RESEARCH ARTICLE

Title: A Systematised Review of the Beighton Score Compared with Other Commonly Used Measurement Tools for Assessment and Identification of Generalised Joint Hypermobility (GJH)

Dr Malini Alexander[1]
[1]University of South Wales

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract. Background.

Objective. A systematised review compared validity and reliability of the Beighton Score to those of other commonly used scores for identification of generalised joint hypermobility (GJH)

Methods. Inclusion criteria: English language, studies on humans, all types of study designs, publications in academic journals, publications from the year two thousand onwards, publications in print and theses. Exclusion criteria: studies not in English, studies measuring single joints only, studies published before the year 2000, cadaveric studies, papers with only abstracts available. An electronic literature search was undertaken of Pub Med/MEDLINE, Embase, Scopus, Cochrane Database, SPORT Discus, Pedro databases, followed by a manual search. The final review included 73 papers. The PRISMA (2021) COSMIN (2010) guidelines and CASP (2019) criteria were used to evaluate methodological quality and bias.

Results. The Beighton Score's Intra-rater and inter-rater reliability ranged between ICC 0.74-0.99 and ICC 0.72-0.98 respectively. The BS has reasonable intra-rater and inter-rater reliability, however validity cannot be accurately determined as incorporation bias was identified as an issue in study methodology, not previously identified in the literature.

Conclusion. Paucity of data prevented accurate assessment of other scoring systems. Urgent research is required to clarify these issues and compare the BS to other tests. No source of funding was received in in undertaking this review. This review was not registered.

Keywords: Generalised Joint Hypermobility, Beighton Score, Hypermobile Ehlers-Danlos Syndrome, Heritable Disorders of Connective Tissue, Connective Tissue Weakness, Hypermobile Spectrum Disorder

1. INTRODUCTION

Joint hypermobility can be defined as the ability of a joint to move beyond the normal range of motion. Joint hypermobility might be limited to a single joint, peripheral, axial, present in a number of joints, or widespread throughout a individual's musculoskeletal system.

Despite prevalence of Generalised Joint Hypermobility (GJH) ranging from 2-57% in the general population, challenges with recognition using the current recommended Beighton Score exist. This leads to problems in conditions where (GJH) is the primary feature such as Hereditable Disorder of Connective Tissue including Joint Hypermobility Spectrum Disorders and Ehlers Danlos Syndrome Disorders.

*Address correspondence to the author by email at:
dr.alexander@rheumination.online

Failure to recognise GJH leads to poor patient outcomes [1]. Therefore, accurate clinical assessment is of paramount importance.

This literature review compares clinimetric properties of the Beighton Score against other scoring systems used to identify GJH, including the Hospital Del Mar/Bulbena Score, The Contompasis Score, The Rotès–Quérol Score, the Upper Limb Hypermobility Assessment Tool, The Lower Limb Hypermobility Assessment Tool, the Sasche Scale and several other tools including the 5pQ questionnaire.

This paper will answer the question: How does the BS compare with other widely-used methods (including complete joint examination) for identifying GJH in patients?

Terminology in GJH literature is imprecise and ambiguous resulting in confusion in clinical practice. In this paper the terms used to describe hypermobility are taken from Castori *et al* [2]. These include the following acronyms:

- GJH – Generalised Joint Hypermobility

- JH – Joint hypermobility
- HSD – Hypermobile Spectrum Disorder
- H-EDS – Hypermobile Ehlers Danlos Syndrome
- BS – Beighton Score
- ROM – Range of motion
- HDTC – Hereditable Disorders of Connective Tissue

Interchangeable use of terms such as ligament laxity, joint hypermobility, soft tissue fragility and joint instability creates confusion for clinicians. This is potentially misleading as these terms are not anatomically equivalent [2,3]. Although closely related these clinical phenomena are influenced by a range of anatomical and physiological mechanisms. Their relationship can be conceptualised schematically in Figure 1.

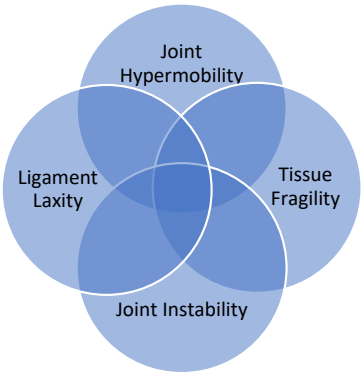


Fig. (1). Four Closely Related, but Non-Equivalent Terms: Ligament Laxity, Joint Hypermobility, Joint Instability and Tissue Fragility.

The BS is a modification of the 1964 Carter and Wilkinson scoring system developed by Beighton and Horan in 1969 to establish prevalence of GJH in an African population [4]. It is the most commonly used method for identifying GJH.

There is no Global consensus, or gold standard cut-off score that defines joint hypermobility [5], nor is there consensus on what constitutes normal ROM therefore the very definition of what constitutes hypermobility varies in the literature as well as clinical practice, however, the Ehlers Danlos Society [6] diagnostic criteria for H-EDS endorses the paper by Castori *et al* [2], recommending a range of age-adjusted cut off scores. Use of a goniometer to assess ROM is recommended by Schlager *et al* [7], and the Ehlers Danlos Society [6].

Some researchers do not recommend the BS as the primary tool for identification of GJH due to its limitations and diagnosis should rely on a comprehensive full body assessment of joint range of motion [2,8,9]. Castori *et al* [2], and Tinkle [10] report use of any single standardized measurement tool proves challenging.

Criticisms of using the BS to establish GJH include:

- Beighton, Solomon and Soskolne [11] did not provide any evidence-based justification for the selection of joints [8]
- Only 4 joint sites are measured [8]
- Validity not adequately researched [12]
- Appropriateness for paediatric populations [13]
- Inability to capture degree of hypermobility [8]
- Developed as an epidemiological tool [8,11]
- Inclusion of ligament laxity measurement [2]
- No consensus-based cut-off values [5]
- Bias towards upper limb hypermobility, that might fail to capture lower limb hypermobility resulting in false negatives [14]
- Only assesses ROM in 2 dimensions. For some joints ROM occurs in multi-dimensions [16]
- There are no consensus values for normal ROM [15–17] and the values chosen in the BS scoring system are based on tradition, rather than evidence

2. MATERIALS AND METHODS

2.1 Databases

A single researcher undertook this review. The search strategy was conducted in accordance with recommendations of Preferred Reporting Items for Systematic Reviews and Meta-Analyses PRISMA [18]. Guidelines for systematic reviews. The author did not use any software to assist in collection of the papers and used a purely manual approach.

Databases were searched during August to November 2021 and included Pub Med/ MEDLINE, Scopus, Cochrane, Database, SPORT Discus and Pedro. The following registers were searched over the same period: Australian New Zealand Clinical Trials Registry, Clinicaltrials.gov, Centrewatch.com, ISRCTN, EU Clinical Trials Register.

A manual search of bibliographies and references from appropriate papers identified, Google and websites listed below were then manually searched to identify additional literature such as theses and publications unavailable in the databases listed above. Websites included: The Ehlers Danlos Society Website, The American College of Rheumatology, EULAR, UpToDate.com and Medscape. The University of South Wales library website was utilised to retrieve papers unavailable in research databases, or in online searches.

No additional literature was sourced by contacting authors or experts.

Studies were grouped according to experimental and quasi-experimental studies, narrative and systematic reviews and grey literature with expert opinion.

2.2 Inclusion Criteria

English language

Humans

Studies on adults and children

Observational studies

Cohort studies

Cross sectional studies

Randomised Control Trials

All publications in academic journals (including expert opinion)

Studies published from the year 2000 onwards

Publications in print

Theses

2.3 Exclusion Criteria

Studies not in English, or not translated into English

Studies measuring a single joint only

Studies using the scoring system to measure a clinical presentation other

than GJH, or H-EDS

Studies published before the year 2000

Grey literature

Papers for which only abstracts are available

2.4 Outcome measures and measures of association

Inter-rater reliability

Intra-rater reliability

Validity

Sensitivity

Specificity

Any other reported statistics relating to test properties

All forms of measurement of strength of association reported in the literature were included

2.5 Search Strategy and Data Extraction

Each search term was manually entered to generate a number of papers. For each search the number was added up to gain the total number for each database. Each search was independently screened and papers whose titles obviously met the eligibility criteria, or exclusion criteria were saved into desktop folders, or excluded from the search accordingly.

This process generated a high number of duplicates each time an individual database was searched. Once the search was complete, the collected papers were then screened more closely to assess eligibility and exclusion criteria and duplicates were removed and deleted.

This resulted in a final list of papers for the electronic search component. For the manual search, narrative and systematic reviews and papers from the EDS Society website were used to search bibliographies of relevant studies. These papers were searched for in Google and, or sourced from the University of South Wales Library and downloaded into folders. Searching in the University library catalogue and Google generated additional papers of interest that were screened for inclusion and exclusion criteria. Duplicates were removed.

Papers not meeting inclusion criteria, or that met exclusion criteria were deleted. This resulted in the final selection of review papers. To access each paper, if it was not available in the database where it was originally cited, then a record was made and upon completion of the review these papers were searched for in other sources such as Research Gate, Academia.edu, Deepdyve, the University of South Wales Library databases, Google Scholar and Google. If free-access full text papers could not be sourced via this process, it was assumed only the abstract was available and such studies were excluded in accordance with exclusion criteria.

This review did not cover emails, other private correspondence, or letters and errata. There are no regulatory reviews relevant to GJH as far as the author is aware.

A number of measures were reported in the literature for assessment of validity and reliability. All measures were included in the results.

GJH has been considered as a diagnosis and therefore the Critical Appraisal Skills Programme (CASP) [19] for diagnostic tests was used to assess quality of papers along with The Joanna Briggs Levels of Evidence for Diagnosis [20] to rate methodological quality of studies. Additionally, use of a goniometer, or other assessment tools was included as part of the review as it is regarded as best practice when assessing ROM [21–23].

A note was made on whether studies referenced recommended Consensus-based Standards for selection of health Measurement Instruments (COSMIN) [24] Guidelines, or any other similar standards. Systematic and narrative reviews were assessed against the updated PRISMA [18] guidelines [25]. These were summarised in a separate table included in Supplementary Material C. A high degree of heterogeneity in results was anticipated, therefore a meta-analysis was not conducted.

Fig. (2). Prisma flow diagram [25].

Where possible this systematic review was conducted in accordance with the PRISMA Statement: An Updated Guideline for following a Systematic Review [25] as well as the PRISMA S Statement an Extension to the Prisma Statement for Reporting Literature Searches in Systematic Reviews [26].

4. RESULTS

The final review consisted of 73 papers. Fifty six papers of experimental and quasi-experimental studies, 3 systematic reviews, 9 narrative reviews, and 5 grey literature and expert opinion. Selected papers from initial screening, but subsequently excluded from the final analysis were recorded and are available from the authors upon request in supplementary material B.

Results are summarised in tables. Table 3.1 provides statistical results of reliability and validity in papers that directly assessed scoring systems. Table 3.1 includes colour coded information on goniometer use, cohort and patient characteristics. Additional information includes whether assessors were blinded and whether any studies discussed their protocols with reference to the COSMIN, or other relevant guidelines.

The Joanna Briggs Level of Evidence [20] for each study is reported. The majority were rated as a Level 2b-3b. Many studies did not report whether consecutive patients were assessed. If not reported it was assumed they were not consecutive patient studies.

If a paper did not explicitly state goniometer use, it was assumed none was. If a paper did not explicitly state whether blinding of researchers, or participants was undertaken, it was assumed no blinding occurred. Blinding in the context of these research scores is with regards to participants knowing which scoring system is being used, or whether examiners were aware of each other's scoring points, or what patients had been scored as previously.

Table 3.2 is a summary of the narrative and systematic review findings. Table 3.3 is table of expert opinion, or grey literature sources included in the study.

The tables where methodological quality was assessed according to the CASP [19] diagnostic tests and the PRISMA [18] checklist for reporting Systematic Reviews are included in Supplementary Material C and D respectively.

5. DISCUSSION

Systematic reviews, mainly included cross sectional studies, some included reviews of case control studies and were therefore given a rating of 3A instead of 2A. Narrative reviews rated 5A as per recommendations from the University of Canberra [27] suggesting narrative reviews are to be considered as expert opinion.

The PRISMA (2020) [25] checklist was used to evaluate both systematic reviews and narrative literature reviews to demonstrate any methodological flaws and group the reviews together.

Certain studies were designed to be association studies, rather than studies focusing specifically on clinimetric statistics and these were excluded from the review.

A statistical synthesis of results was not conducted due to heterogeneity.

Interpretation of reliability statistics can be problematic and non-straightforward [28]. Wide variation is reported for clinimetric statistics with no standardised reporting making direct comparisons difficult. This includes interpretation of effect sizes. Few papers addressed strength of association.

There was inconsistency of reporting inter-rater reliability, but not intra-rater reliability. Only some papers assessed validity and no papers assessed responsiveness.

Compared to the BS there is sparce literature on clinimetric properties of alternative scoring systems.

Intra-rater reliability scores for the BS range from ICC 0.74-0.99, K = 0.82. Inter-rater reliability score ranged from ICC 0.72-0.98, spearman rho = 0.86. For the other scoring systems assessed the score ranges are summarised in table 3.

Johnson, Ward and Simmonds [29] found a strong correlation between LLAS and BS scores (rho = 0.732; $p < 0.001$). The ULHAS had an intra-rater reliability ICC of 0.92 [30]. No data is available on inter-rater reliability, or validity and so a direct comparison between ULHAS and the BS cannot be made.

Vallis, Wray and Smith [31] found both inter-rater reliability and intra-rater reliability of Contompasis inferior to those of BS Inter-rater reliability ICC range 0.58-0.62 and intra-rater reliability ICC 0.73-0.82 versus Inter-rater reliability ICC range 0.72-0.80, intra-rater reliability ICC range 0.71-0.82, However Schlager *et al* [7] found both inter-rater reliability and intra-rater reliability of BS inferior to that of Contompasis Inter-rater reliability 0.82 (95% CI 0.69-0.89) and Intra-rater reliability 0.79 (95% CI 0.57-0.90) vs Inter-rater reliability 0.72 (95% CI 0.55-0.83) and Intra-rater reliability 0.76 (95% CI 0.54-0.88), respectively. Both these studies were performed in healthy adult populations, with Vallis Wray and Smith [31] selecting the sample from physiotherapy students at a university.

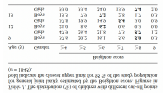
There was only a single paper for reliability statistics for the Rotes-Querol Score by Juul-Kristensen *et al* [32] reporting a Kappa range of 0.31-0.80. This range is lower than the

Table 1. Cross Sectional Studies. JBI Level 4B.

