

## RESEARCH ARTICLE

# Leg length discrepancy: A systematic review on the validity and reliability of clinical assessments and imaging diagnostics used in clinical practice

Martin Alfuth<sup>1,2\*</sup>, Patrick Fichter<sup>2</sup>, Axel Knicker<sup>3</sup>

**1** Faculty of Health Care, Therapeutic Sciences, Niederrhein University of Applied Sciences, Krefeld, Germany, **2** Department of Further Education, M.Sc. Sport Physiotherapy, German Sport University Cologne, Cologne, Germany, **3** Institute of Movement and Neurosciences, German Sport University Cologne, Cologne, Germany

\* [martin.alfuth@hs-niederrhein.de](mailto:martin.alfuth@hs-niederrhein.de)

## Abstract

### Background

A variety of assessments to determine leg length discrepancy (LLD) is used in clinical practice and evidence about validity and reliability may differ.

### Objective

The objective of this systematic review was to identify and describe the validity and reliability of different assessments and imaging diagnostics for the determination of LLD.

### Materials and methods

The review was conducted following the recommendations of Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). The databases Medline (PubMed) and Index to Chiropractic Literature were systematically searched. Studies regarding clinical assessments and imaging diagnostics for the diagnosis of LLD, which reported the clinical properties for assessment of LLD, were included and screened for methodological quality using the Quality Assessment of Studies of Diagnostic Accuracy (QUADAS-2) tool for validity studies and the Quality Appraisal of Diagnostic Reliability (QAREL) tool for reliability studies.

### Results

Thirty-seven articles on clinical assessments and 15 studies on imaging diagnostics met the eligibility criteria. Thirteen studies on the validity of clinical assessments and six studies on the validity of imaging diagnostics had a low risk of bias and low concerns regarding applicability for all domains. One study on the reliability of clinical assessments and one study on the reliability of imaging diagnostics had a low risk of bias. Main limitations were, that an

## OPEN ACCESS

**Citation:** Alfuth M, Fichter P, Knicker A (2021) Leg length discrepancy: A systematic review on the validity and reliability of clinical assessments and imaging diagnostics used in clinical practice. PLoS ONE 16(12): e0261457. <https://doi.org/10.1371/journal.pone.0261457>

**Editor:** Osama Farouk, Assiut University Faculty of Medicine, EGYPT

**Received:** August 27, 2021

**Accepted:** December 2, 2021

**Published:** December 20, 2021

**Copyright:** © 2021 Alfuth et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting Information](#) files.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors declare that no competing interests exist.

analysis of sensitivity and specificity was only performed in a few studies and that a valid reference standard was lacking in numerous studies on clinical assessments.

## Conclusions

For the clinical assessment of LLD, the block test appears to be the most useful method. Full-length standing anteroposterior radiography seems to be the most valid and reliable method and may be used as global reference standard to measure the anatomic LLD when comparing clinical methods and imaging diagnostics.

## Introduction

Clinicians and physiotherapists are frequently confronted with leg length discrepancy (LLD) in their patients [1, 2]. LLD is categorized as an anatomic or functional type [3] and may result in multiple problems for the affected patient. It was declared that when joint pain in the lower extremities arises, any leg length discrepancy should be corrected to prevent increased stresses of weight bearing at other joints that may result in joint dysfunction from relatively harmless movements [4]. Some patients show both types of LLD, which can lead to a compensation or an increase of LLD. The anatomic LLD arises, when the cumulative bone length and the thickness of cartilage significantly differ between both legs. Gurney [5] states that the causes for the anatomic LLD are congenital or acquired. The most prevalent congenital ones are dislocation of the hip and hemiatrophy or hemihypertrophy with skeletal involvement of the lower extremity. Acquired causes may develop from infections, palsy, tumors, surgery, e.g. arthroplasty of the hip or knee, or slipped capital femoral epiphysis [5, 6]. Functional LLD may be caused by contracture of soft tissue, contractures of the hip or knee joints, pelvic obliquity or foot deformities [3, 7]. During loaded standing functional and anatomic LLD may interact.

The prevalence of anatomical LLD in the population is 90% [1]. Forty-one percent (41.3%) of the population demonstrate an anatomic LLD of 0–4 mm, 37.4% of 5–9 mm, 20% of more than 9 mm, 15% of 10–14 mm and 6.4% of more than 14 mm. LLD of > 5 mm is related to an increased risk of osteoarthritis of the hip and knee joints [8] as well as low back pain and lumbar scoliosis [9]. Moreover, a LLD > 6 mm is associated with an increased intensity of low back pain [10] and > 10 mm with an enhanced rate of hip and knee arthroplasties [9].

A valid, reliable and accurate method to measure leg length discrepancy is important to treat the patients effectively. A variety of assessments and imaging diagnostics to determine LLD is used, however, evidence about validity, reliability and diagnostic accuracy may differ [5, 11]. Moreover, it was concluded that the measurement of leg length discrepancy using a tape measure and standing on blocks is not as accurate as the measurement using imaging methods [12]. Therefore, the aim of this systematic review was to identify, describe, and compare the validity and reliability of different clinical assessments and imaging diagnostics for the measurement of LLD. It was hypothesized that imaging diagnostics are more valid, reliable, and accurate than clinical measurement methods.

## Materials and methods

### Study design and search strategy

A systematic review was conducted following the recommendations of Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [13]. Within the scope of a

master's thesis, the databases Medline (PubMed) and Index to Chiropractic Literature were systematically searched until November 2017. Therefore, the present review was not prospectively registered. In order to update the search for the present publication, the search strategy was upgraded and repeated until October 15, 2020. In PubMed the following search terms were combined using the Boolean Operator "OR": "leg length discrepancy, leg length inequality, limb length discrepancy, limb length inequality, long limb and long foot" and limited to humans. Then, the search terms "physical therapy, physiotherapy, orthopaedic, orthopaedics, orthopedic, osteopathy, chiropractic, alternative medicine, podiatry and podiatric" were combined using the Boolean Operator "OR" and limited to humans. Additionally, the search terms "measurement, measure, assessment, test, physical exam, physical examination, clinical examination, imaging diagnostics and diagnostic imaging" were combined and limited using the same procedure. For the updated search we added the following search terms and combined them with the operator "OR" and limited to humans: reliability, reproducibility, validity, accuracy, comparability, comparison, variability, variation, variance. Finally, all combined search combinations using the operator "OR" were then combined using the Boolean Operator „AND“. In Index to Chiropractic Literature the search term "leg length" was combined with the search terms "leg-length", "limb length" and "limb-length using the operator "OR". Furthermore, the reference lists of all eligible diagnostic studies and reviews, as well as the reference lists of reviews with similar topics were screened. A flow diagram of searches for studies is presented in [Fig 1](#).

### Eligibility of studies

Full text studies regarding clinical assessments and imaging diagnostics for the diagnosis of LLD, which reported the clinimetric properties, involving validity or accuracy, and reliability for the assessment of LLD or leg lengths, were included. Further inclusion criteria were English or German language, for studies dealing with clinical assessments a year of publication  $\geq 1983$  and for studies regarding imaging diagnostics a year of publication  $\geq 2000$  to provide a review of updated literature.

### Exclusion criteria

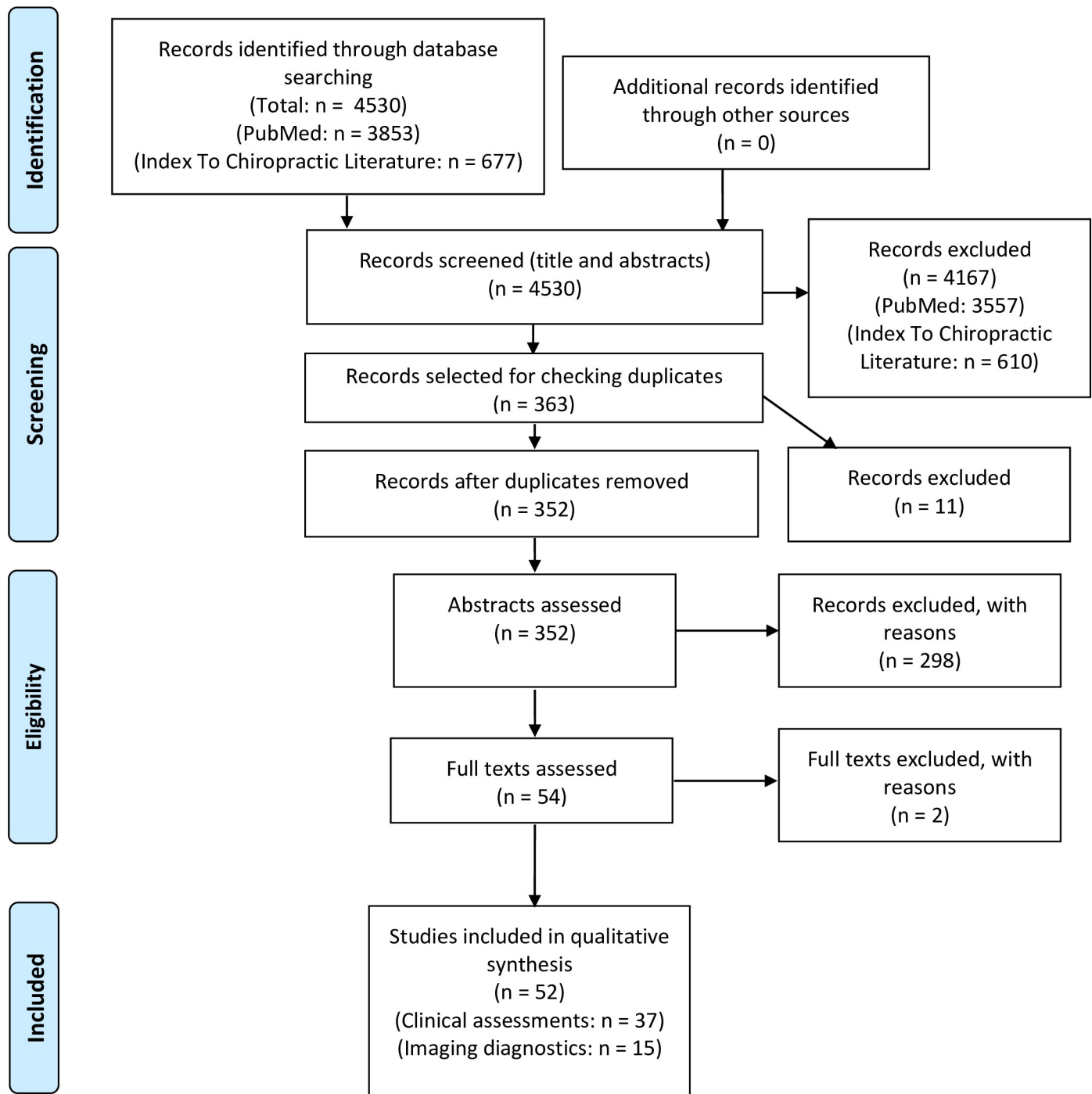
All studies that did not aim at examining LLD or leg lengths in humans and did not meet the aforementioned inclusion criteria were excluded.

