Artificial Intelligence In Medicine 173 (2026) 103332

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

The accuracy, validity and reliability of Theia3D markerless motion capture for studying the biomechanics of human movement: A systematic review

Florent Varcin [a], Mark G. Boocock [a,*]

[a] Ergonomics and Human Factors Group, School of Allied Health, Auckland University of Technology, Private Bag 92006, Auckland, 1142, New Zealand

ARTICLE INFO

Keywords:
Theia3D
Markerless motion capture
Biomechanics
Systematic review
Gait
Kinematics

ABSTRACT

Recent advancements in computer vision recognition combined with the use of pose estimation algorithms has led to a rapid increase in the use of 3D video-based markerless (ML) motion capture to study human movement. One such prominent system is Theia3D. To determine the accuracy, validity, and reliability of Theia3D, a systematic literature review was conducted across five electronic databases using the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines. Studies were included if they investigated the accuracy, validity, or reliability of Theia3D against a standardised method and reported on at least one biomechanical measure. A modified version of COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) and GRADE (Grading of Recommendations Assessment, Development, and Evaluation) were used to evaluate the quality of evidence. Sixteen studies met the inclusion criteria, the majority of which assessed the validity of kinematics during gait or running. Pooled lower limb kinematics showed reasonable accuracy, whilst hip flexion/extension and rotations of the lower limb joints in the transverse plane suggests poor accuracy. Most spatiotemporal gait parameters measured using Theia3D demonstrated excellent validity (Intraclass correlation coefficient (ICC) > 0.9) and inter-session reliability (gait speed, Standard Error of Measurement (SEM) ≤ 0.07 m/s; step/stride length, SEM ≤ 0.06 m; ICC > 0.95). The accuracy, validity, and reliability of Theia3D used in the biomechanical analysis of functional tasks and in different population groups shows promise. However, there is a need for improved methods by which to compare data and a standardisation of biomechanical modelling approaches.

1. Introduction

Traditional approaches to studying three-dimensional (3D) human movement have typically relied on optoelectrical systems to track markers attached to the human body. However, recent advancements in computer vision and recognition have led to 3D video-based motion capture without the need for markers. Referred to as markerless (ML) motion capture, it uses synchronised 2D video recordings to reconstruct 3D kinematics. Computer-based deep-learning algorithms detect changes in pixel values (static or moving pixels) to identify human shape and anatomical landmarks frame-by-frame, often referred to as pose estimation [1].

Several ML-based systems using pose estimation algorithms have been developed (e.g. OpenPose and DeepLabCut) [2,3]. Theia3D ML motion capture, developed by Theia Markerless Inc. (Kingston, ON, Canada), is a commercial AI-driven software platform that uses pose-estimation algorithms to capture and analyse 3D human movement. Using standard RGB video recordings from a series of cameras placed around the person(s) and activity of interest, deep learning models detect human pose in the captured volume. These models typically consist of deep learning convolutional neural networks (CNNs) trained to perform person detection and 2D keypoint localisation in each camera view [4]. The deep learning CNNs automatically identify key visual features of the human body in each image to locate anatomical landmarks, while a tracking module links detections across frames to maintain temporal consistency [2]. Theia3D detects and tracks over 120 anatomical landmarks on the person or persons of interest [5]. The ability to position landmarks on the human body stems from the training given to the models, which is reportedly derived from ‘annotated digital images of humans in the wild’ [5]. During training, the CNNs refine their parameters through backpropagation, improving the accuracy of movement detection [2,6]. Using the 2D positions of landmarks from the different camera views, which are synchronised and calibrated with the known dimensions of an object, an estimation is made of each landmark

---

* Corresponding author.
E-mail address: mark.boocock@aut.ac.nz (M.G. Boocock).

https://doi.org/10.1016/j.artmed.2025.103332
Received 5 January 2025; Received in revised form 30 November 2025; Accepted 1 December 2025

in 3D space. Body segments are constructed from the landmark positions and applied to a skeleton model using a kinematic chain. Within Theia3D, a skeletal model consisting of 17 rigid segments is scaled and optimised to follow the motion of the landmarks [5]. Finally, biomechanical constraints and temporal filtering are applied to obtain smooth and anatomically plausible joint angles [2,6].

Three-dimensional AI-based ML motion capture, such as Theia3D, represents a significant advancement in computer vision recognition and biomechanical analysis, potentially enabling large-scale, non-invasive human movement analysis across a range of medical, sport, and health-related (e.g. ergonomics) disciplines. However, to gain acceptance as a suitable alternative to biomechanical analysis, ML-based motion capture must demonstrate rigor in terms of its accuracy, validity, and reliability. Several studies [4,5,7–10] have evaluated Theia3D against conventional motion capture systems, reporting on these qualities across a range of different motor tasks. Bringing together this literature is important for users as it will provide information on the potential benefits and limitations of ML motion capture. Thus, this systematic review will examine the quality of evidence of those studies reporting on the accuracy, validity, or reliability of Theia3D ML motion capture. Whilst other ML motion capture systems have been developed, this review is restricted to Theia3D. The benefits of Theia3D over other ML systems are considered its broad generalisability for the study of diverse movements and its high degree of accuracy in resolving segmental rotations, enabling a complete pose estimation [3].

2. Methods

2.1. Research question

A systematic review was conducted during May 2024 and June 2024 to address the question: what is the current accuracy, validity and reliability of Theia3D ML motion tracking for studying the biomechanics (e.g. kinematics, kinetics) of human movement in different population groups performing different functional tasks? The question was structured according to the PECO statement [11], as detailed in Table 1.

2.2. Eligibility criteria

To be included in this review, studies could be of an observational or experimental design. They were required to have used Theia3D to study a functional task (e.g. gait, running, or jumping) and derived measures of accuracy, validity or reliability by comparing against an accepted and validated 3D human movement capture system (e.g. marker-based (MB) motion capture system and/or IMUs). Included studies were required to report on at least one biomechanical measure (kinematic, spatiotemporal parameters or kinetics inclusive of a kinematic measure) and involve more than 4 participants (Table 2). The latter criteria was based on the recommendations of Luciano et al. [12] who suggests that to detect subtle differences in human movement kinematics, sample size should range between 5 and 30 participants per group. Studies were restricted to those published in English. Congress abstracts, poster presentations, unpublished papers, and literature reviews were excluded. Although theses and dissertations were excluded based on the unknown peer-review status, an electronic database search of relevant authors was conducted based on the dissertations/thesis retrieved via ProQuest. A date restriction was placed on the year of publication (from 2017) as Theia3D was first launched in 2018 [13]. For all articles meeting the inclusion criteria, a manual search was undertaken of the reference lists.

Table 1
PECO statement used to structure the research question.

PECO framework	Criteria
Population (P)	Human participants with the ability to perform a functional task. Can include healthy or those with a known health condition.
Exposure (E)	Functional tasks or prescribed movement tasks, such as gait, running, or jumping.
Comparison (C)	Theia3D compared against an accepted and validated 3D human movement capture system, such as a marker-based (MB) motion capture and/or Inertial Measurement Units (IMUs).
Outcome (O)	The accuracy, validity or reliability of at least one biomechanical outcome measure, e.g. kinematic, spatiotemporal parameters or kinetics (derived from kinematics).

Table 2
The inclusion and exclusion criteria for this review.

Inclusion criteria	Exclusion criteria
Experimental or observational study designs. Studies using Theia 3D ML motion capture. Studies comparing the accuracy, validity, reliability and/or feasibility of Theia 3D ML motion capture against a validated 3D motion capture method, (e.g. Qualisys, Vicon, IMUs).	Non-human involvement, such as robotics, animal, or face recognition.
Healthy or non-heathy human participants able to undertake a functional task. Human functional tasks, such as gait, running, jumping, balance or manual lifting. Studies published after 2017.	Studies involving fewer than 5 participants.
Full text of peer-reviewed journal articles or those published in conference proceedings. English language publications.	Unpublished works, theses, posters, or abstracts.

2.3. Sources of information and search

This systematic review was developed in accordance with the PRISMA guidelines [14]. Five electronic databases (EBSCO Health (inclusive of Medline + CINAHL + SportsDiscus), SCOPUS, OVID (AMED), ProQuest, PROSPERO and Google Scholar (first 150 references) were searched using a combination of keywords associated with three domains: 1) Theia3D or ML motion capture; 2) human movement analysis, motion tracking or motion capture; and 3) accuracy, validity and reliability. A preliminary search of relevant articles served to inform the list of keywords, along with subject headings from the selected databases. A combination of the Boolean operators ‘AND’ and ‘OR’, along with a six-degree proximity operator (restricted words to within six from each other), enhanced the search strategy. An example of the search strategy used in the EBSCO database can be found in Appendix A, which was modified according to the search protocol of each database. A Senior Liaison Librarian with extensive experience in searching health and rehabilitation-related databases assisted in developing the search strategy. The search was performed independently by the two authors who completed their searches at the end of June 2024. All articles identified from each database were exported to EndNote™ (Version 21.0.1; Clarivate, London, UK) and duplicates removed.

2.4. Study selection

The two reviewers (FV and MB) initially discussed and agreed on the eligibility criteria, applying it to a selection of searched articles (10 %) to determine inter-examiner agreement. On reaching agreement on a study's eligibility, the two reviewers (FV and MB) independently screened articles based on their title and abstract (Stage 1) and articles not related to the topic area were removed. Stage 2 involved reviewing the full text of selected studies against the inclusion criteria to ensure eligibility. Reviewers met to agree on suitability of included studies. Throughout the selection process, reviewers were not blinded to authors' names or journal titles.

2.5. Data collection and extraction

Two reviewers (FV and MB) independently extracted data from studies meeting the inclusion criteria, according to pre-agreed categories and the recommendations of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) (Step 5) (Appendix B) [15]. The data extracted included: study identification (author, year of publication); size and characteristics of the studies population, study design and purpose (e.g. accuracy, validity or reliability); motion capture methods used (e.g. MB, ML, pressure mats); equipment used (e.g. manufacturer, number of cameras); location (laboratory based); statistical analysis; main findings; and conclusions. On completing data extraction, the reviewers met to agree on the appropriateness and completeness of the extracted data.

2.6. Risk of bias and quality assessment

A modified version of COSMIN [15–17] was used to assess the risk of bias (RoB) and methodological quality of selected studies. This evaluates each study against a set of criteria appropriate to intended study objectives ('Boxes'), i.e. Box 6: Reliability; Box 7: Accuracy; and Box 9: Validity. Within each box, a series of questions relating to each criterion are scored as: very good; appropriate; doubtful; inappropriate; or not mentioned. The 'worst score' method is used to determine the overall quality of a study. When determining the objectives of each study, the definitions proposed by Scataglini et al. [18] were used: accuracy was considered a measure of error and determined by absolute agreement between the same measured outcomes; validity (concurrent) was considered to be the relative agreement between variables when the methods are compared simultaneously; and reliability refers to the stability of the measurement across repeatable sessions (e.g. inter-session reliability), trials (e.g. inter-trial reliability), and raters' assessments (e.g. inter-rater reliability).

Content validity and imprecision were also assessed as part of the quality evaluation of each study. Content validity was evaluated using two additional boxes (Appendix C), which applied the same four-point rating scale used in other criteria (i.e. boxes). Box 2a included questions relating to video camera setup [19] based on the video-based motion analysis guidelines of Payton and Bartlett [20], i.e. video and camera settings appropriate to the capture volume (e.g. frame rate, resolution, exposure, number and positioning of cameras); the appropriateness of camera settings relative to the functional task performed; the health status of participants; participants apparel (e.g. clothing); and the environment (e.g. lighting conditions). If a study failed to provide information about these aspects of design, it was rated as 'not mentioned'. To assess the AI methodological quality of each study, a second box (Box 2b) was used to evaluate the transparency of the study in relation to the applied AI system, in this case Theia3D. This assessment was developed according to the recommendation of CONSORT-AI [21] and PROBAST-AI [22], and included criteria relating to: model architecture, training data documentation and methods, interpretation and explanation of AI decision process and outputs, study population characteristics and movement compatibility to the training data, and appropriate referencing of the AI system versions.