Author	Assessment Tools	Participant characteristics + country of study	Goniometer Used	Inter-rater reliability	Intra-rater reliability	Accuracy	Other statistics
Ahlqvist *et al* [33]	**5pQ**	2455 Swedish-speaking pregnant women were consecutively recruited at their first visit for registration in the national antenatal screening programme, mean age 29 years (SD 5 years) (Sweden)	**N/A**	**N/A**	**N/A**	**N/A**	Women with GJH also had higher odds of PGP during the entire pregnancy (adjusted odds ratio (aOR) 1.27: 95% CI 1.11–1.47) and in trimester 1 (aOR 1.54: 95% CI 1.20–1.96), but the associations were not statistically significant in trimester 2 (aOR 1.24: 95% CI 0.82–1.88) or trimester 3 (aOR 1.20: 95% CI 0.99–1.45). The odds of PGP in pregnancy increased with increasing numbers of positive answers to the 5PQ (p for linear trend < 0.001) for the entire pregnancy and in trimester 1 (p for linear trend < 0.001), but not in trimesters 2 or 3 (p = 0.13 and p = 0.06, respectively
Antonio and Magalhaes [34]	**BS** **SF36**	88 healthy volunteers 299 females 89 males Age range 18 -25 years medical and physiotherapy courses (Brazil)	No			Cut off ≥ 4 Prevalence of GJH 26.8%	Rho correlations between BS and SF-36 0.1 to 0.3
Armstrong [35]	**BS** **BRIGHTON CRITERIA**	Female and male rugby players, female netball players, female dancers and male and female age matched controls (UK)	Yes	**N/A**	Intrarater reliability ICC of 0.992 (95% CI 0.979-0.997)	Removal of lumbar flexion from the BS resulted in no change in "not hypermobile" (NH) scores across the three classifications in male rugby and an increase of 5% (n = 2) in male controls (BE and SB). In netball players, the B classification (0-2) increased by 9.8% (n = 6)	in comparison to the BS and BE classification increase of 1.6% (n = 1) while female rugby remained similar at 3% (n = 2) (BE and SB) and 8% (n = 5) (B). Female dancers demonstrated large changes in the B classification 'moderately hypermobile" (MH) (3-4) with an increase of 33.3% (n = 14) in contrast to a decrease -4.8% (n = -2) and increase of 11.9% (n = 5) respectively in the BE (≥4) and SB "hypermobile" (4-6) classifications. This highlights classification system variation and influence of lumbar flexion inclusion
Armstrong and Greig [36]	**BS**	Sixty-five female rugby players, 38 male rugby players, 61 netball players, 42 female dancers, 40 male controls and 40 female controls (UK)	Yes	N/A	Intra-rater reliability ICC 0.992 (CI 0.979-0.997)	N/A	Significant differences existed for group and gender analysis at the left and right 5th metacarpophalangeal joints, left and right thumb, left and right elbow and lumbar spine (p < 0.001). Lumbar flexion demonstrated significant x2 values and large effect sizes for all groups. This effect size was reduced to a moderate effect size when male against female analysis was performed and joint hypermobility was greater in females in comparison to males
Aslan *et al* [37]	**BS**	A total of 72 undergraduate physical therapy students, aged 18 to 25 (29 females and 43 males) (Age range female 20.2±0.9 male 20.5±1.5) (Poland)	Yes	Inter-rater reliability ICC = 0.82	Intra-rater reliability 0.92	N/A	BS is a good score for identifying GJH in healthy subjects
Bale *et al* [38]	**BS** **Brighton Criteria**	120 Children aged 5–16 years old were recruited from secondary paediatric care ?	**No**	N/A	N/A	(28.3%) were classified as having JHS using the Brighton criteria. Of these 58.8% were female. Of the 34, 28 (82.4%) met both major criteria one of which is Beighton score >4. The commonest site of reported pain was in the muscles of the legs (27.5%). 64 (53.3%) children met at least one minor criteria however of these, 28 (43.4%) scored for arthralgia in 1–3 joints.	The other 6 met classification due to one major criteria (Beighton >4) with 2 minor criteria (2 (5.8%) due to arthralgia and skin extensibility, 1 (2.9%) due to arthralgia and dislocations, 1(2.9%) due to arthralgia and dislocations, 1(2.9%) due to skin extensibility and eye signs (myopia), 1 (2.9%) due to skin extensibility and dislocations and 1 (2.9%) due to arthralgia, skin extensibility and multiple dislocations). Of the children meeting criteria for JHS, 12 (33.3%) had a family history. A further 6 cases of JHS could be classified by the Brighton criteria if their family history was confirmed The commonest site of reported pain was in the muscles of the legs (27.5%). 64 (53.3%) children met at least one minor criteria however of these, 28 (43.4%) scored for arthralgia in 1–3 joints.

Boyle, Witt and Riegger-Krugh [39]	**BS**	Students from the Chapel Hill High School soccer team and the University of North Carolina at Chapel Hill physical therapy program. Forty-two (intrarater) and 36 (interrater) female volunteers, aged 15 to 45 years (mean age was 25.4 6 4.2 years) (USA)	Yes	Inter-rater reliability Spearman rho = 0.87 (p<0.0001)	Intra-rater reliability Spearman rho = 0.86 (p<0.0001)	N/A	N/A
Bulbena *et al* [40]	**Screening Questionaire to Detect Hypermobility** **SQCH** **Hospital Del Mar** **5PQ**	158 patients between 18 and 60 years old were consecutively recruited from an anxiety outpatient unit belonging to a general university hospital. 142 females, 16 males (Spain)	No	SQCH: Inter-rater reliability ICC= 0.961 CI (0.922-0.980)	N/A	SQCH: GJH 44.3% PPV 92.79% NPV 46.22% "Wrongly classified 2.48%" The area under the ROC curve of the 7 self-assessed criteria (SQ-CH) is 0.861.Sensitivity and specificity percentages for the 2/3 cut-off point in relation to the Hospital del Mar criteria are 78.0% and 75.8% respectively, obtaining a percentage of 65.75% correctly classified as positive and 85.2% correctly classified as negative. Likelihood ratios are positive and significant for each item according to the cut-off point in relation to the total score in SQ-CH Hospital Del Mar: For presence of GJH 37.3% (rho = 0.745; p<0.001) Sensitivity 78% specificity 75.8% 2/3 cut-off point True positives 65.75% True negatives 85.2% 5PQ: For presence of GJH (rho = 0.857; p<0.001)	N/A
Calhil *et al* [41]	**Marshall test** **BS**	204 healthy patients 2-18 years old presenting to a general paediatric orthopaedic clinic (111 female, 93 male) average age 10.7 years (USA)	No	N/A	N/A	BS: Prevalence of GJH = 13.3% Marshall test: PPV = 34% NPV = 99% Sensitivity =99% Specificity =67% Positive likelihood ratio = 3.3	N/A
Chan *et al* [42]	**BS** **LLAS** **Villefranche vs Brighton Score**	85 dancers from two dance institutions 74 female 26 male Mean age 21.2 (Australia)	No	Villefranch vs Brighton score: 54% disagreement	N/A	BS: Cut off point ≥ 5/9) GJH=72% LLAS: Cut off point ≥ 7/12 GJH= 38% and 42% met the LLAS cut-point on the left and right respectively The Villefranche identified more dancers with JHS/EDS-HT than the Brighton (84% vs 31%, p < 0.001)	72% of dancers have GJH according to BS, while 38% and 42% met have GJH when using LLAS cut-point on the left and right respectively. Between the scoring methods of BS and LLAS there is a disagreement of 48% and 46% respectively, with Beighton classifying more participants as having GJH. A higher Beighton cut-point, e.g. 6/9, to identify true GJH amongst dancers maybe warranted.
Cooper *et al* [43]	**Self-reported line drawings** **Modified BS cut off >=4**	50 volunteer participants 22 male 28 female Median age 49 (UK)	Yes	Inter-rater reliability k= 1.00; (95% CI 1.00-1.00) Self-reported participant-repeatability k =0.91 (95% CI 0.74-1.00)	N/A	Sensitivity of 0.87 (95% CI 0.81, 0.91) Specificity of 0.99 (95% CI 0.98, 1.00)	The self-reported instrument provides a valid and reliable assessment of the presence of generalised joint hypermobility and may have practical use in epidemiological studies

Czaprowski *et al* [44]	**BS** **Straight leg raise (SLR)** **Thomas test for one- (O-JHF)** **two-joint (T-JHF) hip flexors** **fingertip-to-floor (FTF) and lateral trunk flexion (LTF) tests**	249 paediatric patients (136 females and 113 males) Average age 11.7/11.8 SD 0.7/0.8 (Poland)	No	N/A	N/A	A positive result of BS was observed only in 8 females and in 3 males with GJH (22% and 16% all of females and males with GJH, respectively). We suspect that this finding may be due to the influence of both, mobility of the spine and hamstring flexibility (that was low in GJH groups with the mean range of hip flexion 57.3°±7.7 and 53.1°±9.3 for females and males with GJH, respectively) on FTF test result	Clinical examination of the pelvic-hip complex muscles and trunk flexibility by use of SLR, O-JHF, T-JHF, FTF and LTF revealed to be insufficient in diagnosing GJH in children aged 10–13 years. Thus, the Beighton scale should be considered a standard element of physiotherapeutic examination of the musculoskeletal system in children and youth.
Evans, Rome and Peet [45]	**BS** **LLAS**	Participant characteristics: 30 healthy asymptomatic children aged 7-15 years children were recruited as a convenience sample from the Auckland University of Technology podiatry clinic and from staff associated with this clinic (New Zealand)	No	BS: Inter-rater reliability ICC = 0.73 LLAS: nter-rater reliability ICC = 0.78	BS: ntra-rater reliability ICC = 0.96-0.98 LLAS: Intra-rater reliability ICC = 0.90-0.98	N/A	N/A
Farmer *et al* [46]	**5pQ** **BS** **Skin Hyperextensibility** **N/A (skin measured with a harpenden caliper)** **CSES (Corrected Skin Extensibility Score)**	250 healthy volunteers 131 female, with a mean age of 39 years (range 18–89 yrs) and mean body mass index (BMI) 25.4 kg/m2 (range 16.7–45.7 kg/m2). Ethnically, 78.4% were Caucasian, 13.6% were South Asian, 5.6% were Afro-Caribbean, and 2.4% were East Asian (Japanese and Chinese). The mean Beighton score was 1.8 (range 0–8). Forty-four out of 250 (17.6%) of the cohort had a Beighton score ≥ 2 (UK)	No	CSES: Inter-rater reliability ICC = 0.96 Kappa = 0.83		BS: Cut off ≥ 4/9 Sensitivity = 0.853 Specificity = 0.85 AUC = 0.90 (95% CI 0.84–0.96) CSES: Sensitivity = 0.72 Specificity = 0.75 AUC = 0.86	The mean Corrected Extensibility Score was 23.84%/mm in participants found to be hypermobile versus 13.55%/mm in the normal mobility group ($p < 0.0001$). CSES sensitivity was 0.72, specificity 0.75. The κ value for interobserver variability was 0.83.
Glans *et al* [47]	**BS -self reported (Translated into Swedish)** **N/A**	328 participants aged 18–65 years fluent in the Swedish language (Sweden)	N/A	N/A	Intra-rater reliability ICC = 0.92 (95% CI 0.85, 0.96)	Sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72) 80%)ROC = 0.87 (95% CI 0.79, 0.95)	Intra-rater reliability ICC = 0.92 (95% CI 0.85, 0.96)
Hakim and Grahame [48]	**Retraction of hamstrings as an indicator of GJH**	28 participants aged 18–65 years fluent in the Swedish language (Sweden) 212 Consecutive patients from the Joint Hypermobility Clinic 30 were hypermobile, 182 had joint hypermobility syndrome (UK)	No	N/A	N/A	87.5% of sample had retraction of hamstrings Retractions of the triceps surae were found in 90.9% of patients, and retractions of the soles of feet were observed in 95.9% of patients The impact of retraction on the Beighton palms-on-the-floor test is very great indeed 97.8% of patients who present a retraction of the hamstrings of over 45° cannot perform this manoeuvre	The presence of muscle and tendon retractions in the posterior muscle compartments of the lower limbs and the soles of the feet constitute clinical features of Ehlers-Danlos syndrome. They should be addressed with a view to prevention and treatment, mainly through physical therapy
Hansen *et al* [49]	**Compared Parents answers to physicians' assessment using pictures and questionnaires to arrive at a BS measurement**	A total of 188 98 female, 90male from sports clubs in Copenhagen aged 9-13 years (Denmark)	varying kappa values for the individual components of BS between physicians and between physicians and parents. Kappa ranges from <0.44 to 1.0 Lowest agreements was for elbow and knee hyperextensibility Highest kappa ranges were between the 2 experienced physicians, lowest between these physicians and parents	N/A	N/A	The study illustrates that skilled physicians have a low inter observer variability when testing children for joint hypermobility. Untrained parents were unable to identify hypermobility in their children using the BS	

Jannson *et al* [50]	**BS**	1845 children were clinically examined concerning general joint laxity in grades three (n = 573; 317 boys, 256 girls), six (n = 703; 349 boys, 354 girls) and nine (n = 569, 284 boys, 285 girls). The mean ages of the children were 9 y (95% CI: 8.99–9.10) in grade 3, 12 y (95% CI: 11.93– 12.03) in grade 6 and 15 y (95% CI: 14.93–15.05) in grade 9 (Sweden)	No	N/A	N/A	N/A	The largest difference concerning cut-off point is found at the age of 15 y. According to this method, girls were considered to be hypermobile if 8 manoeuvres were performed correctly. The cut-off for boys at the same age would be at 6 manoeuvres, if correctly performed. 
Johnson, Ward and Simmonds [29]	**LLAS**	Thirty-six male, professional footballers aged between 18 and 37 years old (UK)	No	N/A	N/A	cut off ≥4/10 Sensitivity = 67% Specificity =94% Positive Predictive Value (50%) Negative Predictive Value (97%).	
Junge *et al* [51]	**BS (Method A)** **BS (Method B)**	1300 children in the Municipality of Svendborg aged (7-8 years) and fourth grade (10-12 years) (Denmark)	Yes	Method A: kappa = 0.49-0.94 Method B: Kappa= 0.30-0.84	N/A	Cut off ≥5 Both methods need to be tested for their predictive validity at higher cut-off levels, e.g. ≥6 and ≥ 7	
Kulik and Gebska [52]	**BS** **Brighton Criteria**	102 students (60 boys, 42 girls) aged between 6 to 11 years (Poland)	No	N/A	N/A	BS: Prevalence of GJH in sample = 34.3% Mann-Whitney U test, p = 0.085 Brighton Criteria: Prevalence of GJH in sample = 0.98%	The occurrence of joint hypermobility in children using the Beighton score is greater than using the Brighton criteria. The Beighton score and Brighton's criteria are not well correlated, so a standardized method for diagnosing hypermobility should be developed
Kwon *et al* [53]	**BS**	404 healthy female paediatric patients and 266 adult female patients Age range 6-12 in girls and 24-50 in adult females (Korea)	Yes	N/A	N/A	Cut off = 4/9 GJH present in 238 girls (58.9%) and 97 women (36.5%)	Recommend setting 5/9 as cut off in paediatric patients
Lamari, Chueire and Cordeiro [54]	**BS**	1,120 healthy asymptomatic children of mixed racial background (534 boys, 47.7%; 586 girls, 52.3%; age range: 4-7 years) (Brazil)	No	N/A	N/A	Cut off> 4 GJH present in 64.6% of the children	Recommend additional methodological parameters and criteria to characterize joint mobility in addition to BS
McGillis *et al* [55]	**BS**	A sample of 298 patients referred to the Goodhope Ehlers Danlos clinic, 156 with a previous diagnosis of EDS (Canada)	No	N/A	N/A	N/A	Only 46% (n = 51) of HSD/LJH patients scored a BS of 4/9 as assessed by an EDS practitioner (SD = 2.12). In the H-EDS group the average BS was 6/9 (SD = 1.45). The HSD/LJH patients were reviewed for a comparative analysis of BS as assessed by the referring physician compared to the objective assessment by the EDS specialist using a goniometer. 82% (n = 91) of patients had a BS completed by their referring physician on initial referral. It was found that the average BS on referral was 6/9 (SD = 2.17). 81% (n = 74) of referring physicians assessed the BS as higher than the assessment of the EDS practitioner, 14% (n = 13) had the same BS and only 5% (n = 4) had a BS lower on the referral than as assessed in the EDS clinic. The findings call into question the validity of solely applying the BS as a measure of GJH in this population, or alternatively supports the notion that those symptomatic patients with a low BS have a different disorder entirely
Mikołajczyk *et al* [56]	**Sasche Scale**	Participant characteristics: 120 healthy asymptomatic children aged 15 years of age 60 girls and 60 boys (Poland)	No	N/A	N/A	Cut off ≥ 7/13 30% positive	Not compared to BS

