Towards the use of 2-dimensional video-based markerless motion capture to objectively evaluate kinematics during functional capacity evaluation

by

Sarah Remedios

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Kinesiology

Waterloo, Ontario, Canada, 2020

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Background.

Workability, the capacity to perform work, is a concept highly regarded in a return-to-work context and typically measured using functional capacity evaluation (FCE). The aim of FCE is to determine the capacity of an individual such that they can be appropriately matched with job demands. Further, the usage of FCE is popularly used in attempt to proactively reduce the risk of injury in the workplace. Capacity is measured using tests, including maximum safe load tests and manual materials handling tolerance tests. However, the ability to predict workability using FCE has been questioned. Functional capacity evaluation administrators determine workability, in part, by using *a priori* parameters such as “the subject maintains balance," that are determined via subjective observations. Use of subjective observation may explain why the ability to reliably determine effort and capacity by using FCE remain in debate (Trippolini et al., 2014). However, a markerless motion-capture based solution may permit direct measurement of important movement features, and in turn, may improve the predictive utility and reliability of FCE outcomes. But, first, it remains important to evaluate if a 2-dimensional (2D) video-based markerless motion capture solution can generate objective outcomes that match with those generated using existing 3-dimensional (3D) motion capture.

Objective.

To determine the agreement of kinematic outputs calculated from motion data collected via a 2D video-based pose-estimation (markerless motion capture) software and a laboratory-based 3D motion capture for floor-to-waist height lifting task.

The kinematic outputs calculated include peak knee flexion angle, peak trunk flexion angle, peak shoulder flexion and abduction angles, functional stability limits in the anterior-posterior and medial-lateral direction, the distance of the load relative to the center of gravity and mean absolute relative phase angles.

Methods.

Three floor-to-waist height lifts were used for analysis for each participant (N = 20). Participants' lifts were captured using 3D motion capture (Vicon, Oxford, UK) and simultaneously recorded using 2D video (camcorder) in the sagittal plane. The participants lifts were each completed using a light, medium and heavy load dependent on the participants' individual subjective capacities. Post-collection, motion data from 3D motion capture and video-based markerless motion capture were used separately to calculate the specific kinematic metrics of interest. The outcome measures calculated were peak knee flexion angle, peak trunk flexion angle, peak shoulder flexion and abduction angles, center of gravity relative to the load handled, base of support relative to the center of gravity in the anterior-posterior and medial-lateral directions, as well as mean absolute relative phase angles of the hip-knee for the flexion and extension phases of the lift separately. Bland-

Altman analysis and plots were used to calculate agreement as a form of concurrent validity between the two methods.

Results.

For all outcome measures, Bland-Altman analysis did not suggest agreement between outcomes calculated using the 2D pose-estimation method and 3D motion capture method.

Conclusions.

Due to the lack of agreement between the two methods, it is advised that video-based markerless motion capture and 2D pose-estimation be further enhanced prior to use in calculating objective measures of FCE performance.

Acknowledgements

First, I would like to thank my supervisor, Dr. Steven Fischer, for his continued dedication towards advancing my research experience and skills. Thank your unconditional support of my research career at the University of Waterloo beginning during my undergraduate degree and continued into my graduate degree. Your endless guidance, motivation and curiosity has inspired me over the years.

Thank you to my supervisory committee, Dr. Monica Maly and Dr. Andrew Laing, for challenging my original research proposal and final defense document to better the project. Further, thank you for being adaptable to the change in my Thesis during the COVID-19 pandemic.

I would also like to acknowledge my past and current OBEL teammates for their help and guidance: Daniel Armstrong, Justin Davidson, Sheldon Hawley, Nathalie Oomen, Christopher Moore, and Laura Healey. Thank you for challenging my ideas and providing feedback during my practice presentations. Thank you to current and past members of the OBEL team who acted as mentors for different projects that ultimately led to acquisition of the skills required to complete this thesis: Daniel Armstrong, Claragh Pegg, Aleks Budarick, and Nathalie Oomen.