### Selection of studies and analysis

Titles and abstracts of the identified studies were screened independently by two reviewers (P.F. and M.A.). Full texts were reviewed by the two reviewers independently. Any disagreements were discussed between reviewers and clarified. A third author (A.K.) was consulted if agreement of full-text inclusion could not be reached. In that case, the third author served as arbiter for final decision about inclusion or exclusion of the respective study. Additional hand search of literature was conducted continuously by the authors for any further manuscripts dealing with LLD.

### Quality assessment of included studies

The revised "Quality Assessment of Studies of Diagnostic Accuracy (QUADAS-2)" tool was completed to evaluate the risk of bias of the included diagnostic studies and studies investigating concurrent validity [14] by two authors (P.F.; M.A.) independently ([Tables 1 and 2](#); [Figs 2 and 3](#)). This tool involves four domains: patient selection, index test, reference standard, and flow and timing. Each domain is evaluated in the form of risk of bias. Furthermore, a statement



**Fig 1. Flow diagram demonstrating the flow of studies through the review.**

<https://doi.org/10.1371/journal.pone.0261457.g001>

on concerns regarding applicability is provided for the first three domains. Signalling questions are incorporated to assist judge risk of bias. The included studies were categorized as having a “high,” “low,” or “unclear” overall risk of bias or concerns regarding applicability.

For analysis of risk of bias of studies investigating reliability the “Quality Appraisal of Diagnostic Reliability (QAREL)” tool [15] was used (Tables 3 and 4; Figs 4 and 5). The tool

**Table 1. QUADAS 2- tool for the evaluation of quality of the included validity studies on clinical assessments for the measurement of leg length discrepancy (n = 25).**

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Aguilar et al. (2017)	L	H	H	H	L	H	H
Aspegren et al. (1987)	U	H	H	U	U	H	H
Badii et al. (2014)	L	L	L	L	L	L	L
Beattie et al. (1990)	L	L	L	L	L	L	L
Betsch et al. (2019)	L	L	U	L	L	L	U
Cleveland et al. (1988)	L	L	L	L	L	L	L
Cooperstein et al. (2003)	L	L	H	H	L	L	H
Cooperstein et al. (2004)	L	L	H	H	L	L	H
Edeen et al. (1995)	L	L	L	L	L	L	L
Farella et al. (2005)	L	H	H	L	L	H	H
Friberg et al. (1988)	L	L	L	L	L	L	L
Gross et al. (1998)	L	L	L	L	L	L	L
Hanada et al. (2001)	L	L	L	L	L	L	L
Harris et al. (2005)	L	L	L	U	L	U	L
Jamaluddin et al. (2011)	L	L	L	U	L	L	L
Krettek et al. (1996)	L	L	L	L	L	L	L
Lampe et al. (1996)	L	L	L	L	L	L	L
Montgomery et al. (1995)	L	L	L	L	L	U	L
Neelly et al. (2013)	L	L	L	L	L	L	L
Petrone et al. (2003)	L	L	L	L	L	L	L
Piyakunmala & Sangkomkhang (2018)	L	U	L	L	L	U	L
Rhodes et al. (1995a)	L	H	H	L	L	H	H
Rhodes et al. (1995b)	L	L	L	L	L	L	L
Sayed-Noor et al. (2009)	L	L	H	L	L	L	U
Woerman & Binder-MacLeod (1984)	L	L	L	L	L	L	L

L = Low; H = High; U = Unclear.

<https://doi.org/10.1371/journal.pone.0261457.t001>

incorporates 11 items, asking about e.g. representativeness of the sample, blinding of raters, time interval between measurements, application and interpretation of the measurements, statistical methods of agreement, etc. Each item was judged in terms of its significance for the

**Table 2. QUADAS 2- tool for the evaluation of quality of the included validity studies on imaging diagnostics for the measurement of leg length discrepancy (n = 7).**

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Jensen et al. (2017)	L	L	L	L	L	L	L
Khakaria et al. (2011)	L	L	L	L	L	L	L
Kjellberg et al. (2012)	L	L	L	L	L	L	L
Rannisto et al. (2011)	H	U	L	H	H	U	L
Reina-Bueno et al. (2017)	L	L	L	L	L	L	L
Sabharwal et al. (2006)	L	L	L	L	L	L	L
Tipton et al. (2016)	L	L	L	L	L	L	L

L = Low; H = High; U = Unclear.

<https://doi.org/10.1371/journal.pone.0261457.t002>

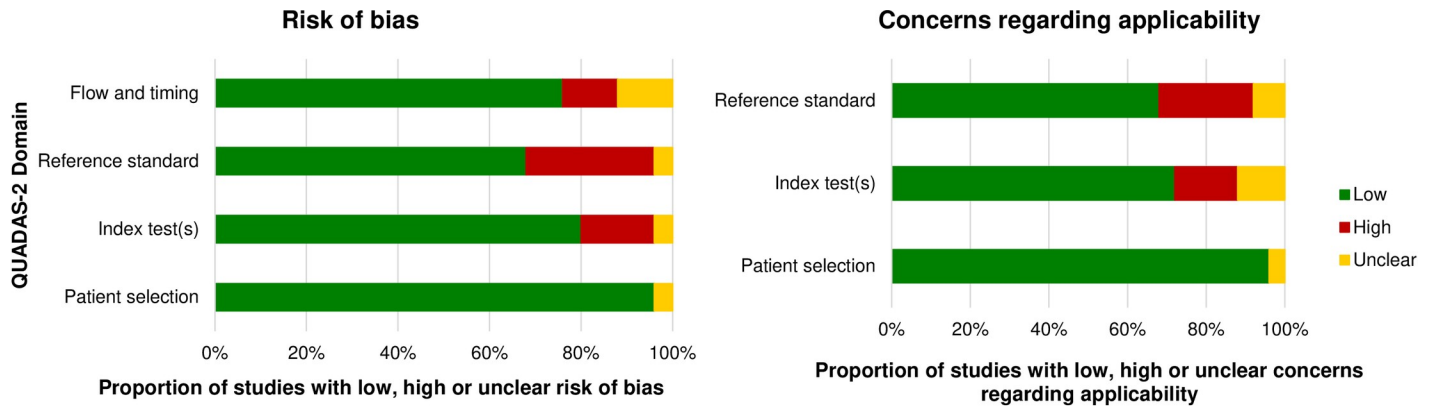


Fig 2. Proportion of studies on clinical assessments with low, high or unclear risk of bias and concerns regarding applicability assessed using the QUADAS-2 tool.

<https://doi.org/10.1371/journal.pone.0261457.g002>

quality of the study [16] by two authors (P.F.; M.A.) independently. If an item was judged with “yes” it was coded with “1”, resulting in a maximal achievable value of “11”. In the present study a value  $\leq 4$  represented a high, a value  $\geq 5$  to 7 a moderate and a value  $\geq 8$  a low risk of bias [17]. Any disagreements were discussed between reviewers and clarified. A third author (A.K.) was asked if agreement of evaluation could not be reached.

## Results

### Flow of studies through the review

The updated search revealed 3853 studies in PubMed and 677 studies in Index to Chiropractic Literature (Fig 1). After titles and abstracts were screened 296 studies were selected in PubMed and 67 studies in Index to Chiropractic Literature. The citations for the selected publications (n = 363) were imported into Endnote software, and duplicates (n = 11) were filtered automatically. After duplicates were removed (n = 11), abstracts of the remaining studies (n = 352) were assessed in depth. After that, 298 articles were excluded. Full texts of the remaining 54 articles were checked for eligibility criteria. Two articles [18, 19] were excluded for the following reasons: measurement of leg length using composite femurs [18]; correlation between anatomical leg length and perception of leg length discrepancy [19]. Of the remaining 52 articles, 37 articles on clinical assessments [20–56] and 15 articles on imaging diagnostics [57–71] were finally included and analyzed.