De Moraes *et al* [57]	**BS**	Portuguese was applied to 2,523 individuals, of whom one was an elementary school student and 2,522 attended the following three universities: USP, at the Ribeirão Preto Campus; Universidade de Franca (Unifran); and Centro Universitário Barão de Mauá, in the city of Ribeirão Preto. The university students attended the following courses: medicine (1st to 4th year, 609 students); nursing; psychology; physical therapy; occupational therapy; law; chemistry; medical physics; and speech therapy (Brazil)	No	N/A	N/A	For cut off ≥4 prevalence of GJH= 48.0% (60.6% and 36.7% in girls and boys, respectively) OR, 1.29; 95% CI, 1.12-1.49 For cut off ≥6, GJH = 18.6% (26.1% and 11.5% in girls and boys, respectively) OR, 1.22; 95% CI, 1.01-1.47	N/A
Morris *et al* [58]	**BS**	1584 participants at 14 years of age taken from pregnancy cohort at King Edwards Menorial Hospital Western Australia (Australia)	No	N/A	N/A	For cut off ≥4 prevalence of GJH= 48.0% (60.6% and 36.7% in girls and boys, respectively) OR, 1.29; 95% CI, 1.12-1.49 For cut off ≥6, GJH = 18.6% (26.1% and 11.5% in girls and boys, respectively) OR, 1.22; 95% CI, 1.01-1.47	
Naal *et al* [59]	**BS-self in FAI (FAI is an indicator for GJH)**		No	N/A	N/A	Cut off ≥4 Prevalence of GJH = 32.7 % (50 % of females and 24.3 % of males) Cut off ≥6 Prevalence of GJH 16.4 % (27.8 % of females and 10.8 % of males)	Correlative relationship between BS self and hip flexion ($r = 0.61$, $p < 0.01$), internal rotation ($r = 0.56$, $p < 0.01$), and external rotation ($r = 0.44$, $p < 0.01$)
Nicholson and Chan [30]	**BS** **ULHAT**	Convenience sample of 112 adult participants (mean age 24.3 ± 5.5years) divided into known hypermobile cases, likely hypermobile cases and controls (Australia)	Yes	BS: Inter-rater reliability ICC2,1 = 0.92	N/A	BS: The cut-point ≥7/12 (sensitivity 0.84, specificity 0.77, +LR 3.7, -LR 0.2)	Beighton score cut-off of ≥4/9 revealed low to moderate levels of agreement with clinical opinion for the control and likely hypermobile groups (48% and 45% respectively), and high levels of agreement only with the known hypermobile group at 91%. Utilizing a Beighton cut-off score of ≥5/9 improved the moderate levels of agreement with clinical opinion for the control and likely hypermobile groups (58% for both), while maintaining high levels of agreement with the known hypermobile group at 88%. Having identified a cut-off score of ≥5/9 for the identification of GJH, the Beighton scoring system identified GJH in 45% control, 60% likely hypermobile and 94% known hypermobile participants. With the more stringent cut-off of ≥5/9 and higher agreement with clinical opinion, the McNemar's test still revealed a significant difference in the agreement between clinical opinion of generalized hypermobility and the Beighton score for the control and likely hypermobile participants ($p < 0.001$), but not in the known hypermobile group ($p = 1.0$)
Öhman, Westblom and Henrikkson [60]	**BS** **Hospital del Mar**	A total of 128 healthy children 74 females 54 males aged 5-8 years (Sweden)	No	N/A	N/A	Cut off ≥4 BS Prevalence of GJH =12% Hospital Del Mar Criteria = 34%	The Hospital del Mar criteria has been developed over a period of years and the version used in this study is not exactly the same as in other studies. This makes comparisons more complicated.
Patel *et al* [13]	**BS** **Brighton Criteria**	200 patients 108 females 92 males aged between 3 and 15 years (mean 10.1) (UK)	No	N/A	N/A	Mean standard deviation (SD) BS was 2.06 (2.2), and the range was 0–8. Comparing males versus females, mean BS SD was 1.71 (2.25) versus 2.36 (2.14); p=0.0378, age was 9.75 versus 10.13, and BS range was 0–7 versus 0–8.64 children (32%) complained of pain in at least one joint, though the mean SD BS in these patients was 1.71 (1.86)	The "Brighton" score, when combined with BS using the higher diagnostic score of 5/9, could be the more reliable scoring system

Pearsall *et al* [61]	**BS** **KT 2000 arthrometer** **N(Ankle arthrometer)**	57 athletes (29 men; 28 woman; age = 20.9 ± 1.45 yr) without a history of previous injury (USA)	No	N/A	N/A	N/A	Non-significant correlations were observed among the test variables for generalized joint laxity (.21 to .37; p > .05) and instrumented ankle and knee joint laxity (.19 to .21; p > .05). When examined by gender, no statistically significant correlations (.05 to .40; p > .05) were found
Phan *et al* [62]	**LLAS**	57 pre-professional and 29 professional ballet dancers (21±4years, 64% female, mean 13.7years training) were recruited (Australia)	No	ICC2,1= .0.85,95% (CI . 0.67 to 0.94, p < 0.001)	N/A	N/A	The right leg was significantly more hypermobile than the left for the whole cohort (44% vs 40% meeting 7/12 for the LLAS; LLAS mean/12(SD): right:5.0(2.4) and 7.6(1.9); left:4.8(2.1) and 6.7(2.0) in pre-professionals and professionals respectively (p ¼ 0.02)). Subtalar pronation (p < 0.001) and hip abduction/external rotation (left:p¼ 0.01; right p < 0.001) were significantly more hypermobile bilaterally in professionals. Three hypermobility profiles on the left and four on the right lower limb were identified.
Romeo *et al* [63]	**Modified BS**	284 healthy preschool children 146 boys 138 girls 26 preschool children with genetic disorders 15 boys and 11 girls Mean age was 33.6 (Italy)	Yes	For cut off score of 6, kappa = 0.75, Kappa = 0.78	N/A	N/A	Revised version of the Beighton score can be used to define generalized hypermobility for children up to 5 years of age
Riley *et al* [64]	**BS**	51 University students 17 male 34 female Average age 23 (USA)	Yes	Interrater reliability ICC =0.52 for intertester reliability at visit 1 and 0.86 at Visit 2	intratester reliability ICC = 0.88 for Tester 1 and 0.71 for Tester 2	N/A	Intraclass correlation coefficients (ICC2,1) were 0.52 for intertester reliability at visit 1 and 0.86 at visit 2, with intratester reliability of 0.88 for Tester 1 and 0.71 for Tester 2 for the BS. Intertester prevalence-adjusted bias-adjusted kappa (PABAK) values for the Beighton GJH cut-off scores were 0.80 e0.84 and 0.80 to 0.92 for intratester reliability.
Sauers *et al* [65]	**Shoulder hypermobility as a marker for GJH using a shoulder arthrometer**	51 adults 28 females 23 men mean age 22 years (USA)	No	N/A	N/A	N/A	No moderate or stronger correlations between laxity, passive range of motion, and generalized joint laxity were seen The BS scores did not correlate highly with any of the passive range of motion values (range, 0.01- 0.48)
Schlager *et al* [7]	**BS** **Hospital Del Mar** **Contompasis**	30 healthy adults aged 18-65 from a convenience sample of a rehabilitation company within primary care Forty-nine adults, 38 women and 11 men, mean (SD) age 39.8 (13.5) years participated in the inter-raterreliability study. Twenty-nine adults, 23 women and 6 men, mean (SD) age 39.9 (12.5) years participated in the intra-raterreliability study (Sweden)	Yes	BS: Inter-rater reliability 0.72 (95% CI 0.55-0.83) Hospital del mar: Inter-rater reliability 0.81 (95% CI 0.69-0.89) Contompasis: Inter-rater reliability 0.82 (95% CI 0.69-0.89)	BS: Intra-rater reliability 0.76 (95% CI 0.54-0.88) Hospital del mar: Intra-rater reliability 0.86 (95% CI 0.73-0.93) Contompasis: Intra-rater reliability 0.79 (95% CI 0.57-0.90)	N/A	N/A
Schlager *et al* [66]	**BS** **BS-Self Reported**	301 pregnant women chosen from a larger sample of 8029 (Sweden)	No	N/A	N/a	BS: Cut off ≥5, (15.9%) indicating a low post-test probability BS Self Reported: Cut off ≥2 AUC 0.73 (95% CI 0.67–0.79) Sensitivity 84.1% (95% CI 69.9–93.4) specificity, 61.9% (95% CI 55.6–67.8). The false-positive rate was 38%. The PPV and NPV ranged between 13.4 and 42.4% and between 90.3 and 99.2%, respectively. The LR+ increased by 2.2 when using a cut-off level of ≥2 for the self-reported 5PQ	There is uncertainty in identifying generalized joint hypermobility in young women using the BS-self reported with a cut-off level of ≥2 when the Beighton score ≥ 5 is used as the reference test. The strength of the BS-self reported is to rule-out women without generalized joint hypermobility

Singh *et al* [67]	**BS**	1000 healthy participants aged 3 to 101 years were recruited via a convenience sample (Australia)	Yes	N/A	N/A	Cut-off of ≥ 4 Sensitivity of 0.8% Specificity of 99.3% ($p < 0.001$) False positive rate of 60%	To lower the risk of a false-positive diagnosis of GJH, further tests of hypermobility need to be utilized A cut-off of54 was only found to be appropriate for females aged 40_59 years and males aged 8_39 years
Sirajudeen *et al* [68]	**BS**	303 children. 161 girls 142 boys Age range 8-14 (Saudi Arabia)	Yes	N/A	N/A	Cut off ≥ 4 Prevalence of GJH =15.2% Cut off ≥ 6 Prevalence of GJH =7.6%	The prevalence reported in this study among school-aged children was comparable with those reported worldwide
Skiwiot *et al* [69]	**BS** **5pQ** **Sachse's criteria**	77 dancers 19 female 58 male Aged 18-25 years from the Polish dance theatre (Poland)	N/A	N/A	N/A	BS: GJH = 64.9% BS $\chi2(1) = 6.485$; $p = 0.011$ 5pQ: GJH = 74% 5pQ–$\chi2(1) = 11.199$; $p = 0.001$ Sasche: GJH = 59.7% Sachse–$\chi2(1) = 11.206$; $p = 0.001$	N/A
Smits-Engelsman, Klerks and Kirby [70]	**BS**	551 paediatric participants 293 females 258 males aged 6 to 12 years (Holland)	Yes	N/A	N/A	Cut off score >= 5 More than 35% of children scored more than 5/9 on the Beighton score	Authors recommended that 7/9 be the cut-off for the Beighton score.
Steinberg *et al* [71]	**BS** **Y** **Range of Motion including Passive En Pointe, Ankle Plantar Flexion/Ankle Dorsiflexion, Hip ROM**	240 non professional female dancers, aged 8 to 16 years, and 226 female nondancers of similar age (Israel)	Yes	Intraclass correlation ICC range between 0.74and 0.95 and for body measurements between 0.90 and 0.95. Kappa = 0.82	Intraclass correlation ICC range between 0.87 and 0.96 and for body measurements between 0.95 and 0.7. Kappa = 0.86	prevalence of JHM is significantly higher among dancers compared with the control subjects ($P < 0.001$). Joints' ROM is higher among dancers with JHM compared with dancers without JHM ($P < 0.05$)	N/A
Van der Giessen *et al* [72]	**BS** **5PQ**	773 Dutch children 378 girls 395 boys aged 4-12 years old (Holland)	No	N/A	N/A	Cut off ≥ 4 GJH present in 26.5% of participants	N/A
Vallis, Wray and Smith [31]	**BS** **Contompasis**	36 physiotherapy students 9 female 27 males mean age 22.7 years (USA)	No	BS: Inter-rater reliability ICC range 0.72-0.80 Contompasis: Inter-rater reliability ICC range 0.58-0.62	BS: Intra-rater reliability ICC range 0.71-0.82 Contompasis: Intra-rater reliability ICC 0.73-0.82	N/A	BS superior to contompasis

Table 2. Case Control Study. JBI LEVEL 3B.

Author	JBI level of evidence	Assessment Tools	Participant characteristics	Goniometer Used	Inter-rater reliability	Intra-rater reliability	Accuracy	Other statistics
Ballenger, Moore-Clingenpeel and Oberle [73]	**JBI LEVEL 3B**	**BS** **Ultrasound findings**	50 paediatric participants average age 16.1, Female 26 males 24 recruited from the same paediatric rheumatology clinic as well as a paediatric ophthalmology clinic and a paediatric dermatology clinic at an academic children's hospital. Ethics approval was obtained from the Nationwide Children's Institutional Review Board (USA)	Yes	N/A	N/A	N/A	H + P knees were more likely to have positive findings noted on MSUS (94% vs. 70% of H-P and 74% of NP knees, p = 0.043). Patellar tendon hyperemia was more common in H + P knees (52%, vs. 19% among H-P and 23% among NP, p = 0.025). Participants who reported taking scheduled non-steroidal anti-inflammatory drugs (NSAIDs) had an increased risk of synovial effusion (RR = 1.83, 95% CI = 1.07–2.30, p = 0.026) and a trend towards increased risk of a higher synovial effusion/ hypertrophy quantitative score (RR = 1.77, 95% CI = 0.92–3.38, p = 0.086)
Cherni *et al* [74]	**JBI LEVEL 3B**	**BS** **Extensometer**	33 pregnant females over the age of 18, BMI <40 with the control group was included to have a baseline value of the laxity before pregnancy (Canada)	N/A	N/A	N/A	Laxity of the metacarpophalangeal joint increased by 11% from the first to the second trimester of pregnancy and stabilized until delivery. The Beighton score was significantly higher in the second trimester of pregnancy ($p < 0.05$)	Moderate correlation was observed between the results given by the extensometer and the Beighton score in both the cases and the control group at first trimester ($r = 0.60$, $p < 0.05$) but none was found for the two hip and lumbar flexibility tests The chosen clinical tests don't seem appropriate to be used alone in pregnant women
Cypel [75]	**JBI LEVEL 3B**	**Glenohumeral abduction** **BS**	110 cases of known EDS. Mean age 30 years (range,6–69 years) overall, 32.5 years (range, 6–66 years) in the 87 females,and 22 years (range, 6–69 years) in the 23 males 100 healthy controls Mean age 48 years 50 females 50 males (France)	Yes	N/A	N/A	Sensitivity 96.4% Specificity 92.5% glenohumeral hypermobility	Increased gleno-humeral abduction may be sufficient to demonstrate joint hypermobilityand to suggest EDS in patients whose personal and family history is consistent with this diagnosis
Heidbreder *et al* [76]	**JBI LEVEL 3B**	**Skin hyperextensibility** **N/A**	17 patients with classical EDS, five patients with vascular type EDS, 17 patients with sCAD without known connective tissue disease and 29 healthy control individuals 38 EDS patients 22 Females 16 males Age range 6-76 29 healthy controles 13 females 16 males Mean age 33.1 (Germany)	N/A	N/A	N/A	N/A	The method assessed is capable of detecting increased skin extensibility in classical type EDS. Increased skin extensibility is not present in sCAD patients Significance of differences in skin extensibility between groups (Student's t-test) Group 1 Group 2 p-value Controls Vascular type EDS 0.967 Controls Classical type EDS <0.001 Vascular type EDS Classical type EDS <0.001 Controls sCAD 0.967 Controls Classical type EDS 0.131 Vascular type EDS sCAD 0.131 Classical type EDS sCAD <0.001 Abbreviations: EDS, Ehlers-Danlos syndrome; sCAD, spontaneous cervical artery dissection. Significance of differences in joint hypermobility point values between groups (Student's t-test) Group 1 Group 2 p-value Controls Vascular type EDS 0.013 Controls Classical type EDS 0.001 Vascular type EDS Classical type EDS 0.31 Controls sCAD 0.34 Classical type EDS sCAD <0.001 Vascular type EDS sCAD 0.016 Abbreviations: EDS, Ehlers-Danlos syndrome; sCAD, spontaneous cervical artery dissection.
Juul-Kristensen [32]	**JBI LEVEL 3B**	**BS** **Brighton Criteria** Rote`s-Que´rol	Patients aged between 18 and 71 yrs. Cases and controls for phases 1 and 2 were randomly selected from files of patients previously referred to the out-patient clinic of the Department of Medical Orthopaedics and Rehabilitation (Denmark)	No	BS: For the total Beighton score, either currently or historically, ICC was 0.91. For GJH, inter-rater reliability ICC =0.88 kappa = 0.74 Brighton Criteria: Inter-rater reliability ICC = 0.93 kappa = 0.84 Rote`s-Que´rol: Kappa between 0.31 – 0.80 currently, or historically	N/A	N/A	Authors state: Further research on the validity of tests and criteria for GJH and BJHS is urgently needed. In spite of a different cut-off level used for the diagnosis of GJH, the present reproducibility of the criterion (0.74) was at the same level as previously reported (0.78)

