracture, or obesity can change the direction of the tape and alter the results. Pelvic obliquity may not actually cause LLD but can affect measurement (10, 19).

Incorrect detection of bony landmarks and inappropriate placement of the tape along the lower limb can introduce errors into the measurement, and therefore the examiner's skill and experience in LLD measurement via TMM affects reliability and validity of the results (21, 23). In most studies, LLD measurement via TMM is practiced by expert examiners, such as physiotherapists, occupational therapist, rheumatologist, orthopedic surgeons, or podiatrists (3, 7, 8, 11, 22). One study evaluated the effect of tester's skill on intertester and intratester reliability of TMM and found a direct relationship between examiner's skill and reliability of the results (21).

TMM was compared to a reference method to evaluate its validity. For this purpose, the examiner measured LLD via TMM for 1 or more than 1 time. The results showed that to enhance the validity of TMM, the tester must measure LLD for at least 2 times and average the results before comparing it to the reference method (10). Beattie et al (10) demonstrated the validity of TMM at the first and second measurements to be 0.359 and 0.786, respectively. They also found the highest validity with dual measurements to be 0.852. The results also indicated that, in general, the second measurement was more valid than the first, perhaps due to increased examiner's skill. However, Neely et al (7) and Jamaluddin et al (8) found that one-time measurement still provides acceptable validity.

Validity of TMM for LLD measurement has also been shown to depend on the length of the 2 limbs. Accordingly, higher errors have been reported for difference of less than 5 mm, while significantly lower errors are reported for cases where the difference exceeds 5 mm (9, 10). Beattie et al (10) found that any difference of smaller than 5 mm in leg length can increase the rate of error in determining the shorter length by the examiner on 4 out of 9 individuals, and the examiner will not make any mistake if the difference exceeded 5 mm.

Different studies have compared TMM with standard procedures, such as radiography, CT scan, and scanography (4). In LLD measurement, using CT scan images, the same bone landmarks as those used in TMM are used and measurements are made in supine position (24). In the CT scan method, biomechanical or structural asymmetry of the leg and ankle is not shown due to weight bearing. In studies where TMM was compared to CT scan, the validity of LLD measurements was higher than that provided by standing radiographs (7, 8, 20).

The results of a reliability study showed that when both examiners were adequately skillful, intertester and intratester reliability were acceptably high. Leard et al (18) reported an intertester reliability of 0.95-0.96 and intratester reliability of 0.99. Also, Gogia et al showed intratester reliability of 0.99. Duff et al (21) showed that intertester error exceeds that of intratester reliability. Jamaluddin et al (8) and Brétas et al (19) found an intertester reliability of 0.807 and an intratester reliability of 0.668.

The current literature review has several limitations. First, as with any systematic review, it is possible that some related articles could not be identified, such as unpublished work or conferences articles. Second, only articles in English were included. Third, studies that evaluated accuracy and precision were not included. Finally, type of participants, examiners, and landmarks that were used for the measurement and the reference methods were not uniform.

Conclusion

In general, according to studies in which validity and re-
The authors study.

Accompanying real applications for those with pelvic tilt or other problems, it is necessary to use a method which is suitable for both the healthy, the obese, and patients. Also, the method should be done in a standing position, should eliminate the need for x-rays, and should not require examiner skill and expertise.

Acknowledgments

The authors would like to thank Iran University of Medical sciences, Tehran, Iran, for financially supporting this study.

Conflict of Interests

The authors declare that they have no competing interests.

References