To determine the overall content validity of each study, the worst-case score across the two boxes (Box 2a and Box 2b) was initially selected. If more than five aspects of assessment across the two boxes were not mentioned, the overall rating was downgraded by one. Five aspects of the assessment were selected as a criterion for downgrading a study, as this was the median number of overall occurrences for which information about study's content validity was not mentioned. This ensured that studies were not disproportionately penalised for lacking certain technical details due to the proprietary nature of Theia3D. Setting the threshold at a median value ensured that the content validity score reflected a balanced cutoff.

The imprecision of a study was based on its sample size. A cut off value of 21 was used to differentiate between 'no risk' and 'serious risk' and represented the median number of participants across studies. This was considered reasonable given that Knudson [23] identified mean sample sizes of between 15 and 42 participants for biomechanics research studies conducted over the last 25 years. Furthermore, Luciano et al. [12] suggests that detecting kinematic differences of less than 10° requires an approximate sample size of between 5 and 30 participants per group. By adopting a value based on the sample sizes of included studies, the assessment of imprecision was anchored to a data-driven, empirically grounded criterion. Given that the sample sizes of included studies in this review ranged between 8 and 45 participants and showed a positive skewed distribution, the selection of a median value was considered appropriate. It ensured a robust measure of central tendency that was not influenced by a small number of very large or very small sample sizes.

When referring to ICC values as measures of validity and reliability, this review adopts the classification criteria of Portney and Watkins [24] (Table 3). For levels of accuracy, the error values of Auer et al. [25] are adopted and in the case of absolute reliability, the recommendations of McGinley et al. [26] are used (Table 3).

Two independent reviewers (FV and MB) assessed the quality of each study. To ensure consistency, reviewers met initially to establish a consensus on the interpretation of each checklist question and determined the studies primary objectives. On completing the modified COSMIN checklist for each study, the two reviewers met to determine possible differences in assessments and agree on a consensus.

2.7. Quality of evidence

A modified version of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) [17,27] tool was used to determine the overall quality of evidence of the pooled results. Modifications were implemented in accordance with COSMIN recommendations [15] and involved downgrading the quality of evidence from high quality depending upon an individual GRADE adjustment for each study. The GRADE adjustment was determined by combining the RoB, imprecision, and content validity ratings (Appendix D).

Study results were pooled according to activities, population characteristics and outcome measures. To determine the overall quality of evidence, adjustments were made to the evidence by assessing the combination of GRADE assessments across studies, with the worst score being adopted (Appendix D). For example, if at least one study was rated as 'serious', the overall quality of evidence was downgraded by one level, 'very serious' by two levels; and 'extremely serious' by three levels. Overall evidence was initial considered 'high' and downgraded to either 'moderate', 'low' or 'very low' [15].

Table 3
Measures, threshold criteria and corresponding descriptors used to evaluate accuracy, validity and reliability of Theia3D against comparable methods.

Measure	Criteria	Descriptor
Validity and reliability[*]		
°ICC	< 0.5	Poor
	≥ 0.5 and < 0.75	Moderate
	≥ 0.75 and < 0.9	Good
	≥ 0.9	Excellent
Accuracy		
°Errors	$< 3°$	Excellent
°Joint angles	$\geq 3°$ and $\leq 7°$	Reasonable
	$> 7°$	Poor
Reliability		
°Errors	$< 2°$	Acceptable
°Joint angles	$\geq 2°$ and $\leq 5°$	Reasonable
	$> 5°$	Not acceptable

[*] Note: validity and reliability criteria adapted from Portney and Watkins [22]; accuracy criteria adapted from Auer et al. [23]; and reliability criteria adapted from McGinley et al. [24].

3. Results

3.1. Study selection

Fig. 1 shows the PRISMA flow diagram of the search strategy and identification of studies meeting the inclusion criteria. A total of 1191 articles were initially identified across the 6 electronic databases. Following removal of duplicates and miscellaneous records, the titles, and abstracts of 742 were screened, of which 126 were considered appropriate for full text review. Ninety-one studies were excluded as they did not involve Theia3D ML tracking. Of the 35 remaining studies, 19 were excluded, of which three studies did not assess accuracy, validity, or reliability, three did not involve a functional task, twelve were master's or PhD theses, unpublished papers, or poster presentations, and one had a sample size of fewer than five participants. Sixteen studies met the inclusion criteria, and no additional studies or potential authors of interest were identified following a check of the reference lists of the included studies.

3.2. Study characteristics

All 16 studies were of a cross-sectional design and involved a total of 303 participants. Of these participants, 180 were adults with 85 females and 95 males. Of the 303 participants, 123 involved children, of which 66 were diagnosed with health-related condition (e.g. cerebral palsy (CP) (n = 46), cavovarus foot (n = 8), abnormal gait (n = 6), Charcot-Marie-Tooth (n = 3), spina bifida (n = 2), spinal arteriovenous malformation (n = 1)). Forty-six adults had a health-related issue, including knee osteoarthritis and various unspecified medical conditions. Three studies [9,10,28] did not present participant information and two studies [4,5] used the same population sample but reported on different outcome measures.

Of the 16 studies, two [4,29] reported on the accuracy, two [5,28] on the validity, and three [6,8,9] on the reliability of Theia3D. Seven studies [7,10,30–34] investigated both the accuracy and validity of Theia3D, and one [35] on the accuracy and reliability of Theia3D. The accuracy, validity and reliability of Theia3D was investigated in one study [36]. Whilst reporting on the validity of Theia3D, one study [28] also assessed its feasibility.

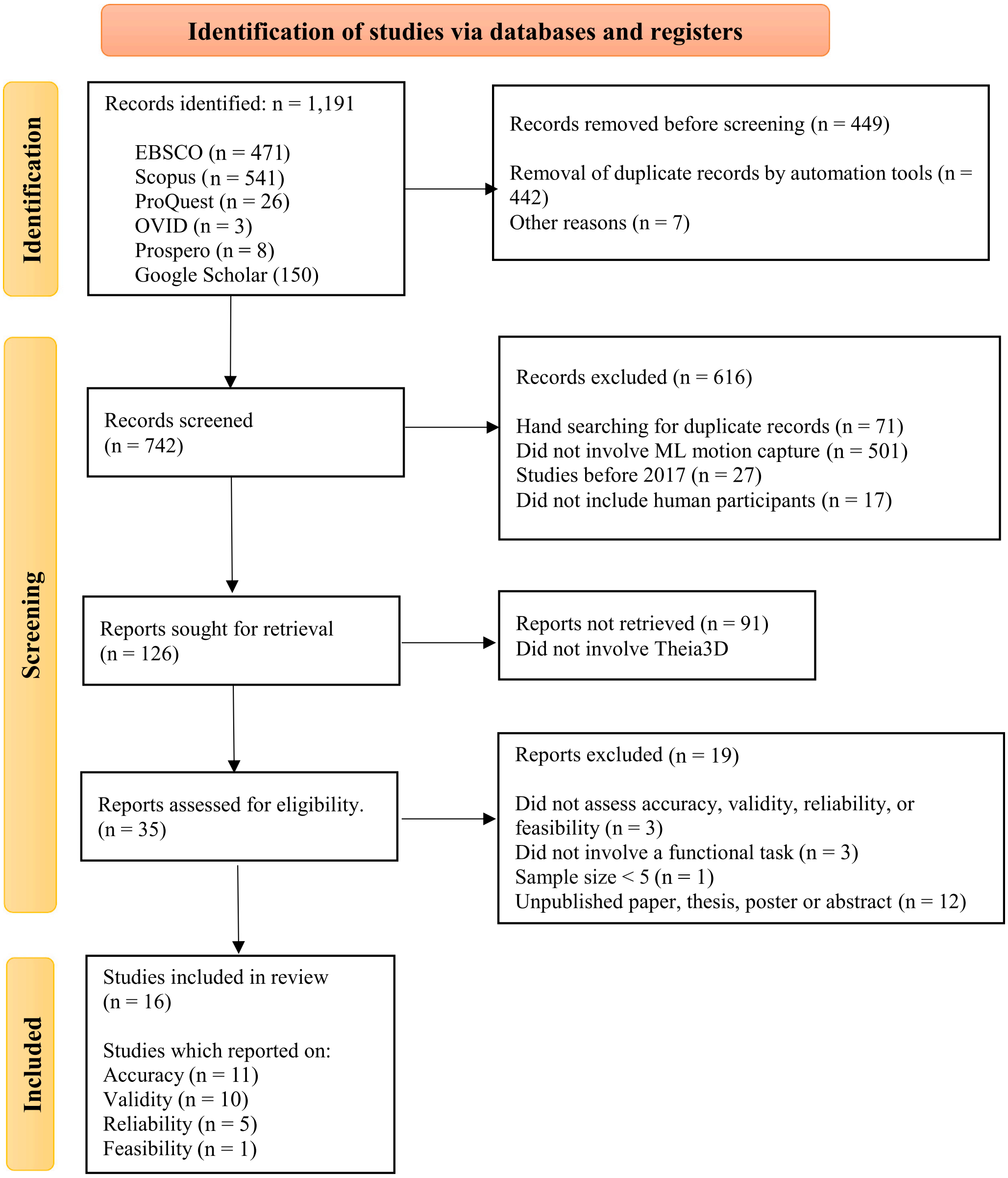


Fig. 1. PRISMA flow chart of the literature search strategy and inclusion of studies.

Studies primarily compared kinematic and kinetic outcome measures against MB motion capture systems, such as Qualysis ($n = 6$) (Qualisys AB, Gothenburg, Sweden), Vicon ($n = 4$) (Vicon Motion Systems Ltd., Oxford, United Kingdom) and Motion Analysis ($n = 1$) (Motion Analysis Corp., Rohnert Park, CA). In two studies [5,28], spatiotemporal parameters derived using Theia3D were compared to measures obtained from pressure-sensitive walkways (PSW) (GAITRite (CIR Systems, Inc., Franklin, NJ) and Stepscan (Charlottetown, PE, Canada)). No studies used IMU's motion capture as a comparison system. The most frequently assessed functional task was gait ($n = 10$), followed by running ($n = 5$) and jumping ($n = 3$). Other activities included loss of balance on a beam, an upper extremity task (box and block test), double leg squatting, and stepping down from a platform. With the exception of three studies [6,35,36], most reported on the software version of Theia3D used to analyse their data (Table 4). The versions of the software in these studies ranged from v2020.3.0.962 to v2023.1.0.3161.

3.3. The risk of bias and quality assessment of studies

Of the sixteen studies meeting the inclusion criteria, one [7] was considered to be of 'very good' overall methodological quality (i.e. RoB), eight were rated 'adequate' [4–6,10,28,30,31,33], five 'doubtful' [8,9,32,35,36] and two [29,34] 'inadequate' (Table 5 and Appendix E). For accuracy and reliability, whilst none of the studies referred to 'blindness of the administration', as data collection often occurred simultaneously, this was not considered a basis on which to downgrade a study. When considering accuracy, if a study only reported on Root Mean Square Difference (RMSD) between the two methods, the overall quality of the study was downgraded by one, due to the potential bias associated with the use of RMSD [10].

When combining the RoB rating with corresponding content validity and imprecision ratings to determine the GRADE adjustment (Table 6), four studies [4,5,28,30] were not considered to present 'serious' concerns, three [7,9,31] 'serious' concerns, six [6,8,10,32,33,36] 'very serious' concerns, and three [29,34,35] 'extremely serious' concerns. When assessing content validity, most studies failed to mention camera shutter speeds or lighting conditions, which are considered important parameters that can influence the accuracy of ML tracking [13]. Only three studies [7,8,31] provided sufficient details about the position and orientation of the camera systems.

Table 4
The version of Theia3D software used and the AI transparency score for included studies.