McCormack *et al* [77]	**JBI LEVEL 3B**	**BS + Contompasis Score + Brighton Criteria**	149 dance students, 85 from the Lower School and 64 from the Upper School, and 71 professional ballet dancers were recruited from the Royal Ballet School and the Royal Ballet Company, London. 36 pupils from a local secondary school and 31 adults working at The Royal Opera House, London (home of the Royal Ballet) were recruited as controls for the senior student and professional dance cohorts, respectively (UK)	Yes	*Table 2.* Prevalence and odds ratios for hypermobility (Beighton score ≥ 4/9) and BJHS, and the means for Contompasis score and right 5th finger dorsiflexion. Variable: Lower School* / Upper School / Upper School Controls / Company / Company Controls Mean Beighton score (range) Female: 5.6 (1–9) / 7.5 (3–9) / 4.3 (0–9) / 7.1 (3–9) / 3.1 (0–7) Male: 4.3 (1–8) / 6.2 (1–9) / 2.5 (0–9) / 6.4 (3–9) / 3.1 (0–7) Prevalence of hypermobility (Beighton score ≥ 4/9) (%) Female: 55/47 (74) / 33/35 (94) / 13/21 (62) / 41/43 (95) / 8/16 (50) Male: 23/38 (82) / 24/29 (83) / 3/15 (20) / 23/28 (82) / 6/15 (40) Combined OR for hypermobility: — / 11.3 (4.1–31.2), p = 0.001 / — / 11.1 (3.8–31.8), p = 0.3 / — Prevalence of BJHS (%) Female: 22/47 (47) / 16/35 (46) / 4/21 (19) / 11/43 (26) / 2/16 (13) Male: 17/38 (45) / 10/29 (35) / 1/15 (7) / 10/28 (36) / 2/15 (15) Combined OR for BJHS: — / 5.9 (1.5–11.3) p = 0.01 / — / 1.7 (0.6–4.7) p = 0.3 / — Nonparametric t test of Contompasis Score (95% CI) Female: — / 41.5 (39.2–43.8) / 28.7 (25.2–32.2) / 38.2 (36.4–40.1) / 26.1 (22.5–29.6) Mean difference: 12.8 (8.6–16.7) / 12.1 (8.2–16.1) Male: — / 36.0 (33.5–38.6) / 23.7 (19.4–28.0) / 35.7 (33.1–38.4) / 24.3 (21.3–27.3) Mean difference: 12.3 (7.4–17.1) / 11.4 (7.6–15.3) Prevalence of hypermobility (Contompasis score ≥ 26) (%) Female: — / 38 (100) / 12/21 (57) / 43 (100) / 9/16 (56) Male: — / 27/29 (93) / 3/15 (20) / 28 (100) / 6/15 (40) Nonparametric t test of right 5th finger dorsiflexion (95% CI) Female: — / 88 (77–85) / 68 (60–76) / 90 (87–93) / 74 (65–84) Mean difference: 21 (12–29) / 16 (8–24) Male: — / 90 (88–92) / 62 (50–74) / 89 (85–94) / 67 (55–79) Mean difference: 28 (16–40) / 22 (12–33) * Observation only and no controls. Scoring systems not validated for children.	An OR of 11.0 (95% CI 3.3–31.8) was found for hypermobility in dancers for both the ballet school and the professional company Odds ratios for BJHS in student dancers were significant, OR = 3.9 (95% CI 1.3–11.3), but not so in professional dancers: OR = 1.7 (95% CI 0.6–4.7)
Meyer *et al* [78]	**LEVEL 3B**	**LLAS**	112 participants in total 100 females 12 males Age range 17-71 years likely hypermobile group were comprised of elite dancers (working as a dancer full-time) from Western Australian Academy of Performing Arts (Edith Cowan University) and the West Australian Ballet (Perth) Control group: healthy individuals who had not undertaken long term training likely to affect joint mobility such as dance, gymnastics and acrobatics from University of Sydney (Australia)	Yes	Inter-rater reliability ICC 2,1 = 0.85, 95%CI (0.67 -0.94) p < 0.001	N/A	N/A	Beighton score (BS) as assessed by primary care practitioner was found to be higher than assessment by EDS practitioner in 81% (n = 74 of 91) of cases. Generalized joint hypermobility was confirmed in only 46% (n = 51 of 111) of patients who had a previous diagnosis of hEDS
Parveneh and Shiari [79]	**LEVEL 3B**	**Parveneh and Shiari Score**	200 paediatric participants aged 3 to 16. 100 children with BJH compared to age and sex matched controls 42 males in each group 64 females in each group (Iran)	No	N/A	N/A	Sensitivity = 100%, specificity = 98%, positive predictive value= 100%, negative predictive values 98% Accuracy was 99% and balanced accuracy was 99%. The area under the ROC = 0.99	Data analysis revealed significant correlation between Beighton and the new criteria

Table 3:0. Randomised Control Trial. Level 1B.

Author	Assessment Tools	Participant characteristics	Goniometer Used	Inter-rater reliability	Intra-rater reliability	Accuracy	Other statistics
Ahn *et al* [80]	**KRSP** **ROM**	91 healthy participants aged >65 31 males 60 females aged ≥65 years (Korea	Yes	KRSP: Inter-rater reliability ICC 0.846(95% CI 0.686-0.931) to 0.986 (95% CI0.972-0.994) ROM: Inter-rater reliability ICC range 0.643 (95% CI 0.486–0.783) to -0.078 (95% CI -0.296–0.494)	N/A	N/A	These results indicate that the Korean protocol can be the reference standard for measuring ROM in clinical settings as an alternative to goniometers

Table 4. Summary of Systematic and Narrative Review Findings.

Systematic Reviews	**Conclusions/Summaries of Findings (directly from the literature)**
Juul-Kristensen *et al* [12] JBI LEVEL 3A	*'Studies demonstrated limited positive to conflicting evidence with regards to the BS. There is a lack of quality evidence with regards to the other scoring systems assessed, so no reliable conclusions were able to be drawn. More research is required to further evaluate validity and reliability of the BS as well as the other scoring systems, however the authors still recommend: For adults BS should be used with a cut off score of 5, but recommend including historical information. For paediatric patients a BS with cut-point of 6 is recommended.'* *'More research is required before evidence-based recommendations can be developed about which assessment method is the best for identifying GJH.'*
Bockhorn *et al* [81] JBI LEVEL 3A	*'The Beighton score is a highly reliable clinical tool that shows substantial to excellent inter- and intrarater reliability when used by raters of variable backgrounds and experience levels. While individual components of risk of bias among studies demonstrated large discrepancy, most of the items were adequate to very good, despite the range in study designs, participant characteristics and different cut off levels used.'*
Palmer *et al* [82] JBI LEVEL 2A	*'Studies reviewed were of high standard in terms of both methodology and reporting. Three of the four studies included reported at least one measure of tissue mechanics that distinguished between adults with hEDS/HSD and healthy participants. These included assessment of skin, muscle (brachioradialis and biceps brachii) and tendon (patellar and Achilles). Only 4 studies have been conducted on this subject and further validation of methodologies reviewed is required including tissue stiffness meters and ultrasound and other techniques of standardisation.'*
Narrative reviews	**Conclusions Summaries of Findings (directly from the literature)**
Malek *et al* [8] JBI LEVEL 5A	*'A consistent finding in the literature reviewed by these demonstrates that the BS is unable to identify GJH in joints not included in the scoring system, making it an unviable scoring system for identifying GJH and it should not be used as the primary tool in differentiating between L-JH and GJH.'*
Desfor [83] JBI LEVEL 5A	*'Criteria for assessment of JH and BJHS are needed, further research is required. Dance educators and clinicians need more information on the relationship between hypermobility and dance training, strength, proprioception, injury and osteoarthritis.'*
Czaprowski, Kotwicki and Stoliński [84] JBI LEVEL 5A	*'Out of the many methods in use, the most common are the Marshall scale, the Beighton scale and the Carter and Wilkinson scale. What makes the value of the Beighton and Carter & Wilkinson scales higher is that they assess mobility in multiple joints, which reduces the risk of false positive or negative results. The Hakim and Grahame questionnaire seems to be a potential alternative, in accordance with the hands-off principle (no need to conduct a direct clinical examination) promoted by the authors.'*
Remvig *et al* [17] JBI LEVEL 5A	*'Using a 6 cut level, intra- and inter-observer kappa scores were 0.75 and 0.78, respectively. Beighton scoring recommendations have been correlated with a global joint mobility index as well as with 2 other scoring systems, the Carter and Wilkinson, and the Rotès-Quérol. All illustrate high concurrent validity with one another.'* **Table 1** Intra- and inter-examiner reliability for positive hypermobility findings for individual manoeuvres, GJH classification, and agreements with overall score Intra-examiner reliability—adults / Inter-examiner reliability: Adults / Children Individual manoeuvres Hyperextension of the fifth finger—Kappa: 0.78 [47], 1.00 [47] / 0.40 [48], 0.61 [48], 0.70 [47], 0.73 [47], 0.79 [6] / 0.43 [49], 0.48 [49], 0.49 [49], 0.69 [49] Apposition of the thumb to the forearm—Kappa: 0.52 [47], 0.71 [47] / 0.87 [6], 0.91 [47], 0.94 [48], 1.00 [47], 1.00 [48] / 0.82 [50], 0.82 [49], 0.84 [49], 0.89 [49], 0.94 [49] Hyperextension of the elbows—Kappa: 0.61 [47], 0.76 [47] / 0.21 [47], 0.34 [48], 0.50 [47], 0.64 [48], 0.92 [6] / 0.68 [50], 0.54 [49], 0.60 [49], 0.63 [49], 0.73 [49] Hyperextension of the knees—Kappa: 0.19 [47], 0.84 [47] / 0.31 [47], 0.40 [47], 0.85 [48], 0.86 [6], 0.90 [48] / 0.30 [49], 0.43 [49], 0.44 [50], 0.62 [49], 0.62 [49] Forward flexion of the trunk—Kappa: 1.00 [47] / 0.88 [48], 0.93 [6], 0.94 [47] / 0.64 [49], 0.82 [50], 0.84 [49] Classification of GJH (BS ≥ 4)—Kappa: 0.78 [47] / 0.54 [47], 0.78 [48] / – Classification of GJH (BS ≥ 5)—Kappa: 0.47 [47] / 0.66 [47], 0.66 [48] / 0.59 [49], 0.64 [49] Agreement with overall score—Spearman's γ: 0.86 [51] / 0.87 [51] Agreement with overall score—Intraclass Correlation Coefficient (ICC): 0.76 [47], 0.84 [52], 0.92 [52] / 0.72 [47], 0.89 [52], 0.94 [52] *'In one article (Bird, 1979), Beighton tests using 0–9 scoring correlated well with a Global Index: r = 0.81 (p < 0.001). A cut off level for GJH was not discussed.'*

<table>
<tr><td>Drabik, Byś and Gawda [85]

JBI LEVEL 5A</td><td>'In order to diagnose EDS-HT, one should be guided not only by the established standard, which is the Beighton score, but also by additional tests that would confirm the diagnostic decision and reduce the risk of error. Extension of diagnostic tests with additional criteria, which are presented in the article, would make diagnoses more accurate and reduce the possibility of false diagnoses.'</td></tr>
<tr><td>Day, Koutedakis and Wyon [86]

JBI LEVEL 5A</td><td>
<table>
<tr><th>Reference</th><th>Number of subjects</th><th>Population</th><th>Age range (yrs)</th><th>Diagnostic test</th><th>Cutoff Point</th><th>Prevalence %</th></tr>
<tr><td>Seckin et al. [73]</td><td>861</td><td>Turkish high school students</td><td>15.36 ± 1.11</td><td>Beighton score</td><td>≥4/9</td><td>11.7%</td></tr>
<tr><td>Al-Rawi et al. [5]</td><td>1 774</td><td>Iraqi university students</td><td>20–24</td><td>Beighton score</td><td>≥4/9</td><td>30.0%</td></tr>
<tr><td>Beighton et al. [9]</td><td>502</td><td>Tswana (South African) villagers</td><td>20–65+</td><td>Beighton score</td><td>≥3/9</td><td>15.7%</td></tr>
<tr><td>Birrell et al. [10]</td><td>204</td><td>Yoruba (Nigeria) villagers</td><td>6–66</td><td>Beighton score</td><td>≥4/9</td><td>43.0%</td></tr>
<tr><td>Pountain [63]</td><td>309</td><td>Cross-section of Omani people</td><td>16–25</td><td>Beighton score</td><td>≥4/9</td><td>20.0%</td></tr>
<tr><td>Didia et al. [16]</td><td>550</td><td>Nigerian undergraduate university students</td><td>17–30</td><td>Beighton score</td><td>≥5/9</td><td>13.0%</td></tr>
<tr><td>Klemp et al. [49]</td><td>Maori = 438, European = 354</td><td>Cross-section of Maori and European New Zealanders</td><td>30–47</td><td>Beighton score</td><td>≥4/9</td><td>Maori = 6.2% European = 4%</td></tr>
<tr><td>Lamari et al. [54]</td><td>1 120</td><td>Brazilian pre-school children</td><td>4–7</td><td>Beighton score</td><td>≥4/9</td><td>64.6%</td></tr>
<tr><td>Subramanyam et al. [76]</td><td>1 000</td><td>South Indian school children</td><td>6–15</td><td>Beighton score</td><td>≥4/9</td><td>17.2%</td></tr>
<tr><td>Gedalia et al. [24]</td><td>260</td><td>North American school children</td><td>5–17</td><td>Beighton score</td><td>≥3/9</td><td>12.0%</td></tr>
<tr><td>Larsson et al. [55]</td><td>660</td><td>North American music school students and faculty members</td><td>14–68</td><td>Carter & Wilkinson test</td><td>≥3/5</td><td>9.7%</td></tr>
<tr><td>Vougioka et al. [80]</td><td>2 432</td><td>Greek school children</td><td>5–14</td><td>Beighton score</td><td>≥3/9</td><td>8.8%</td></tr>
<tr><td>Forleo et al. [20]</td><td>1 005</td><td>Brazilian school children</td><td>5–17</td><td>Beighton score</td><td>≥5/9</td><td>36.3%</td></tr>
<tr><td>Carter and Wilkinson [13]</td><td>285</td><td>English primary school children</td><td>6–11</td><td>Carter & Wilkinson test</td><td>≥4/5</td><td>7.0%</td></tr>
<tr><td>Rikken-Bultman et al. [68]</td><td>1 568</td><td>Dutch secondary school children</td><td>Beighton for 4–17 Biro for 12 = 17</td><td>Biro & Beighton score</td><td>Biro = ≥3/5, Beighton = ≥4/9</td><td>Beighton = 14.2%, Biro = 12.3%</td></tr>
<tr><td>Kirk et al. [45]</td><td>64</td><td>English rheumatology patients and their relatives and healthy people</td><td>5–57</td><td>Carter & Wilkinson test</td><td>≥3/5</td><td>25%</td></tr>
<tr><td>Kirk et al. [45]</td><td>64</td><td>English rheumatology patients and their relatives and healthy people</td><td>5–57</td><td>Measurement in excess of normal ROM in 8 joints</td><td>≥8/17</td><td>41.0%</td></tr>
<tr><td>Seow et al. [74]</td><td>306</td><td>Singaporean Chinese, Malay and Indian adults</td><td>15–39</td><td>Beighton score</td><td>≥5/9</td><td>17.0%</td></tr>
</table>
'The modest amount of research on injury rates in hypermobile dancers confirm that they have substantially more tendon injuries and longer healing times than normal dancers. Further research is required into this vulnerable group of dancers in order to ensure that there is more understanding of the attrition rate from ballet schools to professional companies and health issues of the hypermobile dancer.'
</td></tr>
</table>

Table 5. Grey Literature Summaries.