Thank you to all the volunteers and undergraduate research assistants who assisted in data collection and processing. Specifically, Andrea Disalvia and Nathania Chan for

collections and processing assistance. Your assistance was detrimental for completing this Thesis in a timely manner.

I would further like to thank the administration and tech support teams, specifically Denise Hay, Leanne Varey and Craig McDonald.

Lastly, I would especially like to thank my friends, volleyball teammates and family for all types of support and guidance during my Masters. Thank you to my head coach, Richard Eddy, for being flexible with the time constraints during my Master's degree and trusting me to lead the team while navigating graduate school. Thank you, Daniel Perigo, for your love, encouragement, and patience over the years. Very huge thank you to my parents (Joanne and Steve) and siblings (Jenna and Nicole) for all your unconditional love, motivation, and support. Thank you all and I could not have done it without you all.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Acronyms

2D	2-Dimensional	FSL	Functional stability limit
3D	3-Dimensional	IQR	Interquartile range
AP	Anterior-posterior	LOA	Limit of agreement
BOS	Base of support	MARP	Mean absolute relative phase
CBOS	Center of base of support	ML	Medial-lateral
CNN	Convolutional neural network	SI	Superior-inferior
COG	Center of gravity	WJC	Wrist joint center
CRP	Continuous relative phase		
FCE	Functional capacity evaluation		

1. Introduction

Functional capacity evaluations (FCE) are commonly used to assess the capacity of an individual to complete physically demanding work, like manual materials handling. FCEs can be used in the hiring process, as pre- or post- offer employment tests, or to inform return-to-work decisions regarding an individual's readiness to return to work following a workplace injury. Whether used as a post-offer employment test or to inform return to work, the results are used for job matching purposes (Dempsey, 1998) or to determine general workability (Tengland, 2011). This is completed in order to optimize the fit between the capacity of the worker and job demands.

FCEs have also become popular as a proactive tool to reduce injury rates associated with manual materials handling jobs (Harbin & Olson, 2005), specific for pre-offer or post-offer employment testing. FCEs permit testing of physical capacity (e.g. maximum lifting capacity) and movement competency (e.g. lifting strategy), which may both be implicated in injury pathways. However, there are differing results on the predictive ability of determining worker capacity (Innes, 2006) and movement competency. Perhaps this is because when clinicians make decisions regarding capacity and test endpoints, it is based primarily on biomechanical observations of body mechanics (Sinden, McGillivary, Chapman, & Fischer, 2017). While biomechanical observations are an important factor to consider, such observations are subjective and rely on the experience and eye of the clinician. Often *a priori* parameters are chosen to determine test endpoints, injury risk factors and ultimately

determine the worker's capacity through observation. Observational criteria during manual materials handling tasks in FCEs include muscle recruitment, base of support and balance, posture, and control of movements (e.g. smoothness of movements) (Allison, Galper, Hoyle, & Mecham, 2018; Reneman, Fokkens, Dijkstra, Geertzen, & Groothoff, 2005; Trippolini et al., 2014). Also, the distance of the load in relation to the body has previously been established as an observational risk factor (Smith, 1994). Using these observations to reliably determine effort and capacity remain in debate (Trippolini et al., 2014). This subjective approach is a drawback as there are multiple factors that can influence FCE administration (e.g. skills and attributes of the assessor) (Sinden et al., 2017), thus limiting the utility for clinical decision making (Trippolini et al., 2014). A motion-capture based solution may permit direct measurement of important outcomes and thus improve the predictive utility and reliability of FCE outcomes.

Attempting to introduce and collect 3D motion analysis, force plate data, electromyography etc., in a clinic is expensive and not feasible. Alternatively, 2D motion capture (i.e., video-based markerless motion capture) is more affordable than 3D motion capture analysis but suffers from several shortcomings. Overall, 2D video for 2D biomechanical analysis is less accurate, often only useable in a 2D plane (e.g. sagittal or frontal), and the processing can be labor intensive if manual digitization of the video or image is required (e.g., Dartfish technologies (Dartfish™, Dartfish Inc., Fribourg, Switzerland)). As an alternative, 2D video and pose-estimation (e.g. Cao, Hidalgo Martinez,

Simon, Wei, & Sheikh, 2019) may provide a viable alternative to predict 2D or even 3D whole-body motion data from video-based markerless motion capture.