Authors	Version of Theia3D
Validity	
Chaumeil et al. [4]	v2021.2 (Supplementary material) v2023.1.0.3160 (used in the study)
Kanko et al. [3]	v2020.3.0.962
McGuirk et al. [21]	v2021.2.0.1675
Song et al. [25]	v2022.1.0.2309
Strutzenberger et al. [26]	v2020.6.0.1106
Tang et al. [8]	v2022.1.0.2309
Tang et al. [27]	v2022.1.0.2309
Wren et al. [28]	v2021.2.0.1675
Hansen et al. [30]	NM
Accuracy	
Kanko et al. [5]	v2021.1.0.1450
Kanko et al. [24]	v2022-1-0-2309
Song et al. [25]	v2022.1.0.2309
Strutzenberger et al. [26]	v2020.6.0.1106
Tang et al. [8]	v2022.1.0.2309
Tang et al. [27]	v2022.1.0.2309
Wren et al. [28]	v2021.2.0.1675
Moran et al. [29]	NM
Wishaupt et al. [22]	v2023.1.0.3160
Hansen et al. [30]	NM
Reliability	
Kanko et al. [23]	NM
Outerleys et al. [6]	v2023.1.0.3161
Riazati et al. [7]	v2021.2.0.1675
Moran et al. [29]	NM
Hansen et al. [30]	NM

Abbreviations: NM = not mentioned.

Table 5
Risk of bias (RoB) of included studies.

Authors	Accuracy	Validity	Reliability	RoB
Chaumeil et al. [7]	VG	VG		VG
Hansen et al. [36]	D	VG	D	D
Kanko et al. [6]			A	A
Kanko et al. [5]		A		A
Kanko et al. [4]	A			A
Kanko et al. [30]	A	A		A
McGuirk et al. [28]		A		A
Moran et al. [35]	D		D	D
Outerleys et al. [8]			D	D
Riazati et al. [9]			D	D
Song et al. [31]	A	A		A
Strutzenberger et al. [32]	D	A		D
Tang et al. [33]	A	A		A
Tang et al. [10]	A	A		A
Wishaupt et al. [29]	I			I
Wren et al. [34]	I	I		I

Abbreviations: RoB: risk of bias; I: inadequate; D: doubtful; A: adequate; VG: very good.

Table 6
GRADE adjustment ratings for each study based on their RoB, content validity and imprecision rating.

Authors	RoB	Content validity	Imprecision (sample size)	GRADE adjustment
Chaumeil et al. [7]	VG	A	S	Serious
Hansen et al. [36]	D	D	N	Very serious
Kanko et al. [6]	A	D	S	Very serious
Kanko et al. [5]	A	A	N	No
Kanko et al. [4]	A	A	N	No
Kanko et al. [30]	A	A	N	No
McGuirk et al. [28]	A	A	N	No
Moran et al. [35]	D	I	N	Extremely serious
Outerleys et al. [8]	D	A	S	Very serious
Riazati et al. [9]	D	A	N	Serious
Song et al. [31]	A	A	S	Serious
Strutzenberger et al. [32]	D	D	N	Very serious
Tang et al. [33]	A	D	S	Very serious
Tang et al. [10]	A	D	S	Very serious
Wishaupt et al. [29]	I	D	N	Extremely serious
Wren et al. [34]	I	I	N	Extremely serious

Note: Refer to Table 5 for the RoB rating. Refer to Appendix E for the content validity. Imprecision takes into consideration the sample size of the study, i.e. >20, no serious RoB; ≤21, serious risk of bias.
Abbreviations: RoB: risk of bias; N: no risk; S: serious; D: doubtful; A: adequate; VG: very good.

3.4. Accuracy and validity

For the three main functional tasks (gait, running and jumping), eleven studies [4,7,10,29–36] investigated the accuracy of Theia3D and ten studies [5,7,10,28,30–34,36] assessed the validity of Theia3D against alternative systems. The main findings from these studies were pooled to determine the overall quality of evidence.

Table 7
A summary of the kinematic and spatiotemporal measures used to compare Theia3D with other motion capture systems during gait.

Outcomes	Main findings	Number of studies	GRADE assessment
Kinematics, healthy adults	Hip:		Moderate
	Flex/Ext: $10^\circ \leq \text{RMSD} \leq 11^\circ \mid (R^2 = 0.99)$	2	
	Abd/Add: $2.6^\circ \leq \text{RMSD} \leq 3^\circ$	2	
	Int/Ext rot: $6.9^\circ \leq \text{RMSD} \leq 13.6^\circ$	2	
	Knee:		
	Flex/Ext: $3.3^\circ \leq \text{RMSD} \leq 3.4^\circ \mid (R^2 = 0.965)$	2	
	Abd/Add: $\text{RMSD} = 3.2^\circ$	1	
	Int/Ext rot: $\text{RMSD} = 13.2^\circ$	1	
	Ankle:		
	DFlex/PFlex: $3.4^\circ \leq \text{RMSD} \leq 6.7^\circ$ $(R^2 \geq 0.936)$	2	
	Abd/Add: $\text{RMSD} = 7^\circ$	1	
	Int/Ext rot: $\text{RMSD} = 11.6^\circ$	1	
Kinematics, CP children	Trunk:		Very low
	Flex/Ext: $5.1^\circ \leq \text{RMSD} \leq 11.8^\circ$	2	
	Abd/Add: $1.7^\circ \leq \text{RMSD} \leq 3.1$	2	
	Int/Ext rot: $4.1^\circ \leq \text{RMSD} \leq 5.1$	2	
	Pelvis		
	Flex/Ext: $9.1^\circ \leq \text{RMSD} \leq 11^\circ$	2	
	Abd/Add: $2.1^\circ \leq \text{RMSD} \leq 2.9^\circ$	2	
	Int/Ext rot: $\text{RMSD} \leq 5.2^\circ$	2	
	Hip:		
	Flex/Ext: $7.7^\circ \leq \text{RMSD} \leq 11.9^\circ$	2	
	Abd/Add: $3.9^\circ \leq \text{RMSD} \leq 5^\circ$	2	
	Int/Ext rot: $10^\circ \leq \text{RMSD} \leq 13^\circ$	2	
	Knee:		
	Flex/Ext: $\text{RMSD} \leq 4.8^\circ$	2	
	Abd/Add: $4.7^\circ \leq \text{RMSD} \leq 5.6^\circ$	2	
	Int/Ext rot: $9^\circ \leq \text{RMSD} \leq 10.2^\circ$	2	
	Ankle:		
	Flex/Ext: $4.9^\circ \leq \text{RMSD} \leq 6.4^\circ$	2	
	Foot progression: $3.9^\circ \leq \text{RMSD} \leq 7.8^\circ$	2	
Spatiotemporal, gait parameters	Speed (m/s): $0.97 \leq \text{ICC} \leq 0.99$	2	High
	Step length (m): $0.87 \leq \text{ICC} \leq 0.99$	2	
	Stride length (m): $0.97 \leq \text{ICC} \leq 0.99$	2	
	Stride width (m): $0.44 < \text{ICC} < 0.86$	2	
	Step time (s): $0.82 \leq \text{ICC} \leq 0.99$	2	
	Swing time (s): $0.42 \leq \text{ICC} \leq 0.86$	2	
	Stance time (s): $0.62 \leq \text{ICC} \leq 0.97$	2	
	Double limb support time (s): $0.03 \leq \text{ICC} \leq 0.82$	2	

Abbreviations: Flex: flexion; Ext: extension; Int: interior; Ext: exterior; Rot: rotation; RMSD: root mean square difference; Abd: abduction; Add: adduction; CP: cerebral palsy; R^2: correlation coefficient. DFlex: dorsiflexion; PFlex: plantar flexion; ICC: intraclass correlation coefficient.

3.4.1. *Kinematics*

During gait, four studies [4,29,31,34] considered the accuracy between Theia3D and MB systems for lower limb joint angles (i.e. hip, knee, ankle) (Table 7). Song et al. [31] reported strong correlation ($R^2 \geq 0.94$) between ML and MB for the three lower limb joints (hip, knee and ankle) in the sagittal plane. In healthy adults, Kanko et al. [4] and Song et al. [31] found that the smallest average differences between ML and MB methods occurred for hip abduction/adduction ($2.6^\circ \leq \text{RMSD} < 3^\circ$), indicating excellent agreement (Table 3). In contrast, hip flexion/extension showed the largest average differences ($10^\circ \leq \text{RMSD} \leq 11^\circ$), whilst internal/external rotation exhibited the largest peak differences, corresponding to poor agreement. Overall, the quality of evidence across those studies investigating the accuracy of Theia3D during gait in healthy participants was considered moderate.

Two studies [29,34] investigated the lower kinematics of gait in children with functional impairment (i.e. cerebral palsy) (Table 7). Both studies showed excellent agreement for the accuracy of trunk and pelvis kinematics in the frontal plane ($\text{RMSD} < 3^\circ$). There was reasonable agreement for the accuracy of hip and knee angles in the frontal plane ($3.9^\circ \leq \text{RMSD} \leq 5.6^\circ$), knee and ankle angles in the sagittal plane ($3.8^\circ \leq \text{RMSD} \leq 6.4^\circ$), and pelvis and trunk angles in the transverse plane ($4.1^\circ \leq \text{RMSD} \leq 5.2^\circ$). Larger differences were reported for the trunk, pelvis and hip angles in the sagittal plane ($5.1^\circ \leq \text{RMSD} \leq 11.9^\circ$), as well as for the hip and knee angles in the transverse plane ($9^\circ \leq \text{RMSD} \leq 10^\circ$), indicating overall poor agreement. Based on a statistical parametric mapping (SPM), Wishaupt et al. [29] found a significant difference between Theia3D and MB in sagittal (trunk, pelvis, hip and ankle joints), transverse (hip and knee); and frontal (knee) plane kinematics in children with CP during the gait cycle. The overall quality of evidence for those studies investigating the accuracy of Theia3D in children with CP was considered very low.

Two studies [30,31] examined lower limb kinematics during running (Table 8), whilst two others assessed jumping [31,32] (Table 9), revealing varying levels of agreement between Theia3D and MB systems depending on the joint and plane of motion. The largest differences occurred for hip rotations in the sagittal and transverse planes, for both running and jumping ($5.2^\circ \leq \text{RMSD} \leq 20.6^\circ$). Whereas differences in ankle kinematics were smaller ($3.8^\circ \leq \text{RMSD} \leq 8^\circ$) across the different planes of motion, for the two activities. The quality of evidence associated with those studies investigating the accuracy of Theia3D during running was considered moderate, whilst there was low evidence for those studies investigating jumping activities.

3.4.2. *Spatiotemporal gait parameters*

Two studies [5,28] found close similarity between spatiotemporal measures during gait when comparing Theia3D with pressure sensitive walkways (GAITRite and Stepscan) (Table 7). Intraclass Correlation Coefficients (ICC) (2,1) showed good to excellent agreement (Table 3)

Table 8
A summary of the kinematic and kinetic measures used to compare Theia3D with other motion capture systems during running.

Outcomes	Main findings	Number of studies	GRADE assessment
Kinematics, healthy adults	Hip:		Moderate
	Flex/Ext: 5.2° ≤ RMSD ≤ 13.8°	2	
	Abd/Add: 2.7° ≤ RMSD ≤ 4.5°	2	
	Int/Ext Rot: 9.8° ≤ RMSD ≤ 12.8°	2	
	Knee:		
	Flex/Ext: 4.2° ≤ RMSD ≤ 4.8°	2	
	Abd/Add: RMSD = 7.2°	1	
	Int/Ext Rot: RMSD = 9.4°	1	
	Ankle:		
	DFlex/PFlex: 3.8 ≤ RMSD ≤ 5.1°	2	
	Abd/Add: RMSD = 6°	1	
	Int/Ext Rot: RMSD = 8°	1	
Kinetics, joint moments, healthy adults	RMSD:		Moderate
	Hip Rot < Ankle Flex/Ext < Hip Ab/Ad Knee Flex/Ext < Hip Flex/Ext	2	

Abbreviations: Flex: flexion; Ext: extension; Int: interior; Ext: exterior; Rot: rotation; RMSD: root mean square difference; Abd: abduction; Add: adduction; CP: cerebral palsy; DFlex: dorsiflexion; PFlex: plantar flexion.

Table 9
A summary of kinematic measures comparing Theia3D with other motion capture systems during jumping tasks.