Author, date, JBI Level	Type of paper	Summary of information (directly from the literature)
Castori *et al*, [2] JBI LEVEL 3B	Framework	*'Many EDS patients, however, still remain without a laboratory confirmation and this lack of knowledge contributes to the patients' burden. This is mostly the case for individuals affected by the two largely overlapping conditions previously termed "Ehlers–Danlos syndrome, hypermobility type (EDS-HT)" and "joint hypermobility syndrome (JHS)".'* *'The attribution of GJH as a feature remains partly influenced by the examiner's professional experience and recognition of the need to look at all the joints (certainly at least those in the context of the clinical presentation) and not simply those assessed in these tools'*
Grahame [87] JBI LEVEL 5C	Editorial	'*The true diagnosis in a majority of patients with JHS is overlooked in up to 95% of cases.'* Refers to the Remvig [17] review acknowledging a lack of statistical sophistication in the majority of the papers published at the time and the need for further epidemiological studies to establish standardised joint ROM
Grahame [9] JBI LEVEL 5C	Medscape article	*'Useful as an initial screen, the BS can no longer be considered the gold standard for recognizing hypermobility syndrome in clinical or epidemiological practice. It is an arbitrary all-or-none test that does not take into account the degree of laxity and in some individuals the score diminishes with advancing age, perhaps even reaching zero. Moreover, it samples only five sites, although studies have now shown that pauci-articular hypermobility is more common than polyarticular.'*
Kacheria *et al* [88] JBI LEVEL 2B	Research Poster, International Association for Dental Research, AADR, AACR Annual Meeting	*'Prevalence of hypermobility using the 9 point Beighton was 12.4%, while it was 18.4% using the 4 point scale. The sensitivity was 85% and the specificity was 91%. Stratified by gender, the sensitivity and specificity were 76% and 97%, respectively, for males, and 90% and 87%, respectively, for females. The 4 point method can be accomplished in just over a minute, versus 3 minutes for the full Beighton, and the subject does not need to remove clothing or get out of a chair. Conclusions: The shortened Beighton demonstrated good sensitivity and specificity and can be done easily and quickly. It is more convenient and hence may be used as a substitute for the full Beighton scoring examination in research studies.'*
Cincinnati Children's Hospital Medical Center Joint Hypermobility Team [89] JBI LEVEL 5C	Clinical Practice Guidelines . Comment: Comprehensive and holistic guidelines. Incorrectly defines "joint hypermobility" as a primary diagnosis, however written prior to 2017 nosology changes recommended by Castori [2]	*'A Beighton Scale score of 5/9 or greater is used to classify JH (Scheper 2013 [1b], Junge 2013 [2b], Beighton 1998 [5a]). However, 4/9 was also frequently used to classify JH in the literature reviewed for this guideline.'*

Table 6. Summary Statistics: Ranges of Clinimetric Values for GJH Scores (Taken from Table 3.1 and 3.2).

Scoring System	Intra-rater agreement range	Inter-rater agreement range	Validity Statistics range
BS/BHJS	ICC= range between 0.74 [71] – 0.992 [36], kappa = 0.82 [71]	ICC= 0.72 [7] - 0.96-0.98 [45] Spearman rho = 0.86 (p<0.0001) [39]	The Cronbach's a for defining GJH, EDS-HT and JHS was 0.61, 0.79, and 0.44, respectively [5]
Modified BS	No data	Interrater reliability of kappa = 0.49-0.94 Method A, Kappa= 0.30-0.84 Method B [51]	No data
Carter and Wilkinson Scoring System	No data	No data	Cut off score ≥ 3 / 5 GJH=25% [86]
Sasche Score	No data	No data	GJH = 56.5% Sachse–$\chi2(1) = 11.206$; p = 0.001 [90]
Brighton Critera	No data	Inter-rater reliability ICC = 0.93 kappa = 0.84 [32]	For paediatric patients: GJH present in (28.3%) Of these 58.8% were female. Of the 34, 28 (82.4%) met both major criteria one of which is BS >4. [38]
5pQ	No data	ICC = 0.92 (95% CI 0.85, 0.96) [47]	Sensitivity = 70.9% specificity = 77.4% ROC= 0.786 [57] Sensitivity =84%Specificity 89% HSD: Sensitivity=80%Specificity = 84% [48] presence of GJH (rho = 0.857; p<0.001) [40]
Line Drawings	No data	k= 1.00; (95% CI 1.00-1.00) [43]	Self-reported participant-repeatability k =0.91 (95% CI 0.74-1.00) Sensitivity of 0.87 (95% CI 0.81, 0.91) Specificity of 0.99 (95% CI 0.98, 1.00) [43]
Screening Questionaire to Detect Joint Hypermobility	No data	ICC= 0.961 CI (0.922-0.980) [40]	GJH 44.3% PPV 92.79% NPV 46.22% [40]
BS – self reported	ICC = 0.92 (95% CI 0.85, 0.96) [47]	No data	Patients with FAI [59]: Correlative relationship between BS-self and hip flexion (r = 0.61, p < 0.01), internal rotation (r = 0.56, p < 0.01), and external rotation (r = 0.44, p < 0.01) In pregnant women [66] There is uncertainty in identifying generalized joint hypermobility in young women using the BS-self reported with a cut-off level of ≥2 when the Beighton score ≥ 5 is used as the reference test. The strength of the BS-self reported is to rule-out women without generalized joint hypermobility Cut off ≥2 AUC 0.73 (95% CI 0.67–0.79) Sensitivity 84.1% (95% CI 69.9–93.4) Specificity 61.9% (95% CI 55.6–67.8) The false-positive rate was 38% The PPV and NPV ranged between 13.4 and 42.4% and between 90.3 and 99.2%, respectively The LR+ increased by 2.2 when using a cut-off level of ≥2 for the self-reported 5PQ Paediatric population [47]: Sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72) 80%)ROC = 0.87 (95% CI 0.79, 0.95)
Marshal Test	No data	No data	PPV = 34%

			NPV = 99% Sensitivity = 99% Specificity = 67% Positive likelihood ratio = 3.3 [91]
KRSP	No data	ICC 0.846- 0.986 [80]	No data
Upper Limb Hypermobility Assessment Tool	No data	ICC2,1 = 0.92 [30]	The cut-point ≥7/12 (sensitivity 0.84, specificity 0.77, +LR 3.7, -LR 0.2) [30]
Lower limb Hypermobility Assessment Score	Spearman rho = 0.86 (p<0.0001) [39]	Spearman rho = 0.87 (p<0.0001) [39]	Sensitivity = 67% Specificity =94% Positive Predictive Value (50%) Negative Predictive Value (97%) [39]
Hospital Del Mar	ICC=0.86 (95% CI 0.73-0.93) [7]	Kappa = 0.81 [92] ICC = 0.81 (95% CI 0.69-0.89 [7]	For presence of GJH (rho = 0.745; p<0.001) [7]
Contompasis	ICC = 0.73-0.82 [31] ICC = 0.79 (95% CI 0.57-0.90) [7]	ICC = 0.58-0.62 [31] ICC = 0.82 (95% CI 0.69-0.89) [7]	No data
Rotes-Querol	No data	Kappa = 0.31 – 0.80 [32] Kappa > 0.6 [17]	No data
Parvaneh-Shiari Criteria	No data	No data	Sensitivity = 100%, specificity = 98%, positive predictive value= 100%, negative predictive values 98% Accuracy was 99% and balanced accuracy was 99%. The area under the ROC = 0.99 [79]

*Interpretation of Agreement Scores

ICC [28] less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively.

Kappa [93] Minimally acceptable kappa score 0.8 or greater.

Spearman Ro [94] 1 = complete or perfect correlation, 0=no correlation

Crohnbach alpha [95] >0.7 is acceptable

Chi squared If the value is greater than the significance level, then the null hypothesis is rejected

Table 7. Summary of Patient Cohort Information from Supplementary Material E.

Patient Cohort	Scoring Systems Assessed	Summary
Pregnant Women	**BS, 5PQ, BS-Self Reported, Extensometer**	• BS-Self Reported good rule out test for GJH Cut off ≥2 AUC 0.73 (95% CI 0.67–0.79) Sensitivity 84.1% (95% CI 69.9–93.4) specificity, 61.9% (95% CI 55.6–67.8) The false-positive rate was 38%. The PPV and NPV ranged between 13.4 and 42.4% and between 90.3 and 99.2%, respectively. The LR+ increased by 2.2 when using a cut-off level of ≥2 for the self-reported 5PQ • Relationship between increased score in 5PQ and pelvic girdle pain during pregnancy • BS not good for identifying BS later in pregnancy due to inability to conduct hand to floor test
Healthy Adults	**BS, Modified BS, KSRP, SQCH, 5PQ, Hospital Del Mar, Self-Reported Line Drawings, CSES, ULHAT, Shoulder Hypermobility, Hospital Del Mar, Contompasis, Rote`s-Que´rol, Brighton Criteria, FAI**	• The mean Corrected Extensibility Score was 23.84%/mm in participants found to be hypermobile versus 13.55%/mm in the normal mobility group • ULHAT has The cut-point ≥7/12 (sensitivity 0.84, specificity 0.77, +LR 3.7, -LR 0.2), Inter-rater reliability ICC2,1 = 0.92 • No validity statistics for Rotes Querol, Kappa for interrater reliability varied between 0.31 and 0.81 • Brighton Criteria Inter-rater reliability ICC = 0.93, kappa = 0.84 • Patients with FAI have a prevalence of GJH = 32.7 % (50 % of females and 24.3 % of males) if using BS cut off Cut off ≥4 • BS Inter-rater reliability >0.7 in all studies except Riley *et al* [64] who report Interrater reliability ICC =0.52 for intertester reliability at visit 1 and 0.86 at Visit 2. • BS Intra-rater reliability >0.7 in all studies
Athletes and Dancers	**BS, LLAS, Contompasis + BS + Brighton Criteria, KT arthrometer, Sasche's Criteria, Villefranche,**	• BS Inter-rater reliability ICC>0.7 in all studies, Intra-rater reliability>087, but only 2 studies looked at this statistic • GJH>60% using BS cut off ≥4 in all studies assessing prevalence • LLAS demonstrated GJH= 38% with cut off score ≥ 7/12 • The Villefranche Criteria identified more dancers with JHS/EDS-HT than the Brighton (84% vs 31%, p < 0.001)
Paediatrics	**BS, Modified BS, Hospital Del Mar, Marshall Test, LLAS, BS Self Reported in Swedish, Sasche Scale, 5PQ, Brighton Criteria, Parveneer and Shiari Score**	• Hospital Del Mar GJH ranges between 38% and 70%, Inter-rater reliability Kappa = 0.81PPV = 34% NPV = 99% Sensitivity = 99% Specificity = 67% Positive likelihood ratio = 3.3 • LLAS Inter-rater reliability ICC = 0.78 , Intra-rater reliability ICC = 0.90-0.98 • BS self reported sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72) 80%)ROC = 0.87 (95% CI 0.79, 0.95) • Sasche Scale Cut off ≥ 7/13 (30% positive) • BS cut off Cut off ≥ 4 Prevalence of GJH ranged from 12% to 64.6% • BS Cut off ≥ 6 GJH ranged from 7.6% to 18.6% • Parveneh and Shiari Score Sensitivity = 100%, specificity = 98%, positive predictive value= 100%, negative predictive values 98% Accuracy was 99% and balanced accuracy was 99%. The area under the ROC = 0.99 • Brighton Score Prevalence of GJH in sample = 0.98% (in healthy sample) vs 28.3% in secondary paediatric care sample • BS Self-Reported (in Swedish) Intra-rater reliability ICC = 0.92 (95% CI 0.85, 0.96) , sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72-80%)ROC = 0.87 (95% CI 0.79, 0.95) • Marshall Test PPV = 34% NPV = 99% Sensitivity = 99% Specificity = 67% Positive likelihood ratio = 3.3

EDS/HSD Patients	**Glenohumeral abduction + BS, Retraction of tendons, BS**	• 87.5% of sample had retraction of hamstrings • Retractions of the triceps surae were found in 90.9% of patients, and retractions of the soles of feet were observed in 95.9% of patients • The impact of retraction on the Beighton palms-on-the-floor test is very great indeed 97.8% of patients who present a retraction of the hamstrings of over 45° cannot perform this manoeuvre, which will affect the BS and potentially identification of the condition in this cohort • Skin hypermobility statistically significant for classical EDS patients • Only 46% HSD/LJH patients scored a BS of 4/9 as assessed by an EDS practitioner (SD = 2.12). In the H-EDS group the average BS was 6/9 (SD = 1.45).

Parvaneh and Shiari [79] report a perfect sensitivity of 100% for their score which is essentially a modified BS. There are no reliability statistics available for this scoring system.

Bulbena *et al* [40] found the self-reported questionnaire positively correlates with both the Hospital del Mar criteria (rho = 0. 745; p<0.001) and the self-reporting questionnaire of Hakim and Grahame [48] (rho = 0.857; p<0.001).