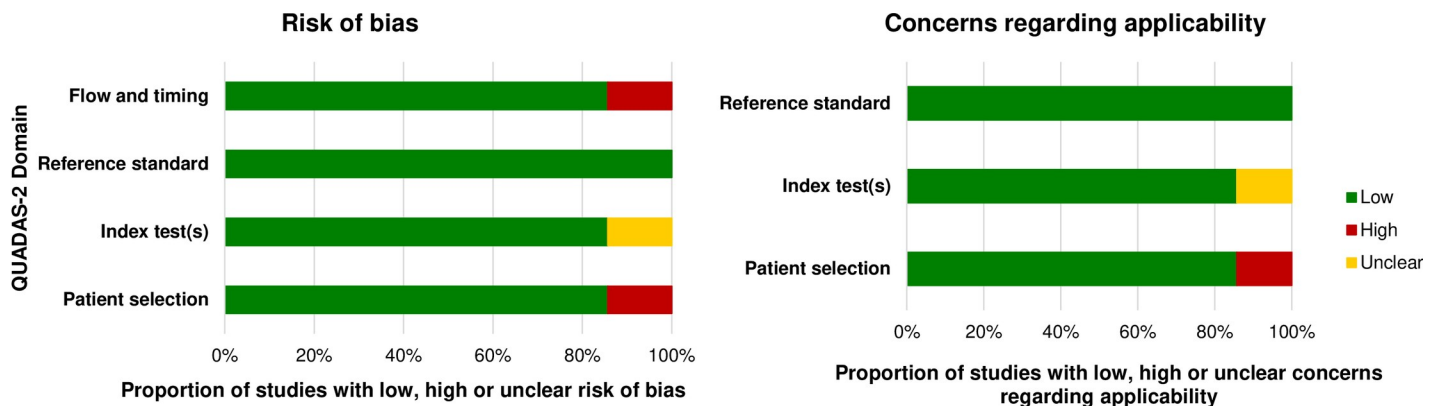


Fig 3. Proportion of studies on imaging diagnostics with low, high or unclear risk of bias and concerns regarding applicability assessed using the QUADAS-2 tool.

<https://doi.org/10.1371/journal.pone.0261457.g003>

**Table 3. QAREL-checklist for the evaluation of quality of the included reliability studies on clinical assessments for the measurement of leg length discrepancy (n = 12).**

Study	Items											Sum	Risk of bias (Quality)
	1	2	3	4	5	6	7	8	9	10	11		
Cooperstein et al. (2017)	N	U	U	U	N/A	U	U	N	N	Y	Y	2	High (Low)
Cooperstein & Lucente (2017)	Y	Y	Y	U	N/A	Y	Y	Y	N	Y	Y	8	Low (High)
De Boer et al. (1983)	N	Y	Y	U	N/A	Y	U	Y	N	Y	Y	6	Moderate (Moderate)
Hellsing (1988)	Y	U	U	U	N/A	U	U	U	N	Y	Y	3	High (Low)
Holt et al. (2009)	Y	Y	Y	N/A	N/A	U	U	Y	U	Y	Y	6	Moderate (Moderate)
Junk et al. (1992)	N	U	U	N/A	N/A	U	U	U	U	Y	N	1	High (Low)
Nguyen et al. (1999)	Y	Y	Y	N/A	N/A	U	U	Y	N	Y	Y	6	Moderate (Moderate)
Schneider et al. (2007)	Y	Y	Y	N/A	N/A	Y	U	U	U	Y	Y	6	Moderate (Moderate)
Schwartzbauer & Hart (2011)	N	Y	Y	N/A	N/A	U	U	N	U	Y	Y	4	High (Low)
Shambaugh et al. (1998)	U	Y	U	U	N/A	U	U	U	U	Y	N	2	High (Low)
Terry et al. (2005)	Y	Y	Y	U	U	Y	U	Y	N	Y	Y	7	Moderate (Moderate)
Woodfield et al. (2011)	N	Y	U	N/A	N/A	N	U	Y	U	Y	Y	4	High (Low)

N = No; Y = Yes; U = Unclear; N/A = Not applicable.

<https://doi.org/10.1371/journal.pone.0261457.t003>

### Quality of the included studies

Characteristics and main results of the included studies on clinical assessments and imaging diagnostics are provided in **S1 and S2 Tables**. Thirteen studies on the validity of clinical assessments [22, 23, 25, 31, 33–35, 41, 42, 44, 46, 48, 55] had a low risk of bias and low concerns regarding applicability for all domains using QUADAS-2 tool (**Table 1**). One study on the reliability of clinical assessments [27] had low risk of bias using the QAREL-checklist (**Table 3**). Six studies on the validity of imaging diagnostics [59–61, 67, 70, 71] had a low risk of bias and low concerns regarding applicability for all domains using QUADAS-2 tool (**Table 2**). Only one study on the reliability of imaging diagnostics [68] had low risk of bias using the QAREL-checklist (**Table 4**).

### Data synthesis

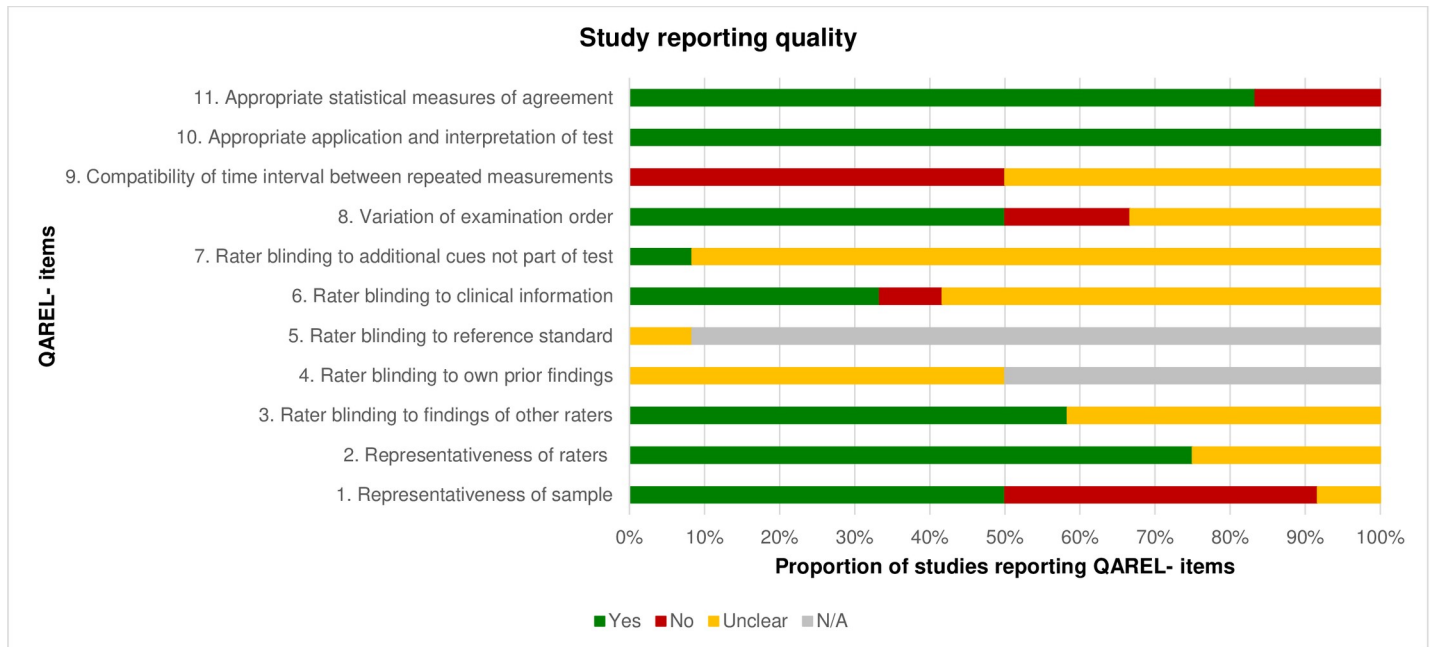
A quantitative data synthesis with analyses of sensitivity and specificity was not performed because of heterogeneity and nonuniformity of data across included studies. Eight relevant

**Table 4. QAREL-checklist for the evaluation of quality of the included reliability studies on imaging diagnostics for the measurement of leg length discrepancy (n = 8).**

Study	Items											Sum	Risk of bias (Quality)
	1	2	3	4	5	6	7	8	9	10	11		
Clavé et al. (2018)	Y	U	U	U	N/A	U	U	U	N	Y	Y	3	High (Low)
Guggenberger et al. (2014)	Y	Y	U	N/A	U	U	U	U	U	Y	Y	4	High (Low)
Konermann & Gruber (2002)	Y	U	Y	U	N/A	U	U	U	U	Y	N	3	High (Low)
Lazennec et al. (2016)	Y	U	U	N/A	N/A	U	U	U	U	Y	Y	3	High (Low)
Meermanns et al. (2011)	Y	Y	U	U	U	U	U	U	Y	Y	N	4	High (Low)
Poutawera & Stott (2010)	Y	Y	U	Y	N/A	U	U	Y	N	Y	Y	6	Moderate (Moderate)
Riad et al. (2010)	Y	Y	Y	Y	N/A	U	U	Y	Y	Y	Y	8	Low (High)
Sabharwal et al. (2007)	Y	Y	Y	U	N/A	U	U	U	N	Y	Y	5	Moderate (Moderate)

N = No; Y = Yes; U = Unclear; N/A = Not applicable.