Outcomes	Main findings	Number of studies	GRADE assessment
Kinematics, healthy adults	Hip:		Low
	Flex/Ext: 12.1° ≤ RMSD ≤ 20.6°	2	
	Abd/Add: RMSD = 2.3°	1	
	Int/Ext Rot: RMSD = 11°	1	
	Knee:		
	DFlex/PFlex: 3.4° ≤ RMSD ≤ 4.4°	2	
		2	
	Ankle:		
	Flex/Ext: 5.0° ≤ RMSD ≤ 5.3°		

Abbreviations: Flex: flexion; Ext: extension; Int: interior; Ext: exterior; Rot: rotation; RMSD: root mean square difference; Abd: abduction; Add: adduction; DFlex: dorsiflexion; PFlex: plantar flexion.

Table 10
A summary of the differences in hip, knee and ankle joint centres when comparing Theia3D with other motion capture systems during gait, running and jumping.

Outcomes	Activities	Main findings RMSD (cm)			Number of studies	GRADE assessment
		Hip	Knee	Ankle		
Joint centre position	Gait	3.6	2.2	2.4	1	Low
	Running	2.96	2.15	2.21	1	
	Jumping	3	2	2.8	1	

Abbreviations: RMSD: root mean square difference.

for most gait parameters (e.g. speed, step and stride lengths), with the exception of stance time (moderate to excellent agreement (ICC ≥ 0.62)); swing time and stride width (moderate to good agreement (0.42 ≤ ICC ≤ 0.86)); and double limp support time (poor to good agreement (0.03 ≤ ICC ≤ 0.82)) (Table 7). The quality of evidence associated with those studies investigating the validity of Theia3D and spatiotemporal measures during gait was considered high.

3.4.3. *Kinetics*

Two studies [30,31] used average RMSD to compare the kinetics of running using Theia3D and MB motion capture (Table 8). These studies expressed joint moments differently, either normalised to body weight (BW; Nm/kg) [30] or as a percentage of height multiplied by body weight (%H * BW) [31]. Both studies reported similar results between approaches, with a RMSD for hip flexion/extension of less than 0.29 Nm/kg and 1.35 %H * BW, respectively, whilst for hip rotation, less than 0.09 Nm/kg and 0.14 %H * BW, respectively. The effect of running speed on kinetics was found to significantly influence RMSD between Theia3D and MB motion capture [10,33]. The overall quality of evidence for those studies investigating the accuracy of Theia3D for measuring the kinetics of running was considered moderate.

3.4.4. *Joint centre position*

Three studies [4,30,32] investigated the accuracy and validity of joint centre position between Theia3D and MB motion capture during three different tasks (gait, running and jumping) (Table 10). Across studies and activities, RMSD values were largest at the hip (range = 2.96 cm – 3.6 cm) and smallest at the knee (range = 2 cm – 2.2 cm). The overall quality of evidence for the accuracy and validity of Theia3D in determining joint centre position was low for gait, running, and jumping.

3.5. *Reliability*

Five studies [6,8,9,35,36] investigated inter-session or intra-trial reliability of Theia3D. Of these, three studies investigated gait [6,8,9], one running [35] and one [36] used the box and block test to assess upper limbs kinematics. Two studies [9,36] investigated the inter-session reliability of Theia3D. One investigated lower limb kinematics in 21 adult participants of varying age (range = 18 to 73 years) during gait [9], whilst the other [36] investigated upper limb kinematics in 19 paediatric participants (aged 7, 9 or 11 years) performing the box and block test. The three studies [6,8,35] investigating both inter-session

Table 11

A summary of biomechanical outcome measures used to determine the reliability of Theia3D during gait and running.

Outcomes	Main findings	Number of studies	Quality of evidence
Gait			
Spatiotemporal parameters	Inter-session		Low
	Gait speed (m/s): $0.04 \leq$ SEM ≤ 0.07	2	
Self-selected speed	Step length (m): $0.01 \leq$ SEM ≤ 0.04	2	
	Stride length (m): $0.24 \leq$ SEM ≤ 0.06	2	
Fast-selected speed	No effect of speed	2	
	ICC > 0.95	1	
Kinematics	Inter-session		Low
Lower joints	$1.5^\circ \leq$ var $\leq 2.8^\circ$ (on average)	2	
	$0.96^\circ \leq$ RMSD $\leq 3.71^\circ$	2	
	(T plane var $>$ F&S planes)		
	Intra-trial		
	$1.15^\circ \leq$ var $\leq 2.5^\circ$	2	
	(T plane var $>$ F&S planes)	2	
	$1.1 \leq$ VarRat ≤ 1.27	2	
Running			
Kinematics	Inter-session		Very low
Lower joints	$0.6^\circ \leq$ var $\leq 1.5^\circ$	1	
	Max knee flex/ext. (2.3°)		
	Intra-trial		
	$1.1 \leq$ var $\leq 2.6^\circ$	1	
	Max knee flex/ext (4.7°)	1	
	$0.55 \leq$ VarRat ≤ 0.88		

Note: 'var' refers to variability described by Schwartz et al. [37].

Abbreviations: SEM: standard error of measurement; var.: variability; VarRat: variability ratio, ICC: intraclass correlation coefficient; T: transverse; S: sagittal; F: frontal.

and intra-trial reliability of Theia3D for determining lower limb kinematics used the variability ratio (intra-trial divided by the inter-session reliability) as an indicator of the average variability [37]. The main findings and the quality of evidence for those studies assessing the reliability of Theia3D to derive kinematics during gait and running are shown in Table 11.

3.5.1. *Kinematics*

When investigating inter-session reliability of lower limb gait kinematics, reasonable variability (Table 3) was found across three studies [6,8,9], with two studies reporting RMSD ranging between 0.96° and 3.7° (Table 11). The inter-session reliability for hip abduction/adduction and knee rotation was inconsistent between two studies [6,9]. Intra-trial variability was lower than inter-session variability [6,8], with multiple session joint angle variability increasing by between 10 % and 27 % compared to the trial variability (Table 11). One study [35] reported inter-session variability of $\leq 1.5^\circ$ and intra-trial variability of $\leq 2.6^\circ$ for running, with an average variability ratio of 0.67 (Table 11). The overall quality of evidence for those studies investigating the reliability of Theia3D for measuring gait kinematics was considered low.

3.5.2. *Spatiotemporal gait parameters*

Two studies [8,9] investigated the reliability of spatiotemporal gait parameters and found Theia3D to have acceptable inter-session reliability (Table 3) for gait speed, step length, and stride length under two speed conditions (SEM ≤ 0.07; ICC ≥ 0.95) (Table 11). However, the methodological quality of these studies was rated as low.

4. Discussion

The aim of this review was to shed light on the accuracy, validity (concurrent), and reliability of Theia3D ML motion capture for the study of human movement. Sixteen studies met the inclusion criteria and were subject to a methodological quality assessment (COSMIN) and overall evidence was evaluated using the GRADE approach.

Studies evaluating the accuracy of lower-limb kinematics during gait, running, and jumping have reported agreement between Theia3D and other MB systems ranging from excellent ($< 3^\circ$; Table 3) to reasonable ($\geq 3^\circ$ and $\leq 7^\circ$; Table 3). This level of consistency was observed in the kinematics of the knee and ankle joints in the sagittal plane, as well as the hip and knee joints in the frontal plane ($2.3^\circ \leq$ RMSD $< 7^\circ$), across different populations. In contrast, hip flexion/extension and joint rotation in the transverse plane showed poor agreement ($> 7^\circ$; Table 3) between Theia3D and MB motion capture systems for gait and running ($6.9^\circ \leq$ RMSD $\leq 13.6^\circ$), with jumping leading to peak RMSD of up to 20.6°. These findings contrast with the systematic review of Scataglini et al. [18], which reported that across various ML-based motion capture systems, kinematic parameters of the hip and knee in the sagittal plane were valid and reliable, whereas, lower limb kinematics in the transverse and frontal planes lacked validity and accuracy. These inconsistencies may be due to the broad range of ML motion capture methods considered in their review, as well as their restriction of only including gait studies.

To understand the potential sources of kinematic variability across studies, it is important to examine the methods employed by ML and MB approaches in deriving joint angles. Studies of kinematic accuracy [4,29,31,32] often reported a hip offset. For example, Wishaupt et al. [29] and Strutzenberger et al. [32] identified significant differences in hip angles between Theia3D and MB approaches when analysed using SPM. These variations are most likely attributable to differences in biomechanical model definitions, the pelvis orientations specified by each model, and the software versions employed. Theia3D aligns joint segments with a global reference system, ensuring a consistent orientation, whilst MB systems define segment orientation according to a 'static' reference pose, which in most cases is a standing posture [38]. This introduces personalised orientations of the segment, such as variations in anterior/posterior tilt of the pelvis between participants. Kanko et al. [30] advocate for adopting the body segment local coordinates used by Theia3D as a global standard for developing future models, which would also remove the need for static trials.

The commercial nature of Theia3D means that there is limited information available on the landmark identifications performed by the CNNs [3]. AI-based pose estimation faces additional challenges in clinical environments, where variations in lighting, clothing, occlusions, and camera positioning can compromise model accuracy. Domain shifts caused by environments or patient populations not represented in the training dataset may also affect ML motion capture performance. Consequently, it remains unclear whether differences reported between Theia3D and MB systems are due to tracking inaccuracies, or the biomechanical models constructed from the movement data.

Differences in hip rotation between Theia3D and MB approaches may primarily reflect the errors associated with MB systems in tracking hip and pelvic bony landmarks. Markers placed on the pelvis and hip are prone to significant errors due to problems in identifying bony landmarks and soft tissue artefacts [3,38,39]. As in most cases, determining the validity and accuracy of any system is dependent upon the method of comparison, with MB approaches being considered 'accepted technology', rather than the 'gold standard' [3,30]. Whilst X-ray reconstruction of moving morphology (XROMM) [40] and dual fluoroscopy (DF) [41] are generally viewed as the gold standards for capturing bone movement in biomechanical analysis [42], this technology is often restricted to a laboratory environment and unable to be used for evaluating gross movements, such as whole body gait [43].

Comparisons of MB systems with bi-planar videography suggests that joint rotations captured using MB system are a concern, leading to large errors [3]. These rotations often involve subtle movements that when capture using MB motion systems are easily obscured or amplified by soft tissue artifacts, small errors in marker placement, different camera perspectives, and marker and body segment occlusions, which is further complicated by differences in how systems define anatomical axes [8]. Consequently, it seems inappropriate to draw definitive conclusions about the accuracy of ML motion capture in tracking joint movements, particular for those in the transverse plane, until more appropriate gold standard approaches become available, or more appropriate comparison systems are employed.

In addition to camera-based systems, other motion capture methods have been developed in recent years. For example, inertia measurement units (IMUs) show promise in capturing and analysing human movement [44,45]. Using hybrid systems that incorporate IMUs could help overcome some of the common problems faced by MB approaches in studying 'real-life' activities of daily living, such as those arising from occlusion-related errors and pose estimation. However, these systems face similar issues to the 'gold standard' MB approaches. For example, studies have raised concerns about the accuracy of IMUs due to calibration issues, noise artifacts, and limitations in sensor fusion algorithms, that affect the precision when recording and modelling human movements [46].

When used to evaluate the gait kinematics of children with CP, Theia3D shows poor accuracy for trunk, pelvis and hip flexion/extension ($5.1^\circ \leq \text{RMSD} \leq 11.9^\circ$), and knee and hip rotation ($9^\circ \leq \text{RMSD} \leq 13^\circ$), when compared to MB approaches. As previously mentioned, joint kinematics related to the pelvis are problematic for MB systems due to errors associated with marker placement, soft tissue artifacts, model definitions and pelvis orientations. These are likely to be further compounded by pathological conditions inherent within CP children. Uncertainties also exist regarding the dataset used to train Theia3D's deep learning model and its applicability to children with CP. Limited diversity in the training dataset would likely contribute to reduced accuracy, particularly in paediatric and pathological populations. This highlights the importance of AI transparency, as clinicians need to understand the limitations of the model's training data and expected performance to appropriately interpret results and implement AI-based motion capture in clinical practice. Thus, knowing the true source of kinematic variability and whether it relates to differences between the systems themselves, the biomechanical models adopted, or the pathology and population characteristics of the participants (e.g. children), remains a challenge when trying to understand the advantages and disadvantages of Theia3D, or other ML motion capture systems.