Pose-estimation is an automated machine learning method to predict human motion data (typically joint centers) in 2D or 3D space. Briefly, pose-estimation uses the human geometrical shape to estimate 3D landmark locations within 2D space (Cao et al., 2019). One example of pose-estimation is *wrnch* (*wrnch*.ai, Montreal, Canada). *Wrnch* is a commercially available software that allows for quick landmark estimation in the 2D and 3D space (McKinnon, Sonne, & Keir, 2020). The utility of the pose-estimation software, *wrnch*, is beneficial for assessing movement (i.e., tracking time-series joint centre trajectories) in the field (e.g. warehouse) or within a clinic. Recently MyAbilities Technologies Inc. (Burlington, Canada) has incorporated *wrnch* by using the estimated 3D joint landmark data as input to perform automated calculations of key joint angles for physical demands description analysis (McKinnon et al., 2020). Use of a validated methodology to assess human movement using artificial technologies, such as pose-estimation, could also improve upon the current state of FCEs.

With access to whole-body motion data, we can objectively quantify the criteria used to assess functional capability. Base of support and balance can be calculated by determining whole body centre of gravity (COG) and the relative movement within the base of support (e.g., functional stability limit) (Holbein & Redfern, 1997). Posture can be represented by intersegmental joint or segment angles specific to postural risk factors (e.g. shoulder

abduction angle, trunk angle, knee angle) (Armstrong, Budarick, Pegg, Graham, & Fischer, 2020; Armstrong, Ross, Graham, & Fischer, 2019; Trafimow & Aruin, 2018). Control of movements is concerned with the smoothness of movements or ones coordination and can be represented using mean absolute relative phase (MARP) angles between the hip and knee (Robin Burgess-Limerick, Abernethy, & Neal, 1993). MARP may be particularly attractive as it is a measure derived from continuous relative phase (CRP) curves with the ability to quantify coordination patterns (i.e. smoothness of movements) during functional movements or tasks (Galgon & Shewokis, 2016). However, the ability to quantify these metrics hinges on the availability of motion capture data, which at present, is difficult to obtain in a clinic setting.

Perhaps, 2D video and emerging pose-estimation algorithms can be used to drive objective data analysis to assist in making decisions regarding workability. This technique has the potential to improve the reliability and validity of FCEs. However, as a first step, it is important to verify if key metrics agree when calculated using motion data obtained from video-based markerless motion capture and pose-estimation relative to motion capture data obtained with a criterion lab-based 3D motion capture system. Video-based markerless motion capture has the potential to assist in proactively identifying candidates at higher risk of musculoskeletal disorders based on movement competency. Therefore, the goal of this study was to assess the concurrent validity of video-based marklerless motion capture relative to a criterion, 3D motion capture.

2. Review of Relevant Literature

2.1 Workability

Workability defines the capacity of a worker to complete their job. Workability has been termed in relation to the normal work of the individual (i.e. what they are trained to do) and in relation to a general kind of work (i.e. that requires little or no training) (Tengland, 2011). Measuring workability involves capacity evaluations that encompasses common work tasks. If an individual is not physically fit to perform a job, then employers will need to compensate to minimize the risk of injury. This can be completed in the form of an FCE, and when administering an FCE there are several different factors that are determined and measured. While it is important in manual materials handling jobs specifically to determine maximum capacity, the clinician or test administrator also looks to assess the employer biomechanically through qualitative assessment (Sinden et al., 2017). Therefore, conclusions specific to an individual's workability can be drawn from their capacity testing administered through an FCE.