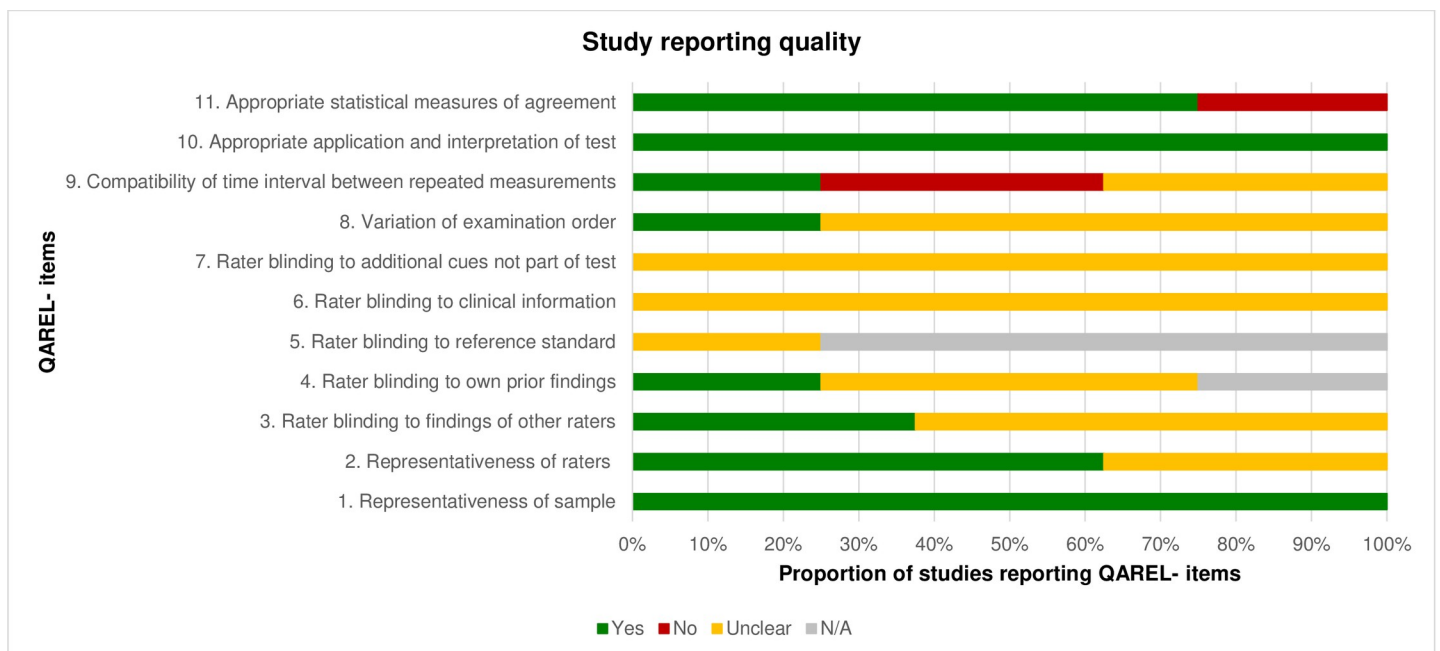
<https://doi.org/10.1371/journal.pone.0261457.t004>



**Fig 4. Assessment of study reporting quality of the included reliability studies on clinical assessments using the QAREL- checklist.** N/A = not applicable.

<https://doi.org/10.1371/journal.pone.0261457.g004>

validity studies on clinical assessments [20, 21, 24, 28, 29, 32, 49, 50] did not involve an adequate reference standard, needed to determine sensitivity and specificity [72]. Therefore, the results were summarized in a descriptive way. Validity and/or reliability of clinical assessments and imaging diagnostics were assessed in the included studies; results are reported in **S1 and S2 Tables**.



**Fig 5. Assessment of study reporting quality of the included reliability studies on imaging diagnostics using the QAREL- checklist.** N/A = not applicable.

<https://doi.org/10.1371/journal.pone.0261457.g005>



## Clinical assessments

**Tape measure method.** Sixteen studies reported on the determination of validity and/or reliability and/or accuracy of the tape measure method for measuring LLD [20, 22–25, 31, 33, 36, 39–42, 44, 50, 54, 55] (S1 Table). Further terms used were variability, variance or variation. The tape measure method was described as following: 1.) measurement from spina iliaca anterior superior (SIAS) to the medial malleolus [20, 23–25, 31, 33, 36, 39, 40, 42, 44, 50, 54, 55]; 2.) measurement of the distance between the SIAS and the lateral malleolus [20, 22, 41, 55]; 3.) measurement from SIAS to navel [55]; 4.) measurement from SIAS to xiphoid [55]; 5.) measurement from os pubis to malleolus medialis [31].

In four studies on tape measurement from spina iliaca anterior superior (SIAS) to the medial malleolus the reference standard CT-scanogram was used [23, 36, 39, 44], in one study the slit scanogram radiographic measurement [54], in one study the split scanogram including the hip, knee and ankle [55]. In eight studies [23–25, 31, 33, 42, 50, 55] different radiographic measurement methods served as the reference standard. In one study [40], an ultrasound measurement was performed, however, it appeared to be another index test rather than a reference standard. In one study on tape measurement from SIAS to the lateral malleolus, the radiography of the pelvis was used as reference test [22], in one study the slit scanogram radiographic measurement [54], in another study the split scanogram including the hip, knee and ankle [55], and in one study the ultrasound and teleroentgenography [41]. The tape measure methods from SIAS to the navel and from SIAS to the xiphoid were compared with the split scanogram including hip, knee and ankle as being the reference standard in one study [55]. In the study using the tape measure method with the landmarks os pubis and malleolus medialis the orthoroentgenography and the pelvic radiography were carried out as reference standards [31]. In one study no reference standard was used [20]. Only in a few studies [36, 39, 44, 54] it was reported, that the tape measure method is reliable and/or valid. In the most studies [22–25, 31, 33, 41, 42, 50, 55] it was concluded, that tape measurement methods were less accurate, revealed wide variation, weak correlation with other methods and disagreement with radiography, lead to miscalculation of small leg length discrepancy, and should be used with caution. There is high evidence that the tape measure method is less valid and reliable.

**Block test.** Validity and/or accuracy and/or reliability of the block test or standardized wooden boards was determined in eleven studies [21, 22, 31, 33, 36, 40–42, 47, 54, 55] (S1 Table). The reference standard was the radiography/roentgenography in eight studies [21, 22, 31, 33, 41, 42, 47, 55], the slit scanogram radiographic measurement in one study [54], the split scanogram including the hip, knee and ankle in one study [55], and the CT-scanogram in one study [36]. In another study, an ultrasound technique and a tape measure method was used, however, it appeared that all measurements were used as index tests rather than reference standard [40]. The block test was considered reliable, accurate and relevant or being superior to tape measure method in five studies [22, 36, 42, 54, 55]. However, poor validity and reliability compared with orthoroentgenography was reported in two studies [31, 47]. Furthermore, it was considered less accurate than ultrasound measurement [41] and radiographic measurement [33]. Two studies seemed methodologically insufficient [21, 40] so that a clear conclusion on accuracy, validity or reliability could not be provided based in their results. Overall, study results differed, so that there is moderate evidence that the block test is or is not valid and reliable for determining LLD.

**“Iliac crest palpation and book correction” method.** The reliability and validity of the “iliac crest palpation and book correction” method (ICPBC) was evaluated in one study [35] (S1 Table). Here, the standing radiographic measurement served as reference standard. Because of high reliability and fair validity of the ICPBC method, it was recommended to use

iliac crest palpation to identify LLD when there is no history of pelvic deformity and the iliac crests can be simply palpated. Based on only one high quality study with low risk of bias, the evidence that the ICPBC method is reliable and fairly valid is strong, however, further studies are needed to confirm this conclusion.

**Prone leg check methods.** In twelve identified publications [26–30, 38, 45, 48, 49, 51–53], reliability and/or validity of the measurement of leg lengths with the participant in prone position was examined (S1 Table). Thereby, the following methods were described: 1.) measurement of the difference between the posterior aspects of the soles using a millimeter ruler [48, 49]; 2.) activator method, i.e. applying pressure in cranial direction through the long axis of the legs using the thumbs and determining whether the leg lengths were equal or whether one leg was shorter compared to the other [45]; 3.) instrumented compressive leg check, i.e. the participants wore a modified surgical boot with a screw mounted at the medial aspects of the wooden left and right soles. During each measurement left or right shoe height was randomly increased using a maximum of 6 shims (0–6; 6 = 9.6 mm) to artificially induce LLD. The difference between the right and left leg length, that was displayed by the screws, was determined using a ruler that was placed at the end of the examination table between the participant's feet [28, 29]. The method was also used in a further study, however, without placing shims beneath the shoe sole [26]; 4.) assessment of leg length discrepancy in prone position by visually comparing the level of the medial malleoli [27]; 5.) determination of LLD with the participant in prone position with the knees extended and then flexed to 90° [30, 38, 51, 52]; 6.) determination of LLD with the participant in prone position with the knees extended and left head rotation [52]; 7.) measurement of LLD with the participant in prone position and the hip in extension [52]; 8.) Derifield-Thompson leg check [53].

A reference standard (standard radiography) was reported only in the studies by Rhodes et al. [48, 49]. The assessment of leg length discrepancy in prone position was predominantly considered reproducible and accurate [26, 28, 29, 38, 45, 49, 51, 53]. In a few studies, reliability was reported poor to fair [27, 30, 52]. The studies investigating validity of the prone leg check showed contradictory results, with one reporting good validity [48], and the other reporting poor validity [49]. There is low evidence that prone leg check methods are or are not valid and reliable for determination of LLD. Further high quality studies are needed to draw a valid conclusion.

**Hand-held devices.** Three studies were identified, where a hand-held device was used for the measurement of LLD [20, 34, 46]. Gross et al. [34] investigated the reliability and validity of using a calibrated pelvic leveling device (S1 Table). The reference standard was a radiographic measurement of the lower extremity. Petrone et al. [46] evaluated the reliability and validity of measuring the difference of iliac crest height (mm) using the PALM (S1 Table). The reference test was a pelvic radiography. Aguilar et al. [20] investigated the relationship between the method measuring the distance between malleolus medialis (DIMG) and lateralis (DEMG) and the ground and other clinical measurements, including the PALM and the pelvimeter. A reference standard was not used. Good reliability and moderate validity for the pelvic leveling device and the PALM were concluded in two studies [34, 46], however, in one of those studies [34] it was pointed out, that the results of reliability and validity do not support clinical use of the device to determine LLD, whereas the device was considered an alternative to radiographic measurement in the other study [46]. The third study [20] did not provide beneficial information on reliability and validity of the measurement of leg length discrepancy using the PALM and pelvimeter. Based on two useful high quality studies with low risk of bias, the evidence that hand-held devices for detecting LLD are moderately valid and reliable is strong. However, further high quality studies are needed to confirm this conclusion.