The majority of studies (4) in this review [4,30,31,34] elected to express accuracy in terms of the absolute magnitude of the average RMSD between Theia3D and MB capture systems. A limitation of using RMSD alone as a measure of accuracy is that whilst it provides important information about the overall magnitude of difference, it does not inform statistical hypothesis testing. In this review, the accuracy of kinematic measures was classified according to a single definition of the error magnitude [25], but it is encouraged to evaluate the underlying numeric value when assessing accuracy, given that this will differ according to the joint studied, the plane of movement and the type of activity. This also has implications for determining the clinical relevance of the kinematic measure. It has been suggested that to gain a true understanding of accuracy, RMSD should be complemented with other appropriate statistical tests, such as SPM analysis [10].

Those studies assessing the inter-session variability of Theia3D across a range of activities reported joint angle differences of less than 3.7° for all joints in both the frontal and sagittal planes [6,8,9]. Whilst McGinley et al. [26] suggests that a reliability range of between 2° and 5° for lower-body kinematics during gait is reasonable, it does warrant clinical consideration. An error exceeding 5° should be regarded as significant and enough to potentially distort interpretation [26]. Based on the inter-session repeatability of those studies included in this review [6,8,9,35], it appears there is reasonable repeatability when measuring lower limb joint angles in the frontal and sagittal planes, whilst greater caution is needed for those measured in the transverse plane. McGinley et al. [26] also considers walking speed to be a potential source of variability associated with the reliability of gait kinematics. Interestingly, the two studies that used Theia3D to investigate the effects of walking speed on joint kinematics found no effects on inter-session reliability [8,9].

It has been suggested that a person's clothing could be a source of inter-session variability when using ML tracking [13]. Although Riazati et al. [9] found that wearing long skirts impacted on the modelling of the pelvis and legs during walking, Keller et al. [47] suggests that clothing has a negligible influence on meaningful clinical interpretations. The continuous updating and training of the deep learning model used by Theia3D will likely reduce the effects of clothing on joint segment tracking [8]. However, wearing well-fitting clothing and apparel that contrasts with the environment is recommended to optimise pose estimation accuracy during ML motion capture [13].

For those studies included in this review, intra-trial variability was found to be less than 2.5° for gait and less than 4.7° for running across all joints and planes of motion. Higher variability for running may be due to movement speed and the effects of blurring. The one study [35] that did investigate the effects of running on ML tracking failed to report on camera settings (e.g. shutter speed), so the effects of these parameters are unclear.

Using the same method of Schwartz et al. [37], Kanko et al. [6] compared reliability measures derived from Theia3D with those reported in the literature using MB approaches. They found that inter-trial variability was slightly higher when using Theia3D and, on average, inter-session variability was lower for Theia3D. Newer versions of Theia3D have shown improvements in overall reliability due to larger training dataset [8]. This underlines the importance of documenting the version of Theia3D within the study protocol, as newer versions will potentially incorporate increased numbers of anatomical landmarks to track movement, be built around larger datasets, use improved tracking algorithms, and adopt automated calibration routines [8]. With the rapid advancement in ML technology, some researchers [6,8] suggest that there will inevitably come a point when minimal gains will be achieved through advancements in imagine detection and recognition. Until such a time, the recommendations appear to be one of keeping known datasets and reprocessing data with newer versions of the software, when appropriate.

Spatiotemporal gait parameters derived from Theia3D have been compared to several PSW, including GAITRite and Stepscan, as well as MB systems. When comparing Theia3D to PSW, some gait parameters (e.g. stride width) have shown variable levels of agreement across studies

$(0.44 \leq ICC \leq 0.86)$ [5,28], which may be due to differences in how the foot is defined by the two approaches. Theia3D uses ankle joint positions [28], whereas, PSW typically use the entire foot. Differences in toe-off times have also been reported, which may be linked to differences in sampling frequency [5,28] and the pressure threshold sensitivities of PSW [28]. However, when comparing Theia3D with MB approach, Kanko et al. [5] found excellent agreement $(ICC \geq 0.93)$ for stride width, highlighting how the selection of the comparison tool can influence outcome measures.

One study [28] investigated the feasibility of using Theia3D in environments other than the laboratory (e.g. parks, community and medical centres). McGuirk et al. [28] captured spatiotemporal data (cadence, speed, step length, step time, stride length, and stride time) from 166 participants in six different locations using Theia3D and a PSW. Theia3D was considered to demonstrate excellent feasibility when assessed according to implementation (96.7 %), practicality (90.5 %), and expansion (100 %). As a result, Theia3D was considered to provide significant advantages in conducting 3D gait analysis in community and clinical environments, offering the potential for improved patient assessment and rehabilitation.

This review has its limitations, most notably due to: the small number of high-quality studies; the heterogeneity in biomechanical models used across studies, the commercial opacity of Theia3D model, the limited number of studies involving participants with pathological conditions, and the limited number of kinetic validation studies. Differences in the types of movements examined, inconsistencies in kinematic and kinetic outcome measures, and variability in the statistical methods used across studies inhibited the synthesis of evidence. Only five studies investigated the reliability of Theia3D and approximately a third were considered to involve low participant numbers $(n < 21)$, with most failing to present information on sample size estimates, leading to 'serious' imprecision. Currently, limited data exists on the optimum setup for ML motion capture, with few studies documenting detailed information to inform the assessment of a study's content validity. The quality appraisal criteria employed in this review to assess content validity was developed from recommendations published in several sources [13,15,16,20]. Future consideration should be given to the standardisation and documentation of necessary conditions for optimum ML tracking. None of the included studies appeared to normalise or correct for differences in biomechanical model definitions, which may have been a source of error, particularly for body segments like the pelvis whose local co-ordinate system can vary significantly according to the adopted model.

As the transparency of AI methodology was largely absent across studies (Appendix E), software developers should be encouraged to provide essential information relating to the training datasets (e.g. population, movements, landmarks), training methods and model structures. If AI-based motion capture systems are to gain acceptance as future technology for analysing human movement, it is important that the design, training data, and validation processes of AI systems are suitability described. Without such transparency, it remains difficult to assess the accuracy, validity, and reliability of its application for studying human movement. When reporting on ML motion capture, future studies should provide a more transparent account of methodological approaches that may influence ML tracking (e.g. camera setup, participant's clothing, and the quality and variability of lighting conditions), adopt appropriate and standardised measures of accuracy, validity and reliability, and report on important AI-specific system characteristics. To understand the potential benefits and limitation of ML tracking for clinical assessment and rehabilitation, more rigorous studies focusing on populations with pathological conditions are required.

5. Conclusions

Theia3D ML motion tracking demonstrates reasonable accuracy compared to MB approaches in measuring lower limb joint angles during functional tasks such as gait, running, and jumping, except for hip flexion/extension and transverse plane rotations, which showed poor agreement. Most spatiotemporal gait parameters (gait speed, step length, and stride length) show excellent reliability and validity, whilst stride width, swing time, and double limb support show poor validity. Theia3D shows reasonable inter-session reliability for lower limb gait kinematics in the frontal and sagittal planes. Overall, the quality of evidence was considered low due to the small number of studies, the heterogeneity of outcome measures reported and the inability to pool data across studies. The biomechanical models adopted by ML and MB motion capture systems need to be standardised to enable more meaningful comparisons between approaches. More high-quality studies are required to validate ML systems for clinical kinematic and kinetic assessment in paediatric patients. Video based data capture and the application of Theia3D to study human movement shows promise as a portable, time-saving, and non-intrusive tool for use in a wide range of clinical, sporting and health-related settings.

CRediT authorship contribution statement

Florent Varcin: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation. **Mark G. Boocock:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

They have no financial and personal relationships with other people or organizations that could inappropriately influence (bias) this work.

Acknowledgements

The valuable assistance and guidance of AUT's Information and Education Services, particularly Andrew South (Senior Liaison Librarian), is greatly appreciated.

Appendix A

An example of the search strategy used in the EBSCO database
Search 1 (Theme – markerless)
markerless OR marker-less OR "Marker Free" OR Theia OR Theia3D
AND (motion OR mobility OR movement) PR(6) (capture OR track* OR analys*)
AND
Search 2 (Theme – accuracy, validity, reliability)

validity OR reliability OR valid* OR reliable OR compare OR comparison OR accuracy OR accurate OR feasibility OR correlation OR correlate OR viability

NB: slight modifications were made to the search strategy depending on the database searched.

Databases searched included: EBSCO (Medline + CINAHL + SportDiscus) + SCOPUS + OVID(AMED) + ProQuest + Prospero + Google Scholar (first 150).

PR(6) — proximity operator used to restrict words within a distance of six from each other.

Appendix B

Summary of participant characteristics, methods and main findings from each included study

Authors	Sample size and characteristics (n)	Approaches compared	Construct	Materials	Methods	Statistical methods	Main results
Gait							
Chaumeil et al. [7]	Healthy (16) Female (7) Male (9)	Theia3D QTM	**Validity** **Accuracy** CoM position XCoM position WBAM	10 Qualysis Miqus M3 (ML) (60 Hz) 10 Qualysis Miqus Video (ML) Synchronise (60 Hz)	Self-selected walking speed on a treadmill (7–20 cycles)	Bland Alman & LOA RMSD Pearson Coeff	**CoM (AP/ML/SI)** RMSD (cm) = 0.66/0.28/1.31 LOA: ≤2 cm | R^2 = 0.98–0.99 **XCoM (AP/ML/SI)** RMSD = 0.76/0.46/1.36 cm LOA: ≤2 cm | R^2 = 0.97–0.99 WBAM (% of the amplitude) RMSD (%) = 3.16/3.87/7.58 R^2 = 0.72–0.99
Kanko et al. [5]	Healthy (30) Female (15) Male (15)	Theia3D QTM	**Validity** Spatio-temporal gait parameter	8 Qualysis Miqus (ML) 7 Qualysis 3+ (MB) Synchronise (85 Hz)	Self-selected walking speed on a treadmill	Bland-Altman methods (LOA) Mean and standard deviation Pearson correlation coeff & ICC (A-1)	**ML compared to MB (mean)** Gait speed diff: 0.00m/s | ICC>0.9 Step & Stride length diff: ≤1 cm ICC > 0.9 | LOA: ±7 cm Stride width diff: ≤1 cm | ICC>0.9 LOA: ±2 cm Time based parameters diff: 0.01s Swing time: ML MB | ICC<0.75 Double limb support time: -0.03s ML > MB | ICC<0.75
		Theia3D GAITRite		8 Sony RX0 II (ML) GAITRite mat (60 Hz)	Self-selected walking speed on an over-ground walkway		**ML compared to Gait mat (mean)** Gait speed diff: 0.02m/s | ICC>0.9 Step & Stride length diff: ≤1 cm 0.75<ICC<0.9 | LOA: ±5 cm Stride width diff: ≤−3.66 cm ICC<0.75 | LOA: −8 cm to 1 cm Step time diff: −0.01 s Double limb support time: 0.07 s ML < GR | ICC<0.5 ML cycle & stance time > GR (LOA) | ICC>0.9
Kanko et al. [6]	Healthy (8) Female (2) Male (6)	Theia3D	**Reliability** Kinematics	8 Sony RX0 II (ML) Synchronise (60 Hz)	Self-selected walking speed on an over-ground walkway	Mean & SD Schwartz et al. [37] method	**Schwartz et al. [37] method:** Inter-trial var <5° Inter session var = 2.8° (average) Variability ratios (1.1)

(*continued on next page*)

(*continued*)