Schlager *et al* [66] found the Hospital del Mar Criteria superior to the BS with regards to both inter-rater and intra-rater reliability statistics (Inter-rater reliability 0.81 (95% CI 0.69-0.89) and Intra-rater reliability 0.86 (95% CI 0.73-0.93) versus Intra-rater reliability 0.76 (95% CI 0.54-0.88) and Intra-rater reliability 0.76 (95% CI 0.54-0.88) respectively. Using a cut off score of ≥4 Öhman, Westblom and Henrikkson [60] found the Hospital Del Mar Criteria superior in its ability to detect GJH at 34% versus 12% for the BS.

No reliability data was available for the Sasche Scale. Skwiot *et al* [90] reported an ability to detect GJH in 56.5% of participants with a $\chi2(1) = 11.206$; $p = 0.001$ (no cut off score mentioned), however Mikołajczyk *et al* [56] reported the Sasche Scale's ability to detect GJH as significantly lower at 30% using a cut off score of ≥ 7/13. This large variance is likely to be due to differences in patient cohorts.

The Rotes-Querol scoring system is the only test that provides non dichotomous scores for hypermobility.

Calhil *et al* [41] found the Marshall test a good rule out test with specificity of 99%, but low sensitivity of 67%.

The BS is brought into question for identifying GJH in the H-EDS patient group by McGillis *et al* [55]. This is the principle method for screening H-EDS when combined with the 5pQ as per the EDS (2017) [6] diagnostic criteria. Joint hypermobility is regarded as the hallmark feature of H-EDS and other EDS subtypes, however authors McGillis *et al* [55], Hamonet and Brock [96] and Malek *et al* [8] report many patients with EDS do not always score highly on the BS.

A reason not identified in current literature is the fact EDS is a disorder of collagen fragility (hence it's systemic nature), however hypermobility has traditionally been utilised as a proxy measure of tissue fragility. As discussed in the introduction, joint hypermobility, whilst related to tissue fragility, is not one in the same phenomenon.

This raises a question regarding nosology of HSD and whether this is truly a condition of hypermobility, or whether it is a disorder of tissue fragility and whether "Connective Tissue Fragility Spectrum Disorders", or similar term is worth considering as alternative nomenclature. It also raises the question in patients who present with clear features of systemic tissue fragility without joint hypermobility should be considered part of this spectrum. Currently such patients do not fall into either the HSD, or EDS nomenclature, but clearly present with connective tissue pathology.

An additional question arises about whether it is possible for some patients to have tissue fragility in the absence of GJH. Currently it is not possible to identify such patients who might be currently misclassified as HSD, or not classified at all.

Degrees of consensus around reliability and validity statistics were reported in systematic and narrative reviews with conflicting conclusions about the usefulness of the BS in detecting GJH.

The majority of reviews including Remvig *et al* [17], Juul-Kristensen *et al* [12], Malek *et al* [8] and Bockhorn *et al* [81] included a high number of association studies. The author determined many of these are not strictly related to clinimetric properties of the scoring systems analysed and therefore should not be included in a review primarily focused on reliability and validity statistics.

Remvig *et al* [17] conclude all scores are comparable, however report the Rotes-Querol has a lower Kappa value compared with other scoring systems assessed.

Malek *et al* [8] conclude use of the BS as a clinical diagnostic tool, particularly within the 2017 International Classification of EDS for the diagnosis of hypermobile EDS (hEDS), remains controversial.

Juul-Kristensen *et al* [12] conclude evidence supports the BS as a reliable clinical tool, however are concerned with a lack of research relating to validity. There is insufficient data to draw conclusions about the other scoring methods assessed in their review.

Drabik, Byś and Gawda [85] report GJH and H-EDS might be more accurately identified when combining the BS with other assessments.

Day, Koutedakis, and Wyon [86] reviewed hypermobility in dance including papers relating to dancers and non-dancers with a difference in detection rates between the Carter and Wilkinson method vs the Beighton Score, however these studies included different types of participants it is difficult to ascertain whether detection rates differ as a result of participant characteristics, or difference in the ability of scores to detect GJH.

Bockhorn *et al* [81] report the BS is excellent in terms of reliability. This review states publication bias and the lack of standardised reporting with use of either composite scores, individual measurements, or variable cut off scores could lead researchers to be influenced to choose the value with the highest level of agreement.

Lack of research to establish validity of the BS raised by Remvig *et al* [17], Malek *et al* [8] and Juul-Kristensen *et al* [12] should be addressed as a priority area in future research.

Most reviews agreed there is lack of high-quality studies, with the exception of the paper by Palmer *et al* [82] a different patient cohort from most papers (known EDS/HSD), hence only four high quality studies were included in their review.

Few papers are high quality. Methodological quality issues include failure to use, or report use of gold standard goniometry devices for assessment of ROM, non-representative patient samples, convenience sampling, non-blinding of participants and assessors, lack of appropriate statistical analysis and, or reporting and failure to compare diagnostic studies to the recommended best practice guidelines. Small sample size is potentially an issue in the studies. Many papers used sample sizes of less than 100, the average sample size ranged from 100-300. 6 studies included a sample size greater than 1000. Some studies did not discuss what cut-off score was used, whilst others did not explain the rationale of cut-off score choice.

In establishing diagnoses, clinicians should rely on evidence-based diagnostic methodology that is standardised according to the COSMIN 2010 [24] criteria, satisfying 3 domains of validity, reliability and responsiveness (the test's ability to detect change over time) [97]. None of the studies reviewed looks at all three of these domains.

Few papers referenced appropriate reporting, or test design standards such as STARD [98] or QUADAS-2 [99]. Some studies were conducted prior to authorship of the COSMIN [24] guidelines, however there is a general lack of referencing to other appropriate standards. Riley *et al* [64] reference the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).

As no standardised reporting of clinimetric statistics exists, comparison is challenging.

Only one randomised control trial (RCT) was identified; Ahn *et al* [80], however this only partially adhered to standards for conducting a RCT.

Specific patient characteristics possibly impacted results of the BS ability to detect GJH. For example, in athlete and dance cohorts. The results and analysis of a study by Riley *et al* [64] noted prevalence of GJH within a sample has an impact on the kappa values of this tool. As reported by Lijmer *et al* [100] diagnostic studies with methodological flaws possibly over-estimate the accuracy of a test and falsely increase the pre-test probability.

A challenge in assessment of validity and reliability studies with regards to the BS, is the fact that most prevalence studies use the BS score itself to establish the presence of GJH. This introduces an inherent level of bias as the scoring method is essentially tested against itself.

This bias is likely to overestimate the sensitivity and specificity values, because these have been based on prevalence calculated using the same test. This is a source of major bias and flaw in study methodology in validity assessment of the BS. This is discussed as a source of bias in validity studies of diagnostic test accuracy by Worster and Carpenter [101] and Kea, Hall and Wang [102]. Incorporation bias would also impact the cut-off values for tests in different cohorts and is an area that requires serious review.

Paucity of literature exists on whether a combination of tests would improve the validity of scoring systems. It is possible the BS in conjunction with other methods yields a higher sensitivity compared with a single scoring system [85,30,54].

5.1 Heterogeneity of Results

High heterogeneity as well as conflicting conclusions drawn by researchers is present.

The complexity of subject material and lack of standardisation of procedures and protocols in clinical practice and current literature are some reasons for heterogeneity.

Other causes of heterogeneity include:

- High variability in study design methodology (case-control, vs cohort study, vs trial vs review)
- Comparison of different range of scores (not all studies compared the same scores, not all studies compared the scores with the gold standard, or included a full clinical assessment)
- Use of different cut off score (most papers use between 4-5 in adult populations and 6-7 in paediatric papers)
- Failure to use goniometer
- Lack of standardised reporting methods
- Lack of standardised statistical parameters for validity and reliability (kappa vs cronbach alpha, vs ICC)
- No standardised protocol for performing BS
- Wide differences in participant characteristics (children, vs pregnant women, vs adults, vs dancers, vs elite athletes)
- Wide differences in study size

5.2 Limitations and Bias in this Review

The single reviewer created possible selection bias and lack of use of systematic review software are the two most important limitations of this review. Other limitations include the search strategy, that did not use meta search terms and did not use the term "psychometric properties" when conducting the search.

The variety of non-standardised terminology potentially poses an issue, however no critical studies have been missed from the review that would alter findings due to a sufficiently broad search criteria and thorough manual search.

Other sources of bias and limitations of this paper include:

- Manual download of papers creates possible reproducibility issues
- Methodological quality of papers used in the review
- Recommended software for conducting a systematic review was not used

- Limitations of inclusion and exclusion criteria
- Broad inclusion criteria
- Variability of reported results
- Limited review of grey literature
- Limited review clinical practice guidelines
- Inclusion of some low-quality papers, grey literature and narrative reports
- English only publications
- Limited to 4 databases
- Only free papers were accessed

Areas for future research should focus on the following areas:

- Standardisation of protocols
- Evaluation between tip of thumb vs more of thumb approximation to forearm
- Consensus on cut off scores
- Lack of individual studies assessing BS specifically for COSMIN [24] criteria for validation
- Explore alternative scores for patients with tight hamstrings -> hand to floor measurement
- Explore alternative scores for patients with ligament laxity as opposed to joint hypermobility
- Whether combining BS with other scoring systems provides increased reliability and validity of findings
- The need to establish a consensus-based gold standard test (or combination of tests) for recognition of GJH
- Investigation of a standardised "global index of joint hypermobility"
- Appropriate study designs to further investigate and compare validity, reproducibility and reliability properties of all widely used scoring systems including questionnaires as these potentially have numerous benefits in a clinical context for the assessment of GJH, including reduced time spent assessing a patient without an examination and the ability to be conducted prior to appointments
- More studies on the Rotes Querol scoring system that measures degrees of hypermobility, rather than the currently dichotomous scoring systems
- A study to determine appropriate statistically derived cut-off scores for the general population vs subgroups such as professional dancers and athletes where a high pre-test probability is likely to affect the scoring systems utilised and affect interpretation of validity and reliability
- A study designed to assess joint ROM in patients with comorbid arthritis with adjusted scoring systems
- Consensus on ROM measurements for hypermobility
- Research into combined assessment methods for establishing a diagnosis of GJH vs single scoring systems
- Systematic review of instruments, software and devices for measurement of joint ROM
- Systematic review of novel associations between various clinical signs and GJH Standardised reporting methods in accordance with the many established guidelines on assessing diagnostic tests, such as COSMIN [24].

CONCLUSION

The Beighton Score is the most commonly used score for identify GJH. This review compared reliability and validity statistics of BS to other scores for identifying BS.

The findings of this systematised review support those of previous similar reviews conducted by Juul-Kristensen *et al* [12], Bockhorn *et al* [81], and Malek *et al* [8] and the systematic and narrative reviews included in this study.

Lack of consensus on terminology and lack of precision of terms creates nosological and clinical dilemmas for researchers, clinicians and patients. In medical literature it creates difficulty identifying relevant literature. As per Castori *et al* [2] the author recommends future research use the term Generalised Joint Hypermobility (GJH) and Hypermobility Spectrum Disorders (HSD) and abandon other approximating terms.

Additionally, imprecise use of terms creates risk that hypermobility is used as a proxy term for tissue fragility, (the true marker of Ehlers Danlos Syndrome and disorders of defective collagen and connective tissue) will potentially limit diagnosis. This is an area that requires exploration in future research.

The Beighton Score has acceptable inter-rater reliability and intra-rater reliability ranges reported in the literature with the lowest ICC reported as 0.72 and 0.74 (respectively). Questions still remain regarding its validity.

Incorporation bias is a major issue of concern not previously addressed in literature to date. As such, The BS should not be used to rule out a diagnosis of GJH. In patients where there is a high level of clinical suspicion who score <4 on the BS, further clinical evaluation is warranted to ensure diagnoses are not missed as delays, or missed diagnoses can have serious negative health outcomes for patients.

There is limited high quality research available hence it is not possible to draw a conclusion about whether other tools are more or less reliable compared to BS, or valid in comparison to the BS.

No large multicentre, blinded, randomised control trial has been conducted comparing the BS to other commonly used methods, or the alternative of combination scores, or global joint ROM index.

Further research is required to quantify validity and reliability of the BS and other scoring systems. Although recommended by systematic and narrative reviews in the past, there remains paucity of literature. This should be a priority as it creates significant challenges for patients who rely on scoring systems to achieve a diagnosis of HSD/EDS and other HDCT involving GJH.

Table 4. Summary of Findings.

• The Beighton Score has demonstrated good intra-rater and inter-rater reliability
• A lack of rigorous study design and methodological flaws prevents an evidence-based conclusion regarding the Beighton Score's validity in assessment of GJH
• The literature is inadequate for determining whether the Beighton Score is Superior to other commonly used scoring systems in clinical use in detecting GJH
• As there is a lack of literature on the validity of the BS, it should not be used to rule out a diagnosis of GJH. In patients where there is a high level of clinical suspicion who score <4 on the BS, additional referral +/- investigations for HSD should be carried out
• There is need for improvement of methodological quality in studies designs to assess clinimetric properties of scoring systems
• Future study designs eliminating incorporation and verification bias for the BS are required
• Novel scoring systems require further evaluation to assess their clinimetric properties and usefulness in clinical practice
• Cut off scores should be chosen with care according to the characteristics of the patient cohort being assessed
• Patients who achieve a low value on any of the scoring systems discussed in this review, should still be assessed for clinical signs of collagen disorders and if these exist, referred or investigated in greater detail to avoid missing the diagnosis of EDS and other collagenopathies
• Further research is urgently required in this important area of musculoskeletal medicine with areas of focus recommended in Table 4.9

REGISTRATION AND PROTOCOL

This paper was not registered as it formed part of a thesis for the University of South Wales, UK.

SUPPORT AND CONFLICTS OF INTEREST

The author received no financial, or other type of support in undertaking this review. The author has no conflict of interest to declare.

AVAILABILITY OF DATA AND OTHER MATERIALS

Supplementary material, and additional information can be obtained by contacting the author at the email address provided.

LIST OF ABBREVIATIONS

GJH – Generalised Joint Hypermobility
JH – Joint hypermobility
HSD – Hypermobile Spectrum Disorder
H-EDS – Hypermobile Ehlers Danlos Syndrome
BS – Beighton Score
ROM – Range of motion
HDTC – Hereditable Disorders of Connective Tissue

CONSENT FOR PUBLICATION

Not relevant for this review as it doesn't contain individual data.

FUNDING

No funding was received in the process of this review.

CONFLICT OF INTEREST

The author has no conflict of interest to declare.

ACKNOWLEDGEMENTS

Thank you to Dr Rizwan Rajak from the University of South Wales for his supervision and support during the undertaking of this review.

REFERENCES

[**1**] Kumar B, Lenert P. Joint Hypermobility Syndrome: Recognizing a Commonly Overlooked Cause of Chronic Pain. Am J Med. 2017;130(6): 640–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002-9343(17)30220-6 [cited: 11th Dec 2021]

[**2**] Castori M, Tinkle B, Levy H, Grahame R, Malfait F, Hakim A. A framework for the classification of joint hypermobility and related conditions. Am J Med Genet C Semin Med Genet. 2017;175: 148–57. Available from: https://doi.org/10.1002/ajmg.c.31539 [cited: 11th Dec 2021]

[**3**] Tofts LJ, Elliott EJ, Munns C, Pacey V, Sillence DO. The differential diagnosis of children with joint hypermobility: a review of the literature. Pediatr Rheumatol Online J. 2009;7(1): 2-10. Available from: https://doi.org/10.1186/1546-0096-7-1 [cited: 11th Dec 2021]

[**4**] Reuter PR, Fichthorn KR. Prevalence of generalized joint hypermobility, musculoskeletal injuries, and chronic musculoskeletal pain among American university students. PeerJ. 2019;7(12): e7625. Available from: https://doi.org/10.7717/peerj.7625 [cited: 11th Dec 2021]

[**5**] Remvig L, Flycht L, Christensen KB, Juul-Kristensen B. Lack of consensus on tests and criteria for generalized joint hypermobility, Ehlers-Danlos syndrome: Hypermobile type and joint hypermobility syndrome. Am J Med Genet A. 2014;164(3): 591–6. Available from: https://pubmed.ncbi.nlm.nih.gov/24464988/ [cited: 11th Dec 2021]

[**6**] The Ehlers Danlos Society. EDS Diagnostics 2017. 2017; [cited: 11 December 2021]. Available from: https://www.ehlers-danlos.com/eds-diagnostics/.