**Distance between the malleolus lateralis or medialis and the floor.** One study using the distance between the malleolus lateralis or medialis and the floor to determine LLD was identified [20]. Here, the relationship using the calculation of correlations between this index test and further clinical assessments measuring LLD in consideration of foot posture was investigated (S1 Table). Further methods were: 1.) tape measure method using the distance between anterior spina iliaca superior and malleolus medialis; 2.) tape measure method using the distance between anterior spina iliaca superior (ASIS) and malleolus lateralis; 3.) measurement of the difference between the left and right pelvic crest using a pelvimeter; 4.) measurement of the asymmetry between the left and right ASIS using a palpation meter. The authors concluded, that the index test was related to other methods to determine leg length asymmetry. However, they further recommended not to use the measurements with the malleolus lateralis and medialis alternatively, because they significantly differed. An appropriate reference standard was not included. There is low evidence that measuring the distance between the malleoli and the floor is a valid method for assessing LLD.

**Supine leg check method.** Six studies were found, in which reliability of LLD measurements with the participants in supine position was investigated [27, 31, 32, 49, 52, 56] (S1 Table). Validity was determined in three of those studies [31, 32, 49]. In one reliability study, the leg check was considered reliable [56]. In another reliability study, no agreement between the prone and supine leg check was concluded [27]. In the last reliability study, poor agreement was found for the supine leg check between examiners [52]. In the studies evaluating validity and reliability, poor validity of the supine leg check method was concluded [32, 49], which was also confirmed by Edeen et al. [31]. In one of those studies, poor reliability was found [32], in the other study reliability was reported excellent [49]. A reference standard was only included in two studies evaluating validity [31, 49]. Overall, there is low to moderate evidence that the supine leg check method is or is not valid and reliable for determining LLD. Further high quality studies are needed to draw a valid conclusion.

**Palpation and visual assessment.** Agreement of the palpation of the unilateral sacral prominence and radiography (reference standard) to determine LLD was investigated by Montgomery et al. [43] (S1 Table). Leg length was measured bilaterally from the upper edge of the acetabulum to the lower edge of the calcaneus. During the clinical measurement patients were positioned in prone and the assessor palpated the inferior angle of the sacrum, in order to identify unilateral sacral prominence. It was concluded that the palpation of the unilateral sacral prominence was as accurate as radiography for determination of an anatomical short leg under 9 mm. One study was identified using a visual assessment of leg length discrepancy during palpation of iliac spines [37], revealing some agreement between the results of three measurements. Based on only one study with acceptable quality, no valid conclusion on the evidence on validity and reliability of this method can be drawn.

## Imaging diagnostics

**Bi-planar imaging system EOS<sup>TM</sup>.** Three studies were included that used the bi-planar imaging system EOS<sup>TM</sup> for the determination of LLD [57, 59, 63]. In one study [59], preview images [radiation dose = 2.99 Gray x square centimeter (mGy x cm<sup>2</sup>)] recorded by the system were compared with diagnostic images (radiation dose = 239.26 mGy x cm<sup>2</sup>). Furthermore, reliability of the methods was determined (S2 Table). No significant difference was found between methods. The authors concluded that preview images are comparable to diagnostic images and could be used for follow-up examination to reduce radiation exposure. In the other studies [57, 63], 2-D- [63] and 3D- measurements [57, 63] using the EOS<sup>TM</sup>- system were considered accurate and highly reliable. However, both studies showed methodological

concerns. A reference standard was included in none of those studies. Based on only one high quality study with low risk of bias, and two low quality studies with high risk of bias, the evidence that the EOS<sup>TM</sup>- system is valid and reliable is low.

**Full-length standing anteroposterior radiography.** Four studies on the measurement of LLD using full-length standing anteroposterior radiography as index test were included [58, 67, 69, 70]. Here, high to almost perfect reliability of the methods was determined (S2 Table). Two studies investigated validity appropriately [67, 70]. The full-length standing radiography was considered valid. In one study it was concluded that full length standing anteroposterior radiography is superior to scanogram (reference standard) for the measurement of LLD [70]. Moreover, the authors recommended to commonly use the standing radiograph for assessing LLD in patients with angular deformities of the lower extremity, because it is similar to the measurement using the scanogram with a lower radiation dose. In summary, there is high quality evidence that full-length standing anteroposterior radiography is a valid method for determination of LLD, however, evidence regarding reliability is low to moderate.

**Scanogram.** One study was identified, where the reliability of LLD- measurements using CT- scanogram was investigated [65]. Although agreement of repeated measurements was almost perfect, CT- scanograms should be carried out more than once and should be double-checked by the surgeon. No reference standard was used. There is moderate evidence, that the scanogram is reliable for measuring LLD. No conclusion on validity can be drawn. High quality studies investigating validity of the scanogram are needed.

**Pelvic radiography with pelvic landmarks.** Two studies were identified where reliability and validity of pelvic radiography with pelvic landmarks was assessed [61, 71] (S2 Table). The method was compared with full-length standing radiograph [71] or CT-scanogram [61] (reference standards). The authors concluded that the use of pelvic radiography with landmarks should be considered cautious when determining LLD because of the limiting comparability to full-length standing radiography [71] and CT-scanogram [61]. A further study was found, where only reliability of the pelvic radiography for the measurement of LLD was assessed [64] (S2 Table). The method using the pelvic reference interteardrop line and center of the femoral head were considered the most accurate methods for the preoperative assessment compared to the measurement of true LLD using full-leg radiography. However, correlation or agreement between the methods were not determined, limiting this conclusion. There is strong evidence, that validity of the pelvic radiography with pelvic landmarks is low. There is low evidence that the method is reliable.

**Picture archiving and communication system (PACS).** One study was found that explored the comparability of accuracy and reproducibility of LLD measurement between PACS and a standard hard-copy radiograph [60]. Measurements were conducted by two raters independently. A comparable reliability was declared for both methods and an excellent agreement was found between both methods. Therefore, a change from hard-copy film to PACS was recommended (S2 Table). Based on one high quality study with low risk of bias, there is high evidence, that measurement of LLD using PACS is valid and reliable. However, further high quality studies are needed to substantiate this conclusion.

**Ultrasound methods.** One study, where the authors investigated the reliability of LLD measurement using the laser-based ultrasound method was found [66]. The ultrasound scanning head of the ultrasound apparatus was fixed at a rod and positioned perpendicular to the tissue interface of the anterior hip region. The distance between the highest point of the femoral head and the floor was measured by two blinded physiotherapists at two consecutive days. Reference standard was the radiographic measurement. The authors concluded an excellent reliability of ultrasound and agreement with radiographic measurement (S2 Table). Therefore, it could be an alternative to the radiographic measurement of LLD, because it is practicable

and non-invasive. Another study investigating the reliability of the sonographic measurement of LLD was identified [62]. A good reproducibility of the determination of total leg length, length of the femur and length of the tibia was concluded, however, the calculation of coefficients of correlation or agreement and the comparison to a reference standard lacked (S2 Table). Based on low quality studies with high risk of bias, there is low evidence, that ultrasound is a valid and reliable method to determine LLD.

**Magnetic resonance imaging.** One study was found where LLD was investigated using magnetic resonance imaging (MRI) [68] (S2 Table). Based on sagittal T1-weighted MRI images of the lower extremity, the length of the pelvis, femur, tibia and calcaneus was measured with the patients positioned in supine and the legs fully extended. The measurements were conducted by two experienced examiners and were repeated after two weeks. Excellent reliability was reported. There is high quality evidence, that magnetic resonance imaging is reliable for assessing LLD. However, high quality studies on the validity of magnetic resonance imaging are needed.

## Discussion

Leg length discrepancy is a common finding in the human population. Numerous methods are used in clinical practice. To the best of the authors' knowledge, this is the first systematic review summarizing study results on the validity and reliability of clinical assessments and imaging diagnostics for the measurement of LLD used in the clinical arena. The main finding was, that the use of almost all clinical assessments is questionable, because of strong evidence that the methods are not valid, reliable and accurate, i.e. tape measure method, or because of little evidence to draw any conclusion. Based on the results of the present review, only the block test can be recommended. Strong evidence for usage was only found for the full-length standing anteroposterior radiography, however, high radiation, high costs, low practicability, and limitations for determination of functional leg length discrepancy, may restrict its use in clinical practice. The use of the block test as well as full-length standing anteroposterior radiography for determination of LLD was also recommended by experienced orthopedists in a recent narrative review [73]. Following, the implication of especially these methods are discussed.

## Clinical assessments—Implication for clinical practice

The block test, which is considered the „indirect” method for measuring LLD [12], showed a higher reliability, sensitivity (55%) and specificity (89%) compared to using a tape measure (sensitivity = 45%; specificity = 56%) [22], which is considered the “direct” clinical measure method [12]. Both clinical assessments were described more reliable and relevant than CT-scanogram measurements for determination of LLD in patients after femoral fracture [36]. In only one out of 20 cases the block test identified the wrong leg as being shorter compared to pelvic radiography [22]. In contrast, other study results suggested the block test being not as accurate as imaging diagnostics and therefore, it should be used for screening of patients under suspicion of having LLD [12]. Furthermore, high values of interobserver variance (mean difference: 1.01 cm; 95% CI: 2.2 cm) led to the conclusion that the block test is unacceptable for clinical decision making [54]. It has to be considered, that contradictory results may have been caused by different ages of participants in the studies (23–85 years [22] versus 1.5–19 years [54]), respectively. Overall, sample sizes differed between studies of clinical assessments and may have influenced statistical analyses and interpretations.