Authors	Sample size and characteristics (n)	Approaches compared	Construct	Materials	Methods	Statistical methods	Main results
Kanko et al. [4]	Healthy (30) Female (15) Male (15)	Theia3D QTM	**Accuracy** Kinematics	8 Qualysis Miqus (ML) 7 Qualysis 3+ (MB) Synchronise (60 Hz)	Self-selected walking speed on a treadmill	RMSD Mean & SD (ML–MB)	**Lower limb joint centre position** Diff ≤3 cm (except for the hip) **Lower limb joint angles (°)** Hip (SP/AP/TP) RMSD = 11/2.6/6.9 Knee (SP/AP/TP) RMSD = 3.3/3.2/13.2 Ankle (SP/AP/TP) RMSD = 6.7/8/11.6
McGuirk et al. [28]	Participants (166)	Theia3D	**Feasibility** 'Outside the lab'	Portable stadiometer digital scale measuring tape Research Electronic Data Capture (REDCap)	Areas of focus: Implementation Practicality Expansion 3-month period Six locations based on space availability	Guidelines outlined by Bowen et al. (2009)	**Success rate:** 96.7% (29/30) **Mobility and data management:** 100% **Practicality:** 90.5% (38/42) Data processing efficiency: 83.3%
	Subset (46) Female (32) Male (14)	Theia3D PSW	**Validity** Spatio-temporal gait parameters	8 video cameras (ML) Synchronised (60 Hz) 8.4 m long PSW (100 Hz)	Six locations self-selected (SS) and fastest comfortable six locations based on space availability (FC) walking Minimum of 12 full gait cycles per leg for each task	MDs LOA ICC (2,1) ICC (3,1)	**Mean diff (SS/FC):** Cadence: ≤0.29 steps/min ICC > 0.9 | LOA: ±2 steps/min Speed diff (SS/FC): −0.01/0.00m/s LOA: ±0.03m/s Step & stride length: −0.6/−0.2 cm ICC > 0.9 | LOA: ±2.0 cm Stride width: −1.2/−1.8 cm ICC >0.75 | LOA: SS: [−5.2;2.9] FC: [−5.3;2.9] **Temporal measures diff**: 0.02 s
Outerleys et al. [8]	Knee osteoarthritis (10) Female (5) Male (5)	Theia3D	**Reliability** Lower body joints kinematic	8 Sony RX0 II (ML)	**1st trial:** Self-selected walking speed **2nd trial:** Self-selected faster walking speed	Mean & SDs ANOVAs ICC SEM	Temporal distance metric highly repeatable **Kinematics** Inter-trial var = 1.2° Inter-session var = 1.5°
Riazati et al. [9]	Participants (age and health characteristics unknown) (21) Female (14) Male (7)	Theia3D	**Reliability** Lower extremity kinematics Spatiotemporal parameters	8 video cameras (ML) (Basler ace acA1300-75gc GigE) (60 Hz)	**1st task:** Self-selected walking speed **2nd task:** Fastest comfortable walking speed	RMSD SEM SEM%	**Kinematics** SEM and RMSD ≤ 5° (majority) Highest RMSD = Knee (TP) Lowest RMSD = Hip (FP) **Spatiotemporal parameters** 2.6% ≤ SEM% ≤ 16.8% (selected speed) 1.9% ≤ SEM% ≤ 12% (fast)
Song et al. [31]	Healthy adults (10) Female (5) Male (5)	Theia3D Motion Analysis Corp.	**Validity** **Accuracy** Lower extremities kinematics and kinetics	12 infrared cameras (Raptor-E × 10) (MB) 8 video cameras (ML) Synchronise (100 Hz)	Over-ground walking	RMSD Pearson coeff	**Kinematics:** Hip (SP/FP/TP): R^2 = 0.99/0.74/−0.2 RMSD = 10.6°/3.0°/13.6° Knee (SP): R^2 = 0.96 | RMSD = 3.9° Ankle (SP): R^2 = 0.97 | RMSD = 3.4°

(*continued on next page*)

(*continued*)

Authors	Sample size and characteristics (n)	Approaches compared	Construct	Materials	Methods	Statistical methods	Main results
							Kinetics: $R^2 = 0.94–0.99$
Wishaupt et al. [29]	Children (45) TD (15) Female (9) Male (6) CP (30)	Theia3D VICON (HBM)	**Accuracy** Clinical Kinematics	7 Blackfly cameras (ML) 12 Vicon cameras (MB) Synchronise (100 Hz)	Comfortable walking speed along the walkway	SPM paired *t*-test RMSD SDs MAD	**TD children*** Trunk (SP) = RMSD = 9.2° Pelvis (SP) = RMSD = 8.9° Hip (SP) = RMSD = 6.8° Knee (FP) = RMSD = 5.1° **CP children*** Trunk (SP) = RMSD = 11.8° Pelvis (SP) = RMSD = 9.1° Hip (SP) = RMSD = 7.7° Knee (FP) = RMSD = 5.6° Knee (TP) = RMSD = 10.2° Ankle (SP) = RMSD = 6.4°
Wren et al. [34]	Paediatric patients (36) Female (16) Male (20) Assistance waling (6)	Theia3D Vicon	**Validity** **Accuracy** Kinematics	10 infrared cameras (T-40S) (MB) (120 Hz) 8 video cameras Miqus (ML) (120 Hz)	Walking trials	Mean % SDs RMSD RMSDoffset	**Kinematics (SP/FP/TP) (mean)** Trunk: RMSD = 5.1°/3.1°/4.1° Pelvis: RMSD = 11°/2.9°/5.2° Hip: RMSD = 11.9°/5.0°/10° Knee: RMSD = 4.8°/4.7°/9.0° Ankle: RMSD = 4.9°/11.2°/19.3° Foot progression (TP): RMSD = 7.8°
Running							
Kanko et al. [30]	Healthy (30) Female (15) Male (15)	Theia3D Qualysis	**Validity** **Accuracy** Kinematics Kinetics	7 Qualisys Oqus (MB) 8 Qualysis Mqus (ML) Synchronise (85 Hz)	Self-selected running speed on a treadmill	RMSD	**Kinematics:** Sagittal plane (ankle/knee/hip) RMSD = 3.8°/4.2°/5.2° Frontal plane: RMSD = 6°/7.2°/2.7° Transverse plane: RMSD = 8°/9.4°/9.8° **Kinetics:** $0.03 \leq$ RMSD (Nm/kg) ≤ 0.29
Moran et al. [35]	Healthy (21) Female (2) Male (7)	Theia3D	**Repeatability** Kinematics Gait events	8 Sony RX0 II (ML) (120 Hz)	Treadmill running. 3 trials with 3 different self-selected speeds	SDs Schwartz's method ICC SEM MDC	**Kinematics:** Intra-trial var: ≤3° Inter-session var: ≤intra-trial var Variability ratio: ≤1 **Discrete metrics** Heel Strike: $0.43° \leq$ SEM $\leq 1.31°$ At toe-off: SEM $\leq 0.4°$ Across all planes, good agreement ($0.77 <$ ICC < 0.85) (on average)
Song et al. [31]	Healthy (10) Female (5) Male (5)	Theia3D Motion Analysis Corp.	**Validity** **Accuracy** Lower extremities kinematics and kinetics	12 IR cameras (Raptor-E × 10) (MB) 8 video cameras (ML)	Over-ground running	RMSD Pearson coeff	**Kinematics:** Hip (SP/FP/TP): $R^2 = 0.99/0.86/0.02$ RMSD = 13.8°/4.5°/12.8° Knee (SP): $R^2 =$

(*continued on next page*)

(*continued*)

Authors	Sample size and characteristics (n)	Approaches compared	Construct	Materials	Methods	Statistical methods	Main results
				Synchronise (100 Hz)			0.95 | RMSD = 4.8° Ankle (SP): R^2 = 0.94 | RMSD = 5.1° **Kinetics:** R^2 = (0.81–0.99)
					Run and cut maneuver		**Kinematics:** Hip (SP/FP/TP): R^2 = 0.99/0.79/0.6 RMSD (°) = 13.1/5.1/15.9 Knee (SP): R^2 = 0.87 | RMSD = 5.9° Ankle (SP): R^2 = 0.9 | RMSD = 5.7° **Kinetics:** R^2 = (0.83–0.99)
Tang et al. [10]	Healthy student (16)		**Validity** **Accuracy** Effect of speed Lower extremity joint moment and power	8 IR cameras (Vicon Bonita) (MB) 8 video cameras (Vicon Vue) (ML) Synchronise (100 Hz)	3 running trials at 2.24 m/s, 2.91 m/s, and 3.58 m/s.	Mean & SDs Kolmogorov-Smirnov test SPM	**Joints moments and power:** ML > MB (increased with speed) Knee and hip joints (swing phase) Ankle plantarflexion (stance phase)
Tang et al. [33]	Healthy (16) Female (7) Male (9)	Theia3D Vicon	**Validity** **Accuracy** Inverse dynamic Lower extremity joint moment and power Joint angles	8 IR camera (Vicon Bonita) (MB) 8 video cameras (Vicon Vue) (ML) Synchronise (100 Hz)	Run on the treadmill for 120 s at 3.58 m/s	Mean & SDs Kolmogorov-Smirnov test Two-tailed paired *t*-test (MB vs ML)	**Sagittal plane:** Hip, knee angles: MB > ML Ankle angles: MB < ML **Joints moments and power:** ML ≠ MB
Jumping							
Chaumeil et al. [7]	Healthy (16) Female (7) Male (9)	Theia3D QTM	**Validity** **Accuracy** CoM position XCoM position WBAM	10 Qualysis Miqus M3 (ML) (60 Hz) 10 Qualysis Miqus Video (ML) Synchronise (60 Hz)	Counter movement jumps (3 repetitions)	Bland Alman & LOA RMSD Pearson Coeff	**CoM (AP/ML/SI)** RMSD = 0.99/0.39/1.61 cm LOA: ≤2 cm | R^2 = 0.98–0.99 **XCoM (AP/ML/SI)** RMSD (cm) = 1.37/0.38/2.09 LOA: >2 cm | R^2 = 0.97–0.99 WBAM (% of the amplitude) RMSD (%) = 3.16/3.87/7.58 R^2 = 0.63(SI); 0.65 (AP)
Song et al. [31]	Healthy (10) Female (5) Male (5)	Theia3D Motion Analysis Corp.	**Validity** **Accuracy** Lower extremities kinematics and kinetics	12 IR cameras (Raptor-E × 10) (MB) 8 video cameras (ML) Synchronise (100 Hz)	Counter movement jumps (no details)	RMSD Pearson Coeff	**Kinematics:** Hip (SP/FP/TP): R^2 = 0.99/0.8/0.09 RMSD = 12.1°/2.3°/11° Knee (SP): R^2 = 0.99 | RMSD = 3.4° Ankle (SP): R^2 = 0.99 | RMSD = 5.3° **Kinetics:** ML = MB
Strutzenberger et al. [32]	Healthy (23) Female (12) Male (15)	Theia3D Qualisys	**Validity** **Accuracy** Kinematics Limb joint positions Jump parameters	7 Qualisys 3+ (MB) 8 Qualisys Mqus (ML) Synchronise (85 Hz)	3 maximal effort countermovement jumps	RMSD ICC SPM	**Jump heights** ICC > 0.997 **Kinematics:** (Sagittal plane) Ankle: RMSD = 4.96° (1.85) ICC = 0.97 Knee: RMSD = 4.44° (1.2) ICC = 0.99 Hip: RMSD =

(*continued on next page*)

(*continued*)