2.2 Functional Capacity Evaluations

FCEs are becoming increasingly popular within society. FCEs are also commonly called 'work capacity evaluation's' or 'physical capacity evaluation's', however, for the purpose of this thesis we will refer to them as an FCE. FCEs include a standardized battery of tests that aim to determine an individual's capacity to do work for the purpose of establishing

general work capacity or specific work capacity as related to possible activity limitations (King, Tuckwell, & Bawett, 1998). A Delphi study (Soer, van der Schans, Groothoff, Geertzen, & Reneman, 2008), including 22 international experts within the field, participated in a survey to determine a globally understood definition of what an FCE is. The following definition comprised a 68% agreement:

> "A FCE is an evaluation of capacity of activities that is used to make recommendations for participation in work while considering the person's body functions and structures, environmental factors, personal factors and health status" (Soer et al., 2008, p.394)

Further examples (Isernhagen, 2009, p.2) defined FCE as the objective measurement of a person's ability to perform functional work activities. Hart et al., (1994) defined an FCE to be able to quantify the safe functional abilities of a person with an impairment, particularly a workplace impairment. While each of these definitions are slightly different, they all identify that the objective of an FCE is to identify the capacity of an individual in a performance context and commonly related to occupational work.

2.2.1 Uses of Functional Capacity Evaluations

The most common reasons why a worker may complete an FCE is to: guide return-to-work/ workplace modification recommendations, establish baseline capacity, inform pre-hire/ post offer employment job matching, and to support vocational rehabilitation (King et al., 1998; Sinden et al., 2017). Within the insurance and workers compensation systems, FCE

assessments are used to inform decisions regarding return to work (Gross & Battie, 2005). Return to work FCEs are assessments that aim to determine if employees that have been previously injured or continue to show symptoms are able to return to their previous job, a different job or a modified job (Genovese & Isernhagen, 2009, p.38). An example of the process when an individual is injured on the job and needs to complete at return to work assessment is addressed in Figure 1. An FCE would take place prior to the return to work decision making process (denoted by a red box (Figure 1)).