[**7**] Schlager A, Ahlqvist K, Rasmussen-Barr E, *et al.* Inter- and intra-rater reliability for measurement of range of motion in joints included in three hypermobility assessment methods. BMC Musculoskeletal Disord. 2018;19(1): 376-86. Available from: https://pubmed.ncbi.nlm.nih.gov/30326873/ [cited: 12th Dec 2021]

[**8**] Malek S, Reinhold EJ, Pearce GS. The Beighton Score as a measure of generalised joint hypermobility. Rheumatol Int. 2021;41(10): 1707–16. Available from: https://doi.org/10.1007/s00296-021-04832-4 [cited: 11th Dec 2021]

[**9**] Grahame R. Joint hypermobility syndrome: a complex constellation of symptoms. Medscape Medical News. 2004 Sept 30; [cited: 11 December 2021]. Available from: https://www.medscape.com/viewarticle/537938.

[**10**] Tinkle BT. Symptomatic joint hypermobility. Best Pract Res Clin Rheumatol. 2020;34(3): 1021-33. Available from: https://doi.org/10.1016/j.berh.2020.101508 [cited: 11th Dec 2021]

[**11**] Beighton P, Solomon L, Soskolne CL. Articular mobility in an African population. Ann Rheum Dis. 1973;32(5): 413–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1006136/ [cited: 11th Dec 2021]

[**12**] Juul-Kristensen B, Schmedling K, Rombaut L, Lund H, Engelbert RHH. Measurement properties of clinical assessment methods for classifying generalized joint hypermobility-A systematic review. Am J Med Genet C Semin Med Genet. 2017;175(1): 116–47. Available from: https://doi.org/10.1002/ajmg.c.31540 [cited: 12th Dec 2021]

[**13**] Patel N, Riddlestone P, Gunalan R, Gough A, Metcalfe J, Holroyd B. Reliability of Beighton's Score in a Pediatric Population. Int J Pediatr Orthop. 2017;3(1): 9–11. Available from: http://ijpoonline.com/reliability-of-beightons-score-in-a-pediatric-population/ [cited: 22th June 2021]

[**14**] Ferrari J, Parslow C, Lim E, Hayward A. Joint hypermobility: The use of a new assessment tool to measure lower limb hypermobility. Clin Exp Rheumatol. 2005;23(3): 413–420. Available from: https://www.clinexprheumatol.org/article.asp?a=2612 [cited: 12th Dec 2021]

[**15**] Soucie JM, Wang C, Forsyth A, *et al.* Range of motion measurements: reference values and a database for comparison studies. Haemophilia. 2011;17(3): 500–7. Available from: https://doi.org/10.1111/j.1365-2516.2010.02399.x [cited: 11th Dec 2021]

[**16**] Moromizato K, Kimura R, Fukase H, Yamaguchi K, Ishida H. Whole-body patterns of the range of joint motion in young adults: masculine type and feminine type. J Physiol Anthropol. 2016; 35(23): 1-12. Available from: https://doi.org/10.1186/s40101-016-0112-8 [cited: 11th Dec 2021]

[**17**] Remvig L, Jensen DV, Ward RC. Epidemiology of general joint hypermobility and basis for the proposed criteria for benign joint hypermobility syndrome: review of the literature. J Rheumatol. 2007;34(4): 804–9. Available from: https://pubmed.ncbi.nlm.nih.gov/17407233/ [cited: 14th Dec 2021]

[**18**] Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). PRISMA 2020 Checklist. 2020 Month Day; [cited: day month year]. Available from: http://prisma-statement.org/

[**19**] CASP. Critical Appraisal Skills Programme 2019 CASP Diagnostic Study Checklist 2019; [cited: day month year]. Available from: https://casp-uk.net/casp-tools-checklists/.

[**20**] Joanna Briggs Institute. JBI Levels of Evidence 2013; [cited:]. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://jbi.global/sites/default/files/2019-05/JBI-Levels-of-evidence_2014_0.pdf.

[**21**] Pratt AL, Ball C. What are we measuring? A critique of range of motion methods currently in use for Dupuytren's disease and recommendations for practice. BMC Musculoskeletal Disord. 2016;17: 20-31. Available from: https://doi.org/10.1186/s12891-016-0884-3 [cited: 11th Dec 2021]

[**22**] Shamsi M, Mirzaei M, Khabiri SS. Universal goniometer and electro-goniometer intra-examiner reliability in measuring the knee range of motion during active knee extension test in patients with chronic low back pain with short hamstring muscle. BMC Sports Sci Med Rehabil. 2019;11(4):1-5. Available from: https://doi.org/10.1186/s13102-019-0116-x [cited: 11th Dec 2021]

[**23**] Mejia-Hernandez K, Chang A, Eardley-Harris N, Jaarsma R, Gill TK, McLean JM. Smartphone applications for the evaluation of pathologic shoulder range of motion and shoulder scores—a comparative study. JSES Open Access. 2018;2(1): 109–14. Available from: https://doi.org/10.1016/j.jses.2017.10.001 [cited: 11th Dec 2021]

[**24**] COSMIN. Checklists for Assessing Study Qualities 2010; [cited: day month year]. Available from: https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/

[**25**] Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;371(71): 1-9. Available from: https://doi.org/10.1136/bmj.n71 [cited: 11th Dec 2021]

[**26**] Rethlefsen ML, Kirtley S, Waffenschmidt S, *et al.* PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10(1): 1-19. Available from: https://doi.org/10.1186/s13643-020-01542-z [cited: 19th Dec 2021]

[**27**] University of Canberra. Evidence-Based Practice in Health. nd; [cited: day month year]. Available from: https://canberra.libguides.com/c.php?g=599346&p=4149721

[**28**] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2): 155–63. Available from: https://doi.org/10.1016/j.jcm.2016.02.012 [cited: 16th Dec 2021]

[**29**] Johnson AP, Ward S, Simmonds J. The Lower Limb Assessment Score: A valid measure of hypermobility in elite football? Phys Ther Sport. 2019;37: 86–90. Available from: https://pubmed.ncbi.nlm.nih.gov/30904749/ [cited: 12th Dec 2021]

[**30**] Nicholson LL, Chan C. The Upper Limb Hypermobility Assessment Tool: A novel validated measure of adult joint mobility. Musculoskelet Sci Pract. 2018;(35): 38-45. Available from: [cited:]

[**31**] Vallis A, Wray A, Smith T. Inter- and intra-rater reliability of the Beighton Score compared to the Contompasis Score to assess generalised joint hypermobility. Int J Ther Rehabil. 2014;23(1-2): 21-27. Available from: https://www.magonlinelibrary.com/doi/abs/10.12968/ijtr.2014.21.Sup7.S8a [cited: 12th Dec 2021]

[**32**] Juul-Kristensen B, Rogind H, Jensen DV, Remvig L. Inter-examiner reproducibility of tests and criteria for generalized joint hypermobility and benign joint hypermobility syndrome. Rheumatology (Oxford). 2007;46(12): 1835–41. Available from: https://pubmed.ncbi.nlm.nih.gov/18006569/ [cited: 12th Dec 2021]

[**33**] Ahlqvist K, Bjelland EK, Pingel R, Schlager A, Nilsson-Wikmar L,

Kristiansson P. The Association of Self-Reported Generalized Joint Hypermobility with pelvic girdle pain during pregnancy: a retrospective cohort study. BMC Musculoskelet Disord. 2020;21(1): 474-84. Available from: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-03486-w [cited: 11th Dec 2021]

[**34**] Antonio DH, Magalhaes CS. Survey on joint hypermobility in university students aged 18–25 years old. Adv Rheumatol. 2018;58(3): 1-7. Available from: https://advancesinrheumatology.biomedcentral.com/articles/10.1186/s42358-018-0008-x#citeas [cited: 12th Dec 2021]

[**35**] Armstrong R. Relative joint contribution to joint hypermobility in rugby players, netballers and dancers: the need for careful consideration of lumbar flexion. Int J Sports Phys Ther. 2018;13(4): 676–86. Available from: https://doi.org/10.26603/ijspt20180676 [cited: 11th Dec 2021]

[**36**] Armstrong R, Greig DM. The Beighton score as a predictor of Brighton criteria in sport and dance. Phys Ther Sport. 2018;32: 145–54. Available from: https://www.semanticscholar.org/paper/The-Beighton-score-as-a-predictor-of-Brighton-in-Armstrong-Greig/ec002f63523ca92712259066f44e4484d3a898b3 [cited: 12th Dec 2021]

[**37**] Aslan U, Çelik E, Cavlak U, Akdağ B. Evaluation of interrater and intrarater reliability of Beighton and Horan Joint Mobility Index'. Fizyoterapi Rehabilitasyon. 2006;17(3): 113–9. Available from: https://www.researchgate.net/publication/287350205_Evaluation_of_interrater_and_intrarater_reliability_of_Beighton_and_Horan_Joint_Mobility_Index [cited: 19th Dec 2021]

[**38**] Bale PJ, MacGregor AJ, Armon K, Easton V, Bacon H, Emma J. The Brighton Criteria Fail to Capture the Clinical Characteristics of Benign Joint Hypermobility Syndrome in Children: Data from the Bendy Study. Rheumatology. 2014;53(1): 167. Available from: https://doi.org/10.1093/rheumatology/keu125.010 [cited: 12th Dec 2021]

[**39**] Boyle KL, Witt P, Riegger-Krugh C. Intrarater and Interrater Reliability of the Beighton and Horan Joint Mobility Index. J Athl Train. 2003;38(4): 281–5. Available from: https://pubmed.ncbi.nlm.nih.gov/14737208/ [cited: 12th Dec 2021]

[**40**] Bulbena A, Mallorquí-Bagué N, Pailhez G, *et al.* Self-reported screening questionnaire for the assessment of Joint Hypermobility Syndrome (SQ-CH), a collagen condition, in Spanish population. Eur J Psychiatry. 2014;28(1): 17–26. Available from: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-61632014000100002 [cited: 12th Dec 2021]

[**41**] Calhil S, Sharkey M, Carter, C. Clinical assessment of generalized ligamentous laxity using a single test: is thumb-to-forearm apposition enough?. J Pediatr Orthop B. 2021;30(3): 296-300. Available from: https://pubmed.ncbi.nlm.nih.gov/32301823/ [cited: 12th Dec 2021]

[**42**] Chan C, Hopper L, Zhang F, Pacey V, Nicholson LL. The prevalence of generalized and syndromic hypermobility in elite Australian dancers. Phys Ther Sport. 2018;32: 15–21. Available from: doi: 10.1016/j.ptsp.2018.02.001 [cited: 12th Dec 2021]

[**43**] Cooper DJ, Scammell BE, Batt ME, Palmer D. Development and validation of self-reported line drawings of the modified Beighton score for the assessment of generalised joint hypermobility. BMC Med Res Methodol. 2018;18(1): 1-8. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-017-0464-8#citeas [cited: 12th Dec 2021]

[**44**] Czaprowski D, Kędra A, Pawłowska P, Kolwicz-Gańko A, Leszczewska J, Tyrakowski M. The Examination of the Musculoskeletal System Based Only on the Evaluation of Pelvic-Hip Complex Muscle and Trunk Flexibility May Lead to Failure to Screen Children for Generalized Joint Hypermobility. PLOS ONE. 2015;10(3): 1-12. Available

from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121360 [cited: 14th Dec 2021]

[**45**] Evans AM, Rome K, Peet L. The foot posture index, ankle lunge test, Beighton scale and the lower limb assessment score in healthy children: a reliability study. J Foot Ankle Res. 2012;5(1): 1757-1146. Available from: https://www.semanticscholar.org/paper/The-foot-posture-index%2C-ankle-lunge-test%2C-Beighton-Evans-Rome/ec70e08fed83e2414acd0fb6002f7860b61855be [cited: 12th Dec 2021]

[**46**] Farmer AD, Douthwaite H, Gardiner S, Aziz Q, Grahame R. A Novel in Vivo Skin Extensibility Test for Joint Hypermobility. J Rheumatol. 2010;37(7): 1513–8. Available from: https://doi.org/10.3899/jrheum.091192 [cited:]

[**47**] Glans M, Humble MB, Elwin M, Bejerot S. Self-rated joint hypermobility: the five-part questionnaire evaluated in a Swedish non-clinical adult population. BMC Musculoskelet Disord. 2020;21(174): 1-8. Available from: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-3067-1#citeas [cited: 12th Dec 2021]

[**48**] Hakim A, Grahame R. A simple questionnaire to detect hypermobility: An adjunct to the assessment of patients with diffuse musculoskeletal pain. Int J Clin Pract. 2003;57(3): 163–6. Available from: https://www.researchgate.net/publication/10779714_A_simple_questionnaire_to_detect_hypermobility_An_adjunct_to_the_assessment_of_patients_with_diffuse_musculoskeletal_pain [cited: 11th Dec 2021]

[**49**] Hansen A, Damsgaard R, Kristensen JH, Bagger J, Remvig L. Interexaminer Reliability of Selected Tests for Hypermobility. J Orthop Med. 2002;24(2): 48–51. Available from: https://www-tandfonline-com.ergo.southwales.ac.uk/action/showCitFormats?doi=10.1080%2F1355297X.2002.11736164 [cited: 12th Dec 2021]

[**50**] Jannson A, Saartok T, Werner S, Renström P. General joint laxity in 1845 Swedish school children of different ages: age and gender specific distributions. Acta Paediatrica. 2004; 93(9): 1202-6. Available from: https://www.google.com/search?client=firefox-b-d&q=Acta+Paediatrics [cited: 12th Dec 2021]

[**51**] Junge T, Jespersen E, Wedderkopp N, Juul-Kristensen B. Inter-tester reproducibility and inter-method agreement of two variations of the Beighton test for determining Generalised Joint Hypermobility in primary school children. BMC Pediatr. 2013;13(214): 1-8. Available from: https://bmcpediatr.biomedcentral.com/articles/10.1186/1471-2431-13-214#citeas [cited: 12th Dec 2021]

[**52**] Kulik. O, Gebska. M. Comparison of Beighton score and Brighton Criterion in order to diagnose joint hypermobility in children. J Educ Health Sport . 2018;8(5): 133–48. Available from: http://ojs.ukw.edu.pl/index.php/johs/article/view/5472 [cited: 12th Dec 2021]

[**53**] Kwon J-W, Lee W-J, Park S-B, Kim MJ, Jang SH, Choi CK. Generalized Joint Hypermobility in Healthy Female Koreans: Prevalence and Age-Related Differences. Ann Rehabil Med. 2013;37(6): 832-8. Available from: https://doi.org/10.5535/arm.2013.37.6.832 [cited: 12th Dec 2021]

[**54**] Lamari NM, Chueire AG, Cordeiro JA. Analysis of joint mobility patterns among preschool children. Rev Paul Med. 2005;123(3): 119–23. Available from: https://www.scielo.br/j/spmj/a/hZPMtBm9H4dJsM7chJbbNnH/?lang=en [cited: 12th Dec 2021]