Authors	Sample size and characteristics (n)	Approaches compared	Construct	Materials	Methods	Statistical methods	Main results
							20.6° (5.22) ICC = 0.8
Other activities							
Chaumeil et al. [7]	Healthy (16) Female (7) Male (9)	Theia3D QTM	**Validity** **Accuracy** CoM position XCoM position WBAM	10 Qualysis Miqus M3 (ML) (60 Hz) 10 Qualysis Miqus Video (ML) Synchronise (60 Hz)	Balance loss: loss of balance requiring one or more recovery steps, 3 repetitions. Walking on a 2 m long, 2.3 cm wide and 5 cm high beam (4 reps)	Bland Alman & LOA RMSD Pearson Coeff	**CoM (AP/ML/SI)** RMSD (cm) = 0.65/0.40/1.29 LOA: ≤ 1.39 cm $\mid R^2 = 0.99$ **XCoM (AP/ML/SI)** RMSD = 0.73/0.52/1.35 cm LOA: ≤ 1.42 cm $\mid R^2 = 0.98–0.99$ **WBAM** (% of the amplitude) RMSD (%) = 2.52/2.38/5.69 $R^2 = 0.87–0.99$
Hansen et al. [36]	Children (21) Male (12) Female (9) 7 yr (8) 9 yr (5) 11 yr (6)	Theia3D Qualysis	**Validity** **Reliability** **Accuracy** Upper limbs kinematics	10 Oqus cameras (170 Hz) (MB) 8 Miqus cameras (85 Hz) (ML)	Box and Blocks test	Means & SD ICCs Bland Altman (LOA) RMSD	**Validity:** $0.75 < ICC < 1$ Except for peak wrist FLEX & RD **Accuracy:** RMSD $\leq 6°$ across all joints **Inter-session reliability:** ML: $0.14 < ICC < 0.8$ MB: $0.16 < ICC < 0.82$
Song et al. [31]	Healthy (10) Female (5) Male (5)	Theia3D Motion Analysis Corp.	**Validity** **Accuracy** Lower extremities kinematics	12 IR cameras (Raptor-E × 10) (MB) 8 video cameras (ML) Synchronise (100 Hz)	Double-leg squat	RMSD Pearson Coeff	**Kinematics:** Hip (SP/FP/TP): $R^2 = 0.9/0.89/0.5$ RMSD = 12.1°/2.8°/8.2° Knee (SP): $R^2 = 0.99 \mid$ RMSD = 2.9° Ankle (SP): $R^2 = 0.98 \mid$ RMSD = 5.5°
					Step down: Stepping down a 20 cm box.		**Kinematics:** Hip (SP/FP/TP): $R^2 = 0.97/0.86/0.4$ RMSD = 8.7°/2.5°/10. 3° Knee (SP): $R^2 = 0.99 \mid$ RMSD = 2.6° Ankle (SP): $R^2 = 0.99 \mid$ RMSD = 4.3°

Note: Information extracted from the sixteen studies included in this review. Data extracted is based on the COSMIN guidelines [15]. Categories separated by a dashed line belong to the same study but represent distinct outcome measures. All kinematic measures are reported in degrees (°) unless otherwise specified.

Abbreviations: QTM: Qualisys Track Manager; ML: markerless; MB: marker-based; IR: infrared; PSW: pressure-sensing walkway; LOA: limits of agreement; RMSD: root mean square difference; MAD: mean absolute deviation. ICC: intraclass correlation coefficient; SEM: standard error of measurement; MDC: minimal detectable change; SPM: statistical parametric mapping; CoM: centre of mass; XCoM: extrapolated center of mass; SP: sagittal plane; FP: frontal plane; TP: transverse plane; WBAM: whole-body angular momentum; SD: standard deviation; AFO: ankle-foot orthosis; TD: typically developing; CP: cerebral palsy.

[*] Only data with significant differences based on SPM analysis across the full gait cycle were included in the table.

Appendix C

COSMIN (Boxes 2a and 2b) quality assessment criteria used to evaluate content validity (adapted from COSMIN [15,17]).

Box 2a
Motion capture recommendations (adapted from Payton and Bartlett [20] and Theia Markerless Inc. [19]).

Criteria and rating scores	Very good	Adequate	Doubtful	Inadequate	N/A — not mentioned
1. Were the number of cameras appropriate? *$\geq$8 cameras = very good* *7 or 6 cameras = adequate* *$<$6 cameras $>$ 3 = doubtful* *2 $\geq$ cameras = inadequate*					
2. Were the cameras synchronised? 'The cameras must capture synchronous videos with identical start times and durations to be used with Theia3D.' *If mentioned = very good* *If evidence that there is no synchronisation = inadequate*					
3. Was the frame rate appropriate for the movement being captured? 25–50 Hz — walking, swimming, stair climbing. 50–100 Hz — running, shot put, high jump. 100–200 Hz — sprinting, javelin throwing, football kick. 200–500 Hz — tennis serves, golf swing, fencing. *Between the range = very good* *Lower limit of a range = adequate* *Under lower limit = doubtful* *In another smaller range = inadequate*					
4. Was the shutter speed (SP) appropriate? General rule: 1 / (2 × frame rate, or higher) seconds *If SP $\geq$ 1 / (2 × frame rate) = very good* *If SP $<$ 1 / (2 × frame rate) = doubtful*					
5. Was the position of the cameras relative to movement being captured appropriate? – Cameras were as close as possible to the capture volume, while ensuring the entire capture volume was within view for all cameras. – Cameras were between 1.2 to 2.5m above the ground. – Cameras were not restricted to partial views of participants, such as lower body or upper body only. – No unusual camera views, especially very high or very low. – Camera setup was symmetrical and surrounded the entire *capture volume, such as a circle, oval, or rectangle.* *All criteria fully satisfied and well-documented = very good* *All criteria satisfied with supporting evidence (figure or video) = adequate* *Some criteria satisfied and partially documented = doubtful* *No criteria satisfied = inadequate*					
6. Was lighting of the capture volume appropriate? *Indoor/lighting stable = very good* *Outdoor/potentially variable lightening = adequate* *Indoor, but lighting inconsistent or partially obstructed = doubtful* *Lighting conditions uncontrolled or highly variable = inadequate*					
7. Participant(s) clothing a. Were the participants wearing fitted clothing? b. Was each limb discernible from the rest of the body? c. Was there good contrast between the participants clothing and the background? *a + b + c = very good* *a + b = very good, if the Q5 Box2 is very good, otherwise use the same rating as Q5 Box2* *If only c = doubtful* *If none = inadequate*					
8. Were the study measurement/outcomes appropriate for the software?					

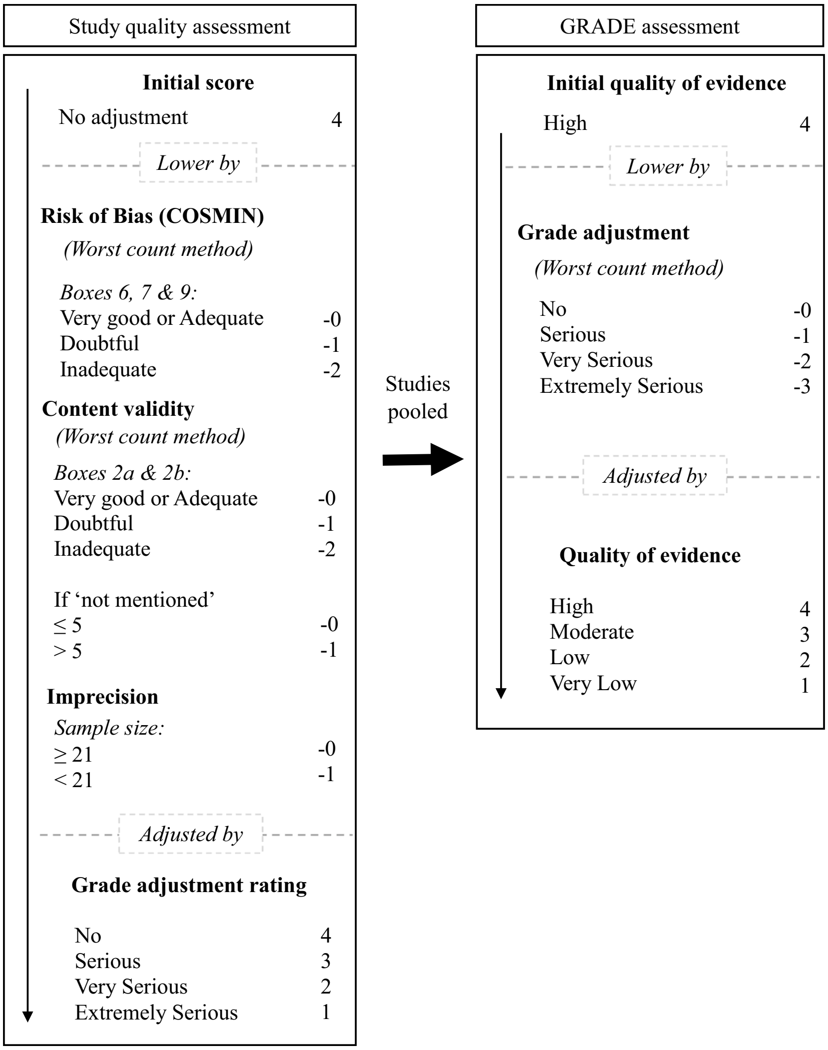
Box 2b

AI system risk of bias and methodological transparency (adapted from CONSORT-AI and PROBAST-AI guidelines [21,22].

Criteria and rating scores	Very good	Adequate	Doubtful	Inadequate	N/A — not mentioned
1. Transparency of the AI model architecture *AI model type, architecture and training methods (e.g. deep-learning, layers, supervised/unsupervised) clearly described = very good* *Model type, architecture and training methods briefly mentioned = adequate* *Unspecified AI model and training methods = doubtful* *Not an AI-based model = inadequate*					
2. Training data documentation *Training dataset origin, size, and diversity (age, gender, ethnicity, movement types) clearly described = very good* *Minimal description of a human-based dataset = adequate* *Unclear as to the human-based dataset used = doubtful* *No mention of a training dataset = inadequate*					
3. Interpretation and explanation *Well described AI estimation processes and outputs (e.g. landmark confidence) = very good* *Some features of AI decision process and outputs described = adequate* *Limited discussion of AI decision process and outputs = doubtful* *Model decisions not interpreted = inadequate*					
4. Study population with respect to AI training data *Close match between study population and the AI training dataset = very good* *Partial match between the study population and the AI training dataset = adequate* *Poor match between study population and the AI training dataset = doubtful* *No mention or inappropriate match between study population and AI training dataset = inadequate*					
5. Movements studied relative to AI training data *Close match between movements studied and those of the AI dataset = very good* *Adequate match between movements studied and those of the AI dataset = adequate* *Limited description of movements or clear mismatch to those of the AI dataset = doubtful* *Movements studied differ markedly from those of the AI training dataset = inadequate*					
6. Was the version of AI system reported? *Comprehensive details of the version number = very good* *Version number provided but missing some information = adequate* *Beta or unofficial version = inadequate* *No reference to the version of the software = not mentioned*					

Appendix D

An overview of the workflow used to assess the quality of each study and the GRADE assessment approach for determining overall evidence.



Appendix E

COSMIN quality assessment ratings for each of the included studies.

Authors	Accuracy (Box 7)						Validity (Box 9)				Reliability (Box 6)						
	2	3	4	5	6	7	1	2	3	4	1	2	3	4	5	6	7
Chaumeil et al. [7]	VG	VG	NM	NM	VG	VG	VG	VG	VG	VG							
Hansen et al. [36]	VG	D	NM	NM	VG	VG	VG	VG	VG	VG	VG	A	D	NM	NM	VG	VG
Kanko et al. [6]											A	VG	A	NM	NM	VG	VG
Kanko et al. [5]							VG	VG	VG	A							
Kanko et al. [4]	VG	A	NM	NM	VG	A											
Kanko et al. [30]	VG	VG	NM	NM	A	A											
McGuirk et al. [28]							VG	A	VG	A							
Moran et al. [35]	D	A	NM	NM	VG	VG					VG	A	A	NM	NM	D	VG
Outerleys et al. [8]											D	A	A	NM	NM	D	VG
Riazati et al. [9]											D	D	D	NM	NM	A	VG
Song et al. [31]	VG	VG	NM	NM	A	A	VG	A	VG	A							
Strutzenberger et al. [32]	VG	A	NM	NM	D	VG	VG	A	VG	A							
Tang et al. [33]	VG	A	NM	NM	A	A	VG	A	A	VG							
Tang et al. [10]	VG	A	NM	NM	A	VG	VG	VG	VG	A							
Wishaupt et al. [29]	VG	VG	NM	NM	I	VG											

(*continued on next page*)

(*continued*)

Authors	Accuracy (Box 7)						Validity (Box 9)				Reliability (Box 6)						
	2	3	4	5	6	7	1	2	3	4	1	2	3	4	5	6	7
Wren et al. [34]	VG	A	NM	NM	I	A	VG	VG	VG	I							