The most included studies on the block test had an acceptable or even high quality, except for the studies by Aspegren et al. [21] and Junk et al. [40], because they showed high risk of

bias. Therefore, the block test seems to be beneficial, as it is completed with the participant in standing position and the difference of functional as well as anatomical leg length is captured simultaneously. However, palpation and visual analysis during measurement may represent potential sources of bias.

Devices, such as the palpation meter or the pelvic leveling device demonstrated good reliability [34, 46], however, for the latter, validity was only moderate as compared with imaging diagnostics (ICC = 0.64) [34]. In contrast to the pelvic leveling device, the determination of LLD using the PALM was considered an alternative to radiographic measurement. Therefore, using devices for assessing LLD may be an additional option in the clinical arena.

### Imaging diagnostics—Implication for clinical practice

Sabharwal & Kumar [12] described that the ideal diagnostic imaging procedure to measure LLD would be accurate, reliable, readily available and affordable. Moreover, it should allow for visualization of the complete lower extremity, should use a low radiation dose and have no magnification error. Using the teleoroentgenogram was considered the first choice for initial examination of patients with LLD [12] for a couple of reasons. The average magnification of the teleoroentgenogram compared to the scanogram was 4.6% [70]. An excellent correlation between measurements of the teleoroentgenogram and CT- scanogram was found. Intra- and interobserver- reliability of measurements of LLD using the teleoroentgenogram were demonstrated [69]. The CT- scanogram provided reliable measurements of LLD using minimal magnification as well [12]. However, radiation exposure of the scanogram was 1.6 to 3.8 times increased compared to the teleoroentgenogram [70].

The PACS may represent a reliable alternative to hardcopy radiography [60]. Although the respective study was rated being at low risk of bias and low concerns of applicability, the conclusion should be considered with caution, because data of standing limb images were analyzed retrospectively, which did not allow for analyzing possible confounding factors that may have been associated with the measurements [74].

Magnetic resonance imaging noninvasively produces high-resolution anatomy images and is useful for the diagnosis of various musculoskeletal conditions [75]. An excellent intra- and interobserver- reliability was observed for examination of LLD [68]. However, a reference standard lacked, so that no information about validity of results was available. Therefore, a comparison of LLD measurements using MRI with the gold standard, which is considered the radiographic scanogram, would be needed [12, 76]. An advantage of MRI is that it does not administer radiation to the patient, however, it may be more expensive, may require sedation in some patients, often needs a longer time to schedule and to carry out the examination, and may be not allowed in patients with specific implanted devices [12].

### Limitations

An analysis of sensitivity and specificity was only performed in four studies [22, 32, 41, 47] with one based on an insufficient reference standard [32]. Furthermore, eight studies on clinical assessments did not include a (valid) reference standard [20, 21, 24, 28, 29, 32, 49, 50]. As the reference standard is a decisive determinant of diagnostic accuracy of a test [77], the studies without a reference standard may not be considered a diagnostic accuracy study (<https://www.equator-network.org/reporting-guidelines/stard/>) and the intended use of QUADAS-2 quality assessment may have been limited. A transient solution for the analysis of studies without a reference standard may be using two critical appraisal tools [78]. Schröder et al. [78] evaluated the methodological quality of the reproducibility studies by assessing reliability using the reliability box of the Consensus-based Standards for the selection of health status

Measurement Instruments (COSMIN) and the validity studies including a reference standard using QUADAS-2. In the present study, the QAREL- tool was used as reported previously [79], because this tool was developed according to QUADAS-2 [14, 77]. Another challenge is the comparison of accuracy of multiple tests to assess the target condition compared to a reference standard (comparative study) and the evaluation of accuracy of a single-test to assess the target condition compared to a reference standard (single-test or non-comparative studies) [72]. A clinically relevant comparative question can result in a mix of single-test (non-comparative) studies and comparative studies and it is up to the review author how to tailor the QUADAS-2 quality assessment for the comparative question. This problem needs new statistical approaches that are under development. In the present study, a complete discrimination of methodological approaches in the analysis of the included studies, such as different study designs, populations, statistical analyses, number and variation of index tests, or use of reference standard could not be met. Furthermore, many terms of measurement quality criteria were used in the included studies. Thereby, it has to be considered, that there are fundamental differences between the concepts of reliability and agreement [80], which may not have been considered comprehensively among analysis of all studies. Therefore, the results of this systematic review on clinical assessments and imaging diagnostics to determine LLD should be considered as preliminary results, that need to be further investigated.

## Conclusion

The block test appears to be the most useful clinical assessment to measure LLD, followed by using devices, such as the pelvimeter. All other clinical tests seem to be not useful and may therefore be avoided by the clinicians and physiotherapists. Full-length standing anteroposterior radiography may be used as global reference standard to measure the anatomic LLD when comparing clinical methods and imaging diagnostics. The alternative measurements PACS and MRI need to be further investigated. The findings of the review may help clinicians to select the most valid, reliable and accurate clinical assessment to determine LLD in their patients and to make a decision, whether a compensation for LLD is indicated or not.

## Supporting information

**S1 Checklist. PRISMA 2009 checklist.**

(PDF)

**S1 Graphical abstract.**

(TIF)

**S1 Table. Characteristics and main results of included studies on clinical assessments (n = 37) for the determination of leg length discrepancy.**

(DOCX)

**S2 Table. Characteristics and main results of included studies on imaging diagnostics (n = 15) for the determination of leg length discrepancy.**

(DOCX)

## Author Contributions

**Conceptualization:** Martin Alfuth, Patrick Fichter.

**Data curation:** Martin Alfuth, Patrick Fichter.

**Formal analysis:** Martin Alfuth, Patrick Fichter, Axel Knicker.

**Investigation:** Martin Alfuth, Patrick Fichter.

**Methodology:** Martin Alfuth, Patrick Fichter.

**Project administration:** Martin Alfuth, Patrick Fichter.

**Resources:** Martin Alfuth, Patrick Fichter.

**Software:** Martin Alfuth, Patrick Fichter.

**Supervision:** Martin Alfuth, Patrick Fichter, Axel Knicker.

**Validation:** Martin Alfuth, Patrick Fichter, Axel Knicker.

**Visualization:** Martin Alfuth, Patrick Fichter, Axel Knicker.

**Writing – original draft:** Martin Alfuth, Patrick Fichter.

**Writing – review & editing:** Martin Alfuth, Patrick Fichter, Axel Knicker.

## References

1. Knutson GA. Anatomic and functional leg-length inequality: a review and recommendation for clinical decision-making. Part I, anatomic leg-length inequality: prevalence, magnitude, effects and clinical significance. *Chiropr Osteopat*. 2005; 13:11. <https://doi.org/10.1186/1746-1340-13-11> PMID: 16026625
2. Troyanovich SJ, Miller KJ. A clinical study of fluctuating asymmetry and leg-length inequality. *Vert Sublux Res* 2011; 12:92–8.
3. Kayani B, Pietrzak J, Hossain FS, Konan S, Haddad FS. Prevention of limb length discrepancy in total hip arthroplasty. *Br J Hosp Med (Lond)*. 2017; 78(7):385–90. <https://doi.org/10.12968/hmed.2017.78.7.385> PMID: 28692359
4. Menell JM. Joint pain. Boston: Little, Brown and Company, 1964:24.
5. Gurney B. Leg length discrepancy. *Gait Posture*. 2002; 15(2):195–206. [https://doi.org/10.1016/s0966-6362\(01\)00148-5](https://doi.org/10.1016/s0966-6362(01)00148-5) PMID: 11869914
6. Engelhardt P. [Slipped capital femoral epiphysis and the "healthy" opposite hip]. *Orthopade*. 2002; 31(9):888–93. <https://doi.org/10.1007/s00132-002-0377-y> PMID: 12232707
7. Knutson GA. Anatomic and functional leg-length inequality: a review and recommendation for clinical decision-making. Part II. The functional or unloaded leg-length asymmetry. *Chiropr Osteopat*. 2005; 13:12. <https://doi.org/10.1186/1746-1340-13-12> PMID: 16080787
8. Harvey WF, Yang M, Cooke TD, Segal NA, Lane N, Lewis CE, et al. Association of leg-length inequality with knee osteoarthritis: a cohort study. *Ann Intern Med*. 2010; 152(5):287–95. <https://doi.org/10.7326/0003-4819-152-5-201003020-00006> PMID: 20194234
9. Tallroth K, Ristolainen L, Manninen M. Is a long leg a risk for hip or knee osteoarthritis? *Acta Orthop*. 2017; 88(5):512–5. <https://doi.org/10.1080/17453674.2017.1348066> PMID: 28665226
10. Rannisto S, Okuloff A, Uitti J, Paananen M, Rannisto PH, Malmivaara A, et al. Leg-length discrepancy is associated with low back pain among those who must stand while working. *BMC Musculoskelet Disord*. 2015; 16:110. <https://doi.org/10.1186/s12891-015-0571-9> PMID: 25943907
11. Mannello DM. Leg length inequality. *J Manipulative Physiol Ther*. 1992; 15(9):576–90. PMID: 1469342
12. Sabharwal S, Kumar A. Methods for assessing leg length discrepancy. *Clin Orthop Relat Res*. 2008; 466(12):2910–22. <https://doi.org/10.1007/s11999-008-0524-9> PMID: 18836788
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009; 62(10):1006–12. <https://doi.org/10.1016/j.jclinepi.2009.06.005> PMID: 19631508
14. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011; 155(8):529–36. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009> PMID: 22007046
15. Lucas N, Macaskill P, Irwig L, Moran R, Rickards L, Turner R, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol*. 2013; 13:111. <https://doi.org/10.1186/1471-2288-13-111> PMID: 24010406
16. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010; 63(8):854–61. <https://doi.org/10.1016/j.jclinepi.2009.10.002> PMID: 20056381