[**55**] McGillis L, Mittal N, Santa Mina D, *et al.* Utilization of the 2017 diagnostic criteria for hEDS by the Toronto GoodHope Ehlers–Danlos syndrome clinic: A retrospective review. Am J Med Genet A. 2020;182(3): 484–92. Available from: https://www.researchgate.net/publication/337955567_Utilization_of_the_2017_diagnostic_criteria_for_hEDS_by_the_Toront

o_GoodHope_Ehlers-Danlos_syndrome_clinic_A_retrospective_review [cited: 12th Dec 2021]

[**56**] Mikołajczyk E, Szczygieł A, Jankowicz-Szymańska A, Kopf P. The occurrence of join hypermobility syndrome (JHS) in 15-year old girls and boys in the context of diagnostic, therapeutic and prophylactic problems. Pol J Physiother. 2012;12(3): 229–40. Available from: https://www.researchgate.net/publication/272893729_The_occurrence_of_join_hypermobility_syndrome_JHS_in_15-year_old_girls_and_boys_in_the_context_of_diagnostic_therapeutic_and_prophylactic_problems [cited: 12th Dec 2021]

[**57**] De Moraes DA de, Baptista CA, Crippa JAS, Louzada-Junior P. Translation into Brazilian Portuguese and validation of the five-part questionnaire for identifying hypermobility. Rev Bras Reumatol. 2011;51(1): 53–69. Available from: https://pubmed.ncbi.nlm.nih.gov/21412606/ [cited: 12th Dec 2021]

[**58**] Morris SL, O'Sullivan PB, Murray KJ, Bear N, Hands B, Smith AJ. Hypermobility and Musculoskeletal Pain in Adolescents. J Pediatr. 2017;181: 213-221.e1. Available from: https://pubmed.ncbi.nlm.nih.gov/27863848/ [cited: 12th Dec 2021]

[**59**] Naal FD, Hatzung G, Müller A, Impellizzeri F, Leunig M. Validation of a self-reported Beighton score to assess hypermobility in patients with femoroacetabular impingement. Int Orthop. 2014;38(11): 2245–50. Available from: https://pubmed.ncbi.nlm.nih.gov/24993650/ [cited: 12th Dec 2021]

[**60**] Öhman A, Westblom C, Henriksson M. Hypermobility among school children aged five to eight years: the Hospital del Mar Criteria gives higher prevalence for hypermobility than the Beighton score. Clin Exp Rheumatol. 2013;32(2): 285–90. Available from: https://www.scirp.org/%28S%28lz5mqp453edsnp55rrgjct55%29%29/reference/referencespapers.aspx?referenceid=1467436 [cited: 12th Dec 2021]

[**61**] Pearsall AW, Kovaleski JE, Heitman RJ, Gurchiek LR, Hollis JM. The relationships between instrumented measurements of ankle and knee ligamentous laxity and generalized joint laxity. J Sports Med Phys Fitness. 2006;46(1): 104–10. Available from: https://pubmed.ncbi.nlm.nih.gov/16596107/ [cited: 12th Dec 2021]

[**62**] Phan K, Nicholson LL, Hiller CE, Chan C. Prevalence and unique patterns of lower limb hypermobility in elite ballet dancers. Phys Ther Sport. 2020;41: 55–63. Available from: https://pubmed.ncbi.nlm.nih.gov/31759239/ [cited: 22nd Jun 2021]

[**63**] Romeo DM, Lucibello S, Musto E, *et al.* Assessing Joint Hypermobility in Preschool-Aged Children. J Pediatr. 2016;176: 162–6. Available from: https://pubmed.ncbi.nlm.nih.gov/27339250/ [cited: 12th Dec 2021]

[**64**] Riley SP, Grimes JK, Calandra K, Foster K, Peet M, Walsh MT. Reliability of the beighton score and impact of generalized joint mobility and resting scapular position on Median Neurodynamic Test 1. J Bodyw Mov Ther. 2020;24(4): 131–7. Available from: https://pubmed.ncbi.nlm.nih.gov/33218501/ [cited: 12th Dec 2021]

[**65**] Sauers EL, Borsa PA, Herling DE, Stanley RD. Instrumented Measurement of Glenohumeral Joint Laxity and its Relationship to Passive Range of Motion and Generalized Joint Laxity. Am J Sports Med. 2001;29(2): 143–50. Available from: https://pubmed.ncbi.nlm.nih.gov/11292038/ [cited: 1th Dec 2021]

[**66**] Schlager A, Ahlqvist K, Pingel R, Nilsson-Wikmar L, Olsson CB, Kristiansson P. Validity of the self-reported five-part questionnaire as an assessment of generalized joint hypermobility in early pregnancy. BMC Musculoskeletal Disord. 2020;21(1): 514-15. Available from: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-03524-7 [cited: 11th Dec 2021]

[**67**] Singh H, McKay M, Baldwin J, *et al.* Beighton scores and cut-offs across the lifespan: cross-sectional study of an

Australian population. Rheumatology (Oxford). 2017;56(11):1857–64. Available from: https://doi.org/10.1093/rheumatology/kex043 [cited: 11th Dec 2021]

[**68**] Sirajudeen MS, Waly M, Alqahtani M, *et al.* Generalized joint hypermobility among school-aged children in Majmaah region, Saudi Arabia. PeerJ. 2020;8: 1-13. Available from: https://pubmed.ncbi.nlm.nih.gov/32864212/ [cited: 11th Dec 2021]

[**69**] Skiwiot M., Śliwiński G, Milanese S, Śliwiński Z. Hypermobility of joints in dancers. PLoS One. 2019;14(2): 1-13. Available from: https://pubmed.ncbi.nlm.nih.gov/30794600/ [cited: 11th Dec 2021]

[**70**] Smits-Engelsman B, Klerks M, Kirby A. Beighton Score: A Valid Measure for Generalized Hypermobility in Children. J Pediatr. 2011;158(1): 119-123. Available from: https://pubmed.ncbi.nlm.nih.gov/20850761 [cited: 11th Dec 2021]

[**71**] Steinberg N, Hershkovitz I, Zeev A, Rothschild B, Siev-Ner I. Joint Hypermobility and Joint Range of Motion in Young Dancers. J Clin Rheumatol. 2016;22(4): 171–8. Available from: https://pubmed.ncbi.nlm.nih.gov/27219302/ [cited: 12th Dec 2021]

[**72**] Van der Giessen LJ, Liekens D, Rutgers KJ, Hartman A, Mulder PG, Oranje AP. Validation of beighton score and prevalence of connective tissue signs in 773 Dutch children. J Rheumatol. 2001;28(12): 2726–30. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6744937/ [cited: 12th Dec 2021]

[**73**] Ballenger LR, Moore-Clingenpeel M, Oberle EJ. Musculoskeletal ultrasound assessment in pediatric knee hypermobility: a case control study. Pediatr Rheumatol Online J. 2021;19(1): 58-66. Available from: https://ped-rheum.biomedcentral.com/articles/10.1186/s12969-021-00546-w [cited: 12th Dec 2021]

[**74**] Cherni Y, Desseauve D, Decatoire A, *et al.* Evaluation of ligament laxity during pregnancy. J Gynecol Obstet Hum Reprod. 2019;48(5): 351–7. Available from: https://pubmed.ncbi.nlm.nih.gov/30794956/ [cited: 12th Dec 2021]

[**75**] Cypel D. Gleno-humeral abduction measurement in patients with Ehlers-Danlos syndrome. Orthop Traumatol Surg Res. 2019;105(2): 287–90. Available from: https://www.researchgate.net/publication/331808395_Gleno-humeral_abduction [cited:]

[**76**] Heidbreder AE, Ringelstein EB, Dittrich R, Nabavi D, Metze D, Kuhlenbäumer G. Assessment of skin extensibility and joint hypermobility in patients with spontaneous cervical artery dissection and Ehlers–Danlos syndrome. J Clin Neurosci. 2008;15(6): 650–3. Available from: https://pubmed.ncbi.nlm.nih.gov/18400501/ [cited: 12th Dec 2021]

[**77**] McCormack M, Briggs J, Hakim A, Grahame R. Joint laxity and the benign joint hypermobility syndrome in student and professional ballet dancers. J Rheumatol. 2004;31(1): 173–8. Available from: https://pubmed.ncbi.nlm.nih.gov/14705238/ [cited: 12th Dec 2021]

[**78**] Meyer KJ, Chan C, Hopper L, Nicholson LL. Identifying lower limb specific and generalised joint hypermobility in adults: validation of the Lower Limb Assessment Score. BMC Musculoskeletal Disord. 2017;18(1): 514-23. Available from: https://doi.org/10.1186/s12891-017-1875-8 [cited: 12th Dec 2021]

[**79**] Parveneh V, Shiari R. Proposed modifications to Beighton criteria for the diagnosis of joint hypermobility in children. Indian Journal of Rheumatology. 2016;11(2): 97–100. Available from: [cited:]

[**80**] Ahn SY, Ko H, Yoon JO, Cho SU, Park JH, Cho KH. Determining the Reliability of a New Method for Measuring Joint Range of Motion Through a Randomized Controlled Trial. Ann Rehabil Med. 2019;43(6): 707–19. Available from:

https://doi.org/10.5535/arm.2019.43.6.707 [cited: 12th Dec 2021]

[**81**] Bockhorn LN, Vera AM, Dong D, Delgado DA, Varner KE, Harris JD. Interrater and Intrarater Reliability of the Beighton Score: A Systematic Review. Orthop J Sports Med. 2021;9(1): 1–11. Available from: https://www.semanticscholar.org/paper/Validation-of-a-self-reported-Beighton-score-to-in-Naal-Hatzung/c98b6340218344392426b6121e35ad8aa5677597 [cited: 14th Dec 2021]

[**82**] Palmer S, Denner E, Riglar M, Scannell R, Webb S, Young G. Quantitative measures of tissue mechanics to detect hypermobile Ehlers-Danlos syndrome and hypermobility syndrome disorders: a systematic review. Clin Rheumatol. 2020; 39: 715–725. Available from: [cited:]

[**83**] Desfor F. Assessment of Joint Hypermobility in Dancers. J Dance Med Sci. 2003;7(1): 17–23. Available from: https://www.ingentaconnect.com/content/jmrp/jdms/2003/00000007/00000001/art00003 [cited: 11th Dec 2021]

[**84**] Czaprowski D, Kotwicki T, Stoliński Ł. Assessment of joint laxity in children and adolescents - a review of methods. Ortopedia Traumatologia Rehabilitacja. 2012;14(5): 407-20. Available from: https://doi.org/10.5604/15093492.1016368 [cited: 12th Dec 2021]

[**85**] Drabik M, Byś A, Gawda P. Difficulties in diagnosis and selection of optimal diagnostic methods for hypermobile Ehlers Danlos Syndrome – literature review. Medycyna Ogólna i Nauki o Zdrowiu. 2020;26(2): 79–86. Available from: http://www.monz.pl/Difficulties-in-diagnosis-and-selection-of-the-optimal-diagnostic-methods-for-hypermobile,122300,0,2.html [cited: 15th Dec 2021]

[**86**] Day H, Koutedakis Y, Wyon MA. Hypermobility and Dance: A Review. Int J Sports Med. 2011;32(7): 485–9. Available from: https://pubmed.ncbi.nlm.nih.gov/21695667/ [cited: 14th Dec 2021]

[**87**] Grahame R. The need to take a fresh look at criteria for hypermobility. J Rheumatol. 2007;34(4): 664–5. Available from: https://pubmed.ncbi.nlm.nih.gov/17407226/ [cited: 11th Dec 2021]

[**88**] Kacheria T, Drangsholt M, Leresche L, Winocur E. Accuracy of a Shortened Beighton Examination Method for Hypermobility. International Association for Dental Research Poster session, 2006 AADR/CADR Annual Meeting, Orlando, Florida, AADR/CADR: 2006.

[**89**] Cincinnati Children's Hospital Medical Center Joint Hypermobility Team. Evidence-based clinical care guideline for Identification and Management of Pediatric Joint Hypermobility, In children and adolescents aged 4 to 21 years old, CCHMC EBDM Website Guideline 43 2014; [cited: 15 December 2021]. Available from: https://www.cincinnatichildrens.org/-/media/cincinnati%20childrens/home/service/j/anderson-center/evidence-based-care/recommendations/type/joint%20hypermobility%20guideline%2043.

[**90**] Skwiot M, Śliwiński G, Milanese S, Śliwiński Z. Hypermobility of joints in dancers. PLOS ONE. 2019;14(2): 1-13. Available from: https://pubmed.ncbi.nlm.nih.gov/30794600/ [cited: 11th Dec 2021]

[**91**] Cahill SV, Sharkey MS, Carter CW. Clinical assessment of generalized ligamentous laxity using a single test: is thumb-to-forearm apposition enough? J Pediatr Orthop. 2021;30(3): 296–300. Available from: https://pubmed.ncbi.nlm.nih.gov/32301823/ [cited: 12th Dec 2021]

[**92**] Bevilacqua DE. Measuring Joint Hypermobility Using the Hospital Del Mar Criteria - A Reliability Analysis Using Secondary Data Analysis. ARAR. 2019;1(1): 1–6. Available from: https://irispublishers.com/arar/fulltext/measuring-joint-hypermobility-using-the-hospital.ID.000502.php [cited: 12th Dec 2021]

[**93**] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med. 2012;22(3): 276–82. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/ [cited: 16th Dec 2021]

[**94**] Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18(3): 91–3. Available from: https://doi.org/10.1016/j.tjem.2018.08.001 [cited: 16th Dec 2021]

[**95**] Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ. 2018;48(6): 1273–96. Available from: https://doi.org/10.1007/s11165-016-9602-2 [cited: 16th Dec 2021]

[**96**] Hamonet C, Brock I. Joint mobility and Ehlers-Danlos syndrome, (EDS) new data based on 232 cases. J Arthritis. 2015;4(2): 1–5. Available from: http://claude.hamonet.free.fr/eng/eds_joint-mobility.htm [cited: 12th Dec 2021]

[**97**] Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3): 236–40. Available from: https://pubmed.ncbi.nlm.nih.gov/21145544/ [cited: 11th Dec 2021]

[**98**] McGrath TA, Alabousi M, Skidmore B, Korevaar DA, Bossuyt PM, Moher D, Thombs B, McInnes MD. Recommendations for reporting of systematic reviews and meta-analyses of diagnostic test accuracy: a systematic review. Syst Rev, 2017;6(194): 1-15. Available from: https://doi.org/10.1186/s13643-017-0590-8 [cited:]

[**99**] Whiting, PF, Rutjes A, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt, P, the QUADAS-2 Group. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies QUADAS, a tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36. Available from: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046. [cited:]

[**100**] Lijmer JG, Mol BW, Heisterkamp S, *et al.* Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests. J Am Med Assoc. 1999;282(11): 1061–6. Available from: https://doi.org/10.1001/jama.282.11 [cited: 16th Dec 2021]

[**101**] Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. CJEM. 2008;10(2): 174–5. Available from: https://www.researchgate.net/publication/5481885_Incorporation_bias_in_studies_of_diagnostic_tests_How_to_avoid_being_biased_about_bias [cited: 21st Dec 2021]

[**102**] Kea B, Hall MK, Wang R. Recognising bias in studies of diagnostic tests part 2: interpreting and verifying the index test. J Accid Emerg Med. 2019;36(8): 501–5. Available https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6693499/ [cited: 21st Dec 2021]