Authors	Box 2a. Motion capture recommendations								Box 2b. AI system risk of bias and methodological transparency						Overall assessment*
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	
Chaumeil et al. [7]	VG	VG	VG	NM	VG	VG	VG	VG	A	NM	NM	NM	NM	VG	A
Hansen et al. [36]	VG	A	VG	NM	A	VG	NM	VG	A	A	NM	NM	NM	NM	D
Kanko et al. [6]	VG	VG	VG	NM	NM	VG	A	VG	A	A	NM	NM	NM	NM	D
Kanko et al. [5]	VG	VG	VG	NM	A	VG	VG	VG	A	A	NM	NM	NM	VG	A
Kanko et al. [4]	VG	VG	VG	NM	A	VG	VG	VG	A	A	NM	NM	NM	VG	A
Kanko et al. [30]	VG	VG	VG	NM	A	VG	VG	VG	A	NM	NM	NM	NM	VG	A
McGuirk et al. [28]	VG	VG	VG	NM	A	A	A	VG	A	NM	NM	NM	NM	VG	A
Moran et al. [35]	VG	VG	VG	NM	A	NM	A	VG	D	NM	NM	NM	NM	NM	I
Outerleys et al. [8]	VG	VG	VG	VG	VG	VG	A	VG	A	NM	NM	NM	NM	VG	A
Riazati et al. [9]	VG	VG	VG	NM	A	A	A	VG	A	NM	NM	NM	NM	VG	A
Song et al. [31]	VG	VG	VG	A	VG	VG	VG	VG	A	A	NM	NM	NM	VG	A
Strutzenberger et al. [32]	VG	VG	VG	NM	NM	NM	NM	VG	A	A	NM	NM	NM	VG	D
Tang et al. [33]	VG	VG	VG	NM	D	VG	VG	VG	A	NM	NM	NM	NM	VG	D
Tang et al. [10]	VG	VG	VG	NM	NM	VG	VG	VG	A	NM	NM	NM	NM	VG	D
Wishaupt et al. [29]	A	VG	VG	NM	A	VG	VG	VG	NM	NM	NM	NM	NM	VG	D
Wren et al. [34]	VG	A	NM	NM	I	A	VG	VG	NM	NM	NM	NM	NM	VG	I

Abbreviations: VG: very good; A: adequate; D: doubtful; I: inadequate; NM: not mentioned.

[*] Worst score and downgraded by 1 if NM > 5.

References

[1] Mathis A, Schneider S, Lauer J, Mathis MW. A primer on motion capture with deep learning: principles, pitfalls, and perspectives. Neuron Oct 14 2020;108(1):44–65. https://doi.org/10.1016/j.neuron.2020.09.017.

[2] Roggio F, Trovato B, Sortino M, Musumeci G. A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: a narrative review. Heliyon Nov 15 2024;10(21):e39977. https://doi.org/10.1016/j.heliyon.2024.e39977.

[3] Wade L, Needham L, McGuigan P, Bilzon J. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. PeerJ 2022;10:e12995. https://doi.org/10.7717/peerj.12995.

[4] Kanko RM, Laende EK, Davis EM, Selbie WS, Deluzio KJ. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. J Biomech 2021;127:110665. https://doi.org/10.1016/j.jbiomech.2021.110665.

[5] Kanko RM, Laende EK, Strutzenberger G, Brown M, Selbie WS, DePaul V, et al. Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. J Biomech 2021;122:110414. https://doi.org/10.1016/j.jbiomech.2021.110414.

[6] Kanko RM, Laende E, Selbie WS, Deluzio KJ. Inter-session repeatability of markerless motion capture gait kinematics. J Biomech May 24 2021;121:110422. https://doi.org/10.1016/j.jbiomech.2021.110422.

[7] Chaumeil A, Lahkar BK, Dumas R, Muller A, Robert T. Agreement between a markerless and a marker-based motion capture systems for balance related quantities. J Biomech Feb 21 2024;165:112018. https://doi.org/10.1016/j.jbiomech.2024.112018.

[8] Outerleys J, Mihic A, Keller V, Laende E, Deluzio K. Markerless motion capture provides repeatable gait outcomes in patients with knee osteoarthritis. J Biomech May 2024;168:112115. https://doi.org/10.1016/j.jbiomech.2024.112115.

[9] Riazati S, McGuirk TE, Perry ES, Sihanath WB, Patten C. Absolute reliability of gait parameters acquired with markerless motion capture in living domains. Front Hum Neurosci 2022;16:867474. https://doi.org/10.3389/fnhum.2022.867474.

[10] Tang H, Munkasy B, Li L. Differences between lower extremity joint running kinetics captured by marker-based and markerless systems were speed dependent. J Sport Health Sci 2024. https://doi.org/10.1016/j.jshs.2024.01.002.

[11] Morgan RL, Whaley P, Thayer KA, Schunemann HJ. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environ Int Dec 2018;121(Pt 1):1027–31. https://doi.org/10.1016/j.envint.2018.07.015.

[12] Luciano F, Ruggiero L, Pavei G. Sample size estimation in locomotion kinematics and electromyography for statistical parametric mapping. J Biomech Jun 9 2021;122:110481. https://doi.org/10.1016/j.jbiomech.2021.110481.

[13] Theia Markerless Inc. About Theia Markerless, Inc.. Accessed 27/03/24, https://www.theiamarkerless.ca/about/.

[14] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021;18(3):e1003583. https://doi.org/10.1371/journal.pmed.1003583.

[15] Mokkink LB, Boers M, van der Vleuten C, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020. https://doi.org/10.1186/s12874-020-01179-5.

[16] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res May 2018;27(5):1171–9. https://doi.org/10.1007/s11136-017-1765-4.

[17] Prinsen CAC, Mokkink LB, Patrick DL, Alonso J, Bouter LM, de Vet HCW, et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res May 2018;27(5):1147–57. https://doi.org/10.1007/s11136-018-1798-3.

[18] Scataglini S, Abts E, Van Bocxlaer C, Van den Bussche M, Meletani S, Truijen S. Accuracy, validity, and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: a systematic review and meta-analysis. Sensors Jun 6 2024;24(11). https://doi.org/10.3390/s24113686.

[19] Theia Markerless Inc. Camera system requirements. Accessed 03/04/2024, https://www.theiamarkerless.ca/docs/cameras.html.

[20] Payton CJ, Bartlett RM. Biomechanical evaluation of movement in sport and exercise: the British Association of Sport and Exercise Sciences guide. J Sports Sci Med 2008:218.

[21] Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Spirit AI, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med Sep 2020;26(9):1364–74. https://doi.org/10.1038/s41591-020-1034-x.

[22] Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ Mar 24 2025;388:e082505. https://doi.org/10.1136/bmj-2024-082505.

[23] Knudson D. Twenty-year trends of authorship and sampling in applied biomechanics research. Percept Mot Skills Feb 2012;114(1):16–20. https://doi.org/10.2466/11.Pms.114.1.16-20.

[24] Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Pearson/Prentice Hall; 2009.

[25] Auer S, Suss F, Dendorfer S. Using markerless motion capture and musculoskeletal models: an evaluation of joint kinematics. Technol Health Care 2024;32(5):3433–42. https://doi.org/10.3233/THC-240202.

[26] McGinley JL, Baker R, Wolfe R, Morris ME. The reliability of three-dimensional kinematic gait measurements: a systematic review. Gait Posture Apr 2009;29(3):360–9. https://doi.org/10.1016/j.gaitpost.2008.09.003.

[27] GRADE. GRADE handbook — handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013.

[28] McGuirk TE, Perry ES, Sihanath WB, Riazati S, Patten C. Feasibility of markerless motion capture for three-dimensional gait assessment in community settings. Front Hum Neurosci 2022;16:867485. https://doi.org/10.3389/fnhum.2022.867485.

[29] Wishaupt K, Schallig W, van Dorst MH, Buizer AI, van der Krogt MM. The applicability of markerless motion capture for clinical gait analysis in children with cerebral palsy. Sci Rep May 24 2024;14(1):11910. https://doi.org/10.1038/s41598-024-62119-7.

[30] Kanko RM, Outerleys JB, Laende EK, Selbie WS, Deluzio KJ. Comparison of concurrent and asynchronous running kinematics and kinetics from marker-based

and markerless motion capture under varying clothing conditions. J Appl Biomech Jan 18 2024:1–9. https://doi.org/10.1123/jab.2023-0069.

[31] Song K, Hullfish TJ, Silva RS, Silbernagel KG, Baxter JR. Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements. J Biomech 2023;157:111751. https://doi.org/10.1016/j.jbiomech.2023.111751.

[32] Strutzenberger G, Kanko R, Selbie S, Schwameder H, Deluzio K. Assessment of kinematic CMJ data using a deep learning algorithm-based markerless motion capture system. ISBS Proc Arch 2021;39(1):236.

[33] Tang H, Pan J, Munkasy B, Duffy K, Li L. Comparison of lower extremity joint moment and power estimated by markerless and marker-based systems during treadmill running. Bioengineering 2022;9(10):574. https://doi.org/10.3390/bioengineering9100574.

[34] Wren TA, Isakov P, Rethlefsen SA. Comparison of kinematics between Theia markerless and conventional marker-based gait analysis in clinical patients. Gait Posture 2023. https://doi.org/10.1016/j.gaitpost.2023.05.029.

[35] Moran MF, Rogler IC, Wager JC. Inter-session repeatability of marker-less motion capture of treadmill running gait. Appl Sci 2023;13(3):1702. https://doi.org/10.3390/app13031702.

[36] Hansen RM, Arena SL, Queen RM. Validation of upper extremity kinematics using markerless motion capture. Biomed Eng Adv 2024;7. https://doi.org/10.1016/j.bea.2024.100128.

[37] Schwartz MH, Trost JP, Wervey RA. Measurement and management of errors in quantitative gait data. Gait Posture Oct 2004;20(2):196–203. https://doi.org/10.1016/j.gaitpost.2003.09.011.

[38] Chèze L. Biomécanique du mouvement et modélisation musculo-squelettique. Technol Bioméd 2014/06;2014. https://doi.org/10.51257/a-v1-med8050.

[39] Flux E, van der Krogt MM, Cappa P, Petrarca M, Desloovere K, Harlaar J. The human body model versus conventional gait models for kinematic gait analysis in children with cerebral palsy. Hum Mov Sci Apr 2020;70:102585. https://doi.org/10.1016/j.humov.2020.102585.

[40] Brainerd EL, Baier DB, Gatesy SM, Hedrick TL, Metzger KA, Gilbert SL, et al. X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. J Exp Zool Pt A Ecol Integr Physiol Jun 1 2010;313(5):262–79. https://doi.org/10.1002/jez.589.

[41] Ye D, Sun X, Zhang C, Zhang S, Zhang X, Wang S, et al. In vivo foot and ankle kinematics during activities measured by using a dual fluoroscopic imaging system: a narrative review. Front Bioeng Biotechnol 2021;9:693806. https://doi.org/10.3389/fbioe.2021.693806.

[42] Miranda DL, Rainbow MJ, Crisco JJ, Fleming BC. Kinematic differences between optical motion capture and biplanar videoradiography during a jump-cut maneuver. J Biomech Feb 1 2013;46(3):567–73. https://doi.org/10.1016/j.jbiomech.2012.09.023.

[43] Selbie S. April 13. https://www.youtube.com/watch?v=YfE5zBzi0Ts.

[44] Mihcin S. Simultaneous validation of wearable motion capture system for lower body applications: over single plane range of motion (ROM) and gait activities. Biomed Tech (Berl) Jun 27 2022;67(3):185–99. https://doi.org/10.1515/bmt-2021-0429.

[45] Mihcin S, Ciklacandir S, Kocak M, Tosun A. Wearable motion capture system evaluation for biomechanical studies for hip joints. J Biomech Eng Apr 1 2021;143(4). https://doi.org/10.1115/1.4049199.

[46] Vargas-Valencia LS, Elias A, Rocon E, Bastos-Filho T, Frizera A. An IMU-to-body alignment method applied to human gait analysis. Sensors (Basel) Dec 10 2016;16(12). https://doi.org/10.3390/s16122090.

[47] Keller VT, Outerleys JB, Kanko RM, Laende EK, Deluzio KJ. Clothing condition does not affect meaningful clinical interpretation in markerless motion capture. J Biomech 2022;141:111182. https://doi.org/10.1016/j.jbiomech.2022.111182.