17. Konieczka C, Gibson C, Russett L, Dlot L, MacDermid J, Watson L, et al. What is the reliability of clinical measurement tests for humeral head position? A systematic review. *J Hand Ther.* 2017; 30(4):420–31. <https://doi.org/10.1016/j.jht.2017.06.010> PMID: 28802538
18. Affatato S, Toni A. Leg length measurement: a new method to assure the correct leg length in total hip arthroplasty. *Med Eng Phys.* 2000; 22(6):435–9. [https://doi.org/10.1016/s1350-4533\(00\)00058-8](https://doi.org/10.1016/s1350-4533(00)00058-8) PMID: 11086255
19. Lazennec JY, Folinais D, Florequin C, Pour AE. Does Patients' Perception of Leg Length After Total Hip Arthroplasty Correlate With Anatomical Leg Length? *J Arthroplasty.* 2018; 33(5):1562–6. <https://doi.org/10.1016/j.arth.2017.12.004> PMID: 29310921
20. Aguilar EG, Dominguez AG, Pena-Algaba C, Castillo-Lopez JM. Distance Between the Malleoli and the Ground. A New Clinical Method to Measure Leg-Length Discrepancy. *J Am Podiatr Med Assoc.* 2017; 107(2):112–8. <https://doi.org/10.7547/15-013> PMID: 28394688
21. Aspegren DD, Cox JM, Trier KK. Short leg correction: a clinical trial of radiographic vs. non-radiographic procedures. *J Manipulative Physiol Ther.* 1987; 10(5):232–8. PMID: 3694061
22. Badii M, Wade AN, Collins DR, Nicolaou S, Kobza BJ, Kopec JA. Comparison of lifts versus tape measure in determining leg length discrepancy. *J Rheumatol.* 2014; 41(8):1689–94. <https://doi.org/10.3899/jrheum.131089> PMID: 25028369
23. Beattie P, Isaacson K, Riddle DL, Rothstein JM. Validity of derived measurements of leg-length differences obtained by use of a tape measure. *Phys Ther.* 1990; 70(3):150–7. <https://doi.org/10.1093/ptj/70.3.150> PMID: 2304973
24. Betsch M, Michalik R, Graber M, Wild M, Krauspe R, Zilkens C. Influence of leg length inequalities on pelvis and spine in patients with total hip arthroplasty. *PLoS One.* 2019; 14(8):e0221695. <https://doi.org/10.1371/journal.pone.0221695> PMID: 31454389
25. Cleveland RH, Kushner DC, Ogden MC, Herman TE, Kermond W, Correia JA. Determination of leg length discrepancy. A comparison of weight-bearing and supine imaging. *Invest Radiol.* 1988; 23(4):301–4. <https://doi.org/10.1097/00004424-198804000-00010> PMID: 3372194
26. Cooperstein R, Holzworth M, O'Brien A. Intra- and interexaminer reliability of compressive leg checking and correlation with the sit-stand test for anatomic leg length inequality. *Chiropractic Journal of Australia.* 2017; 45(2):184–95.
27. Cooperstein R, Lucente M. Comparison of Supine and Prone Methods of Leg Length Inequality Assessment. *J Chiropr Med.* 2017; 16(2):103–10. <https://doi.org/10.1016/j.jcm.2017.01.001> PMID: 28559750
28. Cooperstein R, Morschhauser E, Lisi A, Nick TG. Validity of compressive leg checking in measuring artificial leg-length inequality. *J Manipulative Physiol Ther.* 2003; 26(9):557–66. <https://doi.org/10.1016/j.jmpt.2003.08.002> PMID: 14673405
29. Cooperstein R, Morschhauser E, Lisi AJ. Cross-sectional validity study of compressive leg checking in measuring artificially created leg length inequality. *J Chiropr Med.* 2004; 3(3):91–5. [https://doi.org/10.1016/S0899-3467\(07\)60092-5](https://doi.org/10.1016/S0899-3467(07)60092-5) PMID: 19674629
30. DeBoer KF, Harmon RO Jr., Savoie S, Tuttle CD. Inter- and intra-examiner reliability of leg-length differential measurement: a preliminary study. *J Manipulative Physiol Ther.* 1983; 6(2):61–6. PMID: 6619669
31. Edeen J, Sharkey PF, Alexander AH. Clinical significance of leg-length inequality after total hip arthroplasty. *Am J Orthop (Belle Mead NJ).* 1995; 24(4):347–51. PMID: 7788314
32. Farella M, Michelotti A, Pellegrino G, Giani U, Martina R. Interexaminer reliability and validity for diagnosis of temporomandibular disorders of visual leg measurements used in dental kinesiology. *J Orofac Pain.* 2005; 19(4):285–90. PMID: 16279479
33. Friberg O, Nurminen M, Korhonen K, Soininen E, Manttari T. Accuracy and precision of clinical estimation of leg length inequality and lumbar scoliosis: comparison of clinical and radiological measurements. *Int Disabil Stud.* 1988; 10(2):49–53. <https://doi.org/10.3109/09638288809164098> PMID: 3403498
34. Gross MT, Burns CB, Chapman SW, Hudson CJ, Curtis HS, Lehmann JR, et al. Reliability and validity of rigid lift and pelvic leveling device method in assessing functional leg length inequality. *J Orthop Sports Phys Ther.* 1998; 27(4):285–94. <https://doi.org/10.2519/jospt.1998.27.4.285> PMID: 9549712
35. Hanada E, Kirby RL, Mitchell M, Swuste JM. Measuring leg-length discrepancy by the "iliac crest palpation and book correction" method: reliability and validity. *Arch Phys Med Rehabil.* 2001; 82(7):938–42. <https://doi.org/10.1053/apmr.2001.22622> PMID: 11441382
36. Harris I, Hatfield A, Walton J. Assessing leg length discrepancy after femoral fracture: clinical examination or computed tomography? *ANZ J Surg.* 2005; 75(5):319–21. <https://doi.org/10.1111/j.1445-2197.2005.03349.x> PMID: 15932444
37. Helsing AL. Leg length inequality. A prospective study of young men during their military service. *Ups J Med Sci.* 1988; 93(3):245–53. <https://doi.org/10.3109/03009738809178550> PMID: 2977001

38. Holt KR, Russell DG, Hoffmann NJ, Bruce BI, Bushell PM, Taylor HH. Interexaminer reliability of a leg length analysis procedure among novice and experienced practitioners. *J Manipulative Physiol Ther.* 2009; 32(3):216–22. <https://doi.org/10.1016/j.jmpt.2009.02.009> PMID: 19362232
39. Jamaluddin S, Sulaiman AR, Imran MK, Juhara H, Ezane MA, Nordin S. Reliability and accuracy of the tape measurement method with a nearest reading of 5 mm in the assessment of leg length discrepancy. *Singapore Med J.* 2011; 52(9):681–4. PMID: 21947147
40. Junk S, Terjesen T, Rossvoll I, Braten M. Leg length inequality measured by ultrasound and clinical methods. *Eur J Radiol.* 1992; 14(3):185–8. [https://doi.org/10.1016/0720-048x\(92\)90083-l](https://doi.org/10.1016/0720-048x(92)90083-l) PMID: 1563425
41. Krettek C, Koch T, Henzler D, Blauth M, Hoffmann R. [A new procedure for determining leg length and leg length inequality using ultrasound. II: Comparison of ultrasound, telerradiography and 2 clinical procedures in 50 patients]. *Unfallchirurg.* 1996; 99(1):43–51. PMID: 8850079
42. Lampe HI, Swierstra BA, Diepstraten AF. Measurement of limb length inequality. Comparison of clinical methods with orthoradiography in 190 children. *Acta Orthop Scand.* 1996; 67(3):242–4. <https://doi.org/10.3109/17453679608994680> PMID: 8686460
43. Montgomery DM, Egan IF, Pollard HP. Palpable unilateral sacral prominence as a clinical sign of lower limb anisomelia: a pilot study. *J Manipulative Physiol Ther.* 1995; 18(6):353–6. PMID: 7595109
44. Neelly K, Wallmann HW, Backus CJ. Validity of measuring leg length with a tape measure compared to a computed tomography scan. *Physiother Theory Pract.* 2013; 29(6):487–92. <https://doi.org/10.3109/09593985.2012.755589> PMID: 23289961
45. Nguyen HT, Resnick DN, Caldwell SG, Elston EW Jr., Bishop BB, Steinhouer JB, et al. Interexaminer reliability of activator methods' relative leg-length evaluation in the prone extended position. *J Manipulative Physiol Ther.* 1999; 22(9):565–9. [https://doi.org/10.1016/s0161-4754\(99\)70015-1](https://doi.org/10.1016/s0161-4754(99)70015-1) PMID: 10626698
46. Petrone MR, Guinn J, Reddin A, Sutlive TG, Flynn TW, Garber MP. The accuracy of the Palpation Meter (PALM) for measuring pelvic crest height difference and leg length discrepancy. *J Orthop Sports Phys Ther.* 2003; 33(6):319–25. <https://doi.org/10.2519/jospt.2003.33.6.319> PMID: 12839206
47. Piyakunmala K, Sangkomkamhang T. Measurement of Patient's Perception on Limb-Length Discrepancy Compared With Weight-Bearing Orthoroentgenography in Total Hip Arthroplasty: A Prospective Study. *J Arthroplasty.* 2018; 33(7):2301–5. <https://doi.org/10.1016/j.arth.2018.02.024> PMID: 29544973
48. Rhodes DW, Mansfield ER, Bishop PA, Smith JF. The validity of the prone leg check as an estimate of standing leg length inequality measured by X-ray. *J Manipulative Physiol Ther.* 1995; 18(6):343–6. PMID: 7595107
49. Rhodes DW, Mansfield ER, Bishop PA, Smith JF. Comparison of leg length inequality measurement methods as estimators of the femur head height difference on standing X-ray. *J Manipulative Physiol Ther.* 1995; 18(7):448–52. PMID: 8568426
50. Sayed-Noor AS, Hugo A, Sjoden GO, Wretenberg P. Leg length discrepancy in total hip arthroplasty: comparison of two methods of measurement. *Int Orthop.* 2009; 33(5):1189–93. <https://doi.org/10.1007/s00264-008-0633-9> PMID: 18677480
51. Schneider M, Homonai R, Moreland B, Delitto A. Interexaminer reliability of the prone leg length analysis procedure. *J Manipulative Physiol Ther.* 2007; 30(7):514–21. <https://doi.org/10.1016/j.jmpt.2007.07.001> PMID: 17870420
52. Schwartzbauer M, Hart J. Inter-examiner reliability of leg length inequality assessments. *Annals of Vertebral Subluxation Research.* 2011; 3:51–6.
53. Shambaugh P, Sclafani L, Fanselow D. Reliability of the Derifield-Thompson test for leg length inequality, and use of the test to demonstrate cervical adjusting efficacy. *J Manipulative Physiol Ther.* 1988; 11(5):396–9. PMID: 3235927
54. Terry MA, Winell JJ, Green DW, Schneider R, Peterson M, Marx RG, et al. Measurement variance in limb length discrepancy: clinical and radiographic assessment of interobserver and intraobserver variability. *J Pediatr Orthop.* 2005; 25(2):197–201. <https://doi.org/10.1097/01.bpo.0000148496.97556.9f> PMID: 15718901
55. Woerman AL, Binder-Macleod SA. Leg length discrepancy assessment: accuracy and precision in five clinical methods of evaluation\*. *J Orthop Sports Phys Ther.* 1984; 5(5):230–9. <https://doi.org/10.2519/jospt.1984.5.5.230> PMID: 18806409
56. Woodfield HC, Gerstman BB, Olaisen RH, Johnson DF. Interexaminer reliability of supine leg checks for discriminating leg-length inequality. *J Manipulative Physiol Ther.* 2011; 34(4):239–46. <https://doi.org/10.1016/j.jmpt.2011.04.009> PMID: 21621725
57. Clave A, Maurer DG, Nagra NS, Fazilleau F, Lefevre C, Stindel E. Reproducibility of length measurements of the lower limb by using EOS. *Musculoskelet Surg.* 2018; 102(2):165–71. <https://doi.org/10.1007/s12306-017-0518-4> PMID: 29094321

58. Guggenberger R, Pfirrmann CW, Koch PP, Buck FM. Assessment of lower limb length and alignment by biplanar linear radiography: comparison with supine CT and upright full-length radiography. *AJR Am J Roentgenol*. 2014; 202(2):W161–7. <https://doi.org/10.2214/AJR.13.10782> PMID: 24450698
59. Jensen J, Mussmann BR, Hjarbaek J, Al-Aubaidi Z, Pedersen NW, Gerke O, et al. Microdose acquisition in adolescent leg length discrepancy using a low-dose biplane imaging system. *Acta Radiol*. 2017; 58(9):1108–14. <https://doi.org/10.1177/0284185116682381> PMID: 28068825
60. Khakharia S, Bigman D, Fragomen AT, Pavlov H, Rozbruch SR. Comparison of PACS and hard-copy 51-inch radiographs for measuring leg length and deformity. *Clin Orthop Relat Res*. 2011; 469(1):244–50. <https://doi.org/10.1007/s11999-010-1460-z> PMID: 20625949
61. Kjellberg M, Al-Amiry B, Englund E, Sjoden GO, Sayed-Noor AS. Measurement of leg length discrepancy after total hip arthroplasty. The reliability of a plain radiographic method compared to CT-scanogram. *Skeletal Radiol*. 2012; 41(2):187–91. <https://doi.org/10.1007/s00256-011-1166-7> PMID: 21491155
62. Konermann W, Gruber G. [Ultrasound determination of leg length]. *Orthopade*. 2002; 31(3):300–5. <https://doi.org/10.1007/s00132-001-0298-1> PMID: 12017860
63. Lazennec JY, Brusson A, Rousseau MA, Robbins CB, Pour AE. Do Patients' Perceptions of Leg Length Correlate With Standing 2- and 3-Dimensional Radiographic Imaging? *J Arthroplasty*. 2016; 31(10):2308–13. <https://doi.org/10.1016/j.arth.2016.03.065> PMID: 27235329
64. Meermans G, Malik A, Witt J, Haddad F. Preoperative radiographic assessment of limb-length discrepancy in total hip arthroplasty. *Clin Orthop Relat Res*. 2011; 469(6):1677–82. <https://doi.org/10.1007/s11999-010-1588-x> PMID: 20878559
65. Poutawera V, Stott NS. The reliability of computed tomography scanograms in the measurement of limb length discrepancy. *J Pediatr Orthop B*. 2010; 19(1):42–6. <https://doi.org/10.1097/BPB.0b013e32833390b4> PMID: 19898257
66. Rannisto S, Paalanne N, Rannisto PH, Haapanen A, Oksaaja S, Uitti J, et al. Measurement of leg-length discrepancy using laser-based ultrasound method. *Acta Radiol*. 2011; 52(10):1143–6. <https://doi.org/10.1258/ar.2011.110268> PMID: 22090464
67. Reina-Bueno M, Lafuente-Sotillos G, Castillo-Lopez JM, Gomez-Aguilar E, Munuera-Martinez PV. Radiographic Assessment of Lower-Limb Discrepancy. *J Am Podiatr Med Assoc*. 2017; 107(5):393–8. <https://doi.org/10.7547/15-204> PMID: 29077503
68. Riad J, Finnbogason T, Brostrom E. Leg length discrepancy in spastic hemiplegic cerebral palsy: a magnetic resonance imaging study. *J Pediatr Orthop*. 2010; 30(8):846–50. <https://doi.org/10.1097/BPO.0b013e3181fc35dd> PMID: 21102211
69. Sabharwal S, Zhao C, McKeon J, Melaghari T, Blacksin M, Wenekor C. Reliability analysis for radiographic measurement of limb length discrepancy: full-length standing anteroposterior radiograph versus scanogram. *J Pediatr Orthop*. 2007; 27(1):46–50. <https://doi.org/10.1097/01.bpo.0000242444.26929.9f> PMID: 17195797
70. Sabharwal S, Zhao C, McKeon JJ, McClemens E, Edgar M, Behrens F. Computed radiographic measurement of limb-length discrepancy. Full-length standing anteroposterior radiograph compared with scanogram. *J Bone Joint Surg Am*. 2006; 88(10):2243–51. <https://doi.org/10.2106/JBJS.E.01179> PMID: 17015603
71. Tipton SC, Sutherland JK, Schwarzkopf R. The Assessment of Limb Length Discrepancy Before Total Hip Arthroplasty. *J Arthroplasty*. 2016; 31(4):888–92. <https://doi.org/10.1016/j.arth.2015.10.026> PMID: 26614748
72. Leeflang MMG, Reitsma JB. Systematic reviews and meta-analyses addressing comparative test accuracy questions. *Diagn Progn Res*. 2018; 2:17. <https://doi.org/10.1186/s41512-018-0039-0> PMID: 31093565
73. Vogt B, Gosheger G, Wirth T, Horn J, Rodl R. Leg Length Discrepancy- Treatment Indications and Strategies. *Dtsch Arztebl Int*. 2020; 117(24):405–11. <https://doi.org/10.3238/arztebl.2020.0405> PMID: 32865491
74. Sessler DI, Imrey PB. Clinical Research Methodology 1: Study Designs and Methodologic Sources of Error. *Anesth Analg*. 2015; 121(4):1034–42. <https://doi.org/10.1213/ANE.0000000000000815> PMID: 26378703
75. Dean Deyle G. The role of MRI in musculoskeletal practice: a clinical perspective. *J Man Manip Ther*. 2011; 19(3):152–61. <https://doi.org/10.1179/2042618611Y.0000000009> PMID: 22851878
76. Leitzes AH, Potter HG, Amaral T, Marx RG, Lyman S, Widmann RF. Reliability and accuracy of MRI scanogram in the evaluation of limb length discrepancy. *J Pediatr Orthop*. 2005; 25(6):747–9. <https://doi.org/10.1097/01.bpo.0000173246.12184.a5> PMID: 16294130

77. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003; 3:25. <https://doi.org/10.1186/1471-2288-3-25> PMID: 14606960
78. Schroder FF, Post CE, Wagenaar FBM, Verdonshot N, Huis In't Veld R. MRI as Diagnostic Modality for Analyzing the Problematic Knee Arthroplasty: A Systematic Review. *J Magn Reson Imaging.* 2020; 51(2):446–58. <https://doi.org/10.1002/jmri.26874> PMID: 31332865
79. MacKay JW, Low SBL, Smith TO, Toms AP, McCaskie AW, Gilbert FJ. Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis. *Osteoarthritis Cartilage.* 2018; 26(9):1140–52. <https://doi.org/10.1016/j.joca.2017.11.018> PMID: 29550400
80. Hernaez R. Reliability and agreement studies: a guide for clinical investigators. *Gut.* 2015; 64(7):1018–27. <https://doi.org/10.1136/gutjnl-2014-308619> PMID: 25873640