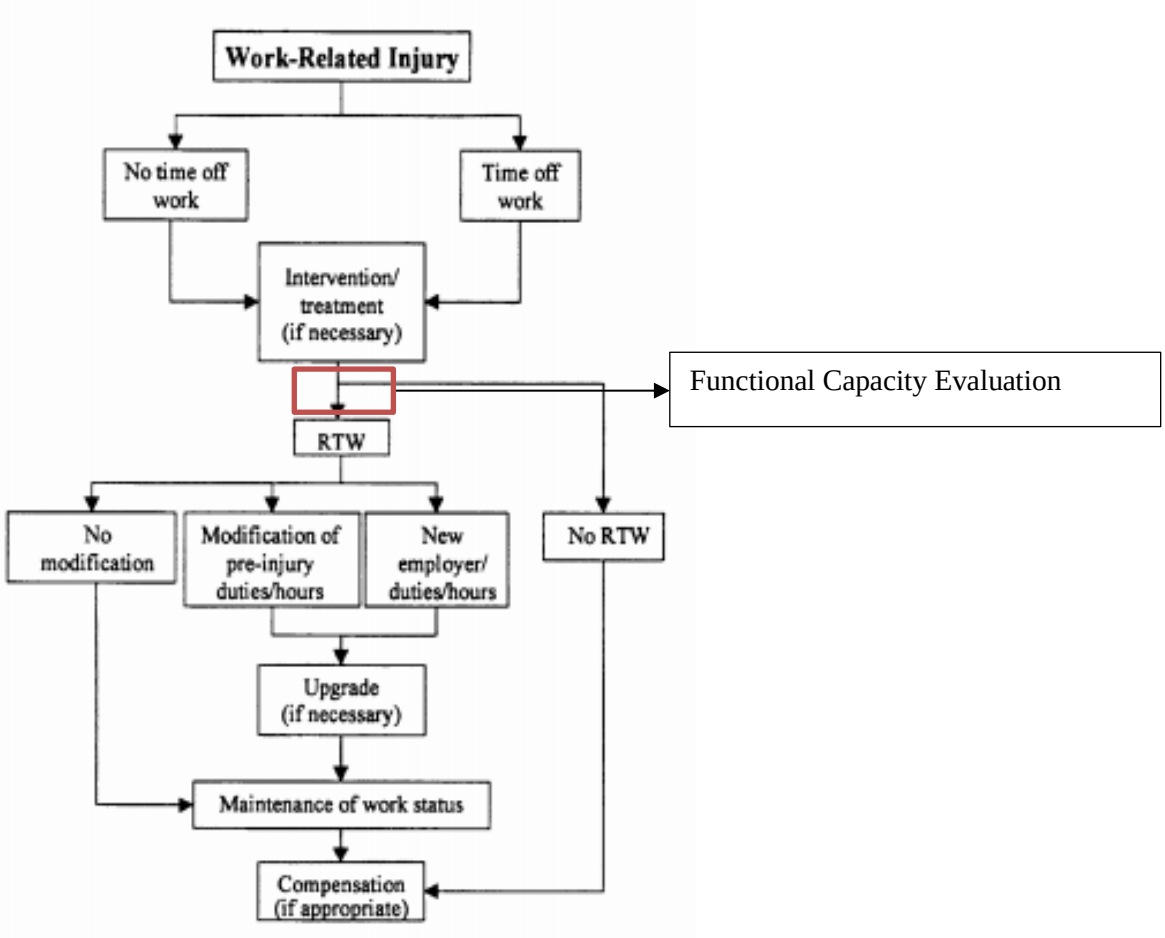


Figure 1: Overview of the occupational rehabilitation process in return to work scenario (Adapted from Innes & Straker, 1998)

Further, pre-employment screens were initially used in World War II to place workers more efficiently into appropriate jobs (Isernhagen, 2009, p.3; Harbin & Olsen, 2005). However, the purpose of pre-employment screening began to creep beyond the initial job-matching intent, where it was believed that such screening could also decrease back injury claims by evaluating work capacities including strength, flexibility and fitness (Bigos et al.,

1992). The FCE industry rationalizes the use of FCE for injury prevention by citing a series of studies demonstrating that individuals lacking the physicality to perform a job, had significant increase in incidence of low back injuries (Don B Chaffin, 1974; Don B Chaffin, Herrin, & Keyserling, 1978; Don B Chaffin & Park, 1973). Whether an FCE is administered for return to work or pre-/ post-offer employment testing, the results of the FCE are to determine the physical capacity of an individual to support job matching (Armstrong, Pegg, & Fischer, 2019). Job matching has become a tool to enhance return to work decisions, such as determining what the employer can complete safely that will proactively reduce injury rates during physically demanding jobs (Harbin & Olson, 2005) once the individual has returned to work. Although there has not yet been an established FCE protocol used for both return to work and pre-/post-offer evaluations, the FCEs used for each purpose often share common elements (Sinden et al., 2017).

2.2.2 Administering Functional Capacity Evaluations

There are several different FCE batteries that are commonly used to assess a worker's capacity. Since the increase in utilizing FCEs as a tool to assess workability for disability claims, and job matching, there has been an increase in individualized commercially available FCEs (Isernhagen, 2009, p.6). However, this has created confusion within the FCE referral space due to lack of set standards for identifying if a specific FCE protocol or model (Isernhagen, 2009, p.6). The referral space for FCEs includes physicians, vocational counsellors, employers, workers compensation insurers, disability insurers, attorneys, and the

patients themselves (Isernhagen, 2009, p.6) and are administered by health care professionals including kinesiologists, chiropractors, occupational therapists, and physiotherapists as some examples (Hart, Isernhagen, & Matheson, 1993; Sinden et al., 2017).

Over the years there has been much work done on identifying the commercially available FCEs for comparison and analysis (V Gouttebarge, Wind, Kuijer, & Frings-Dresen, 2004; Innes, 2006; Innes & Straker, 1999a, 1999b; King et al., 1998; Sinden et al., 2017; Wind, Gouttebarge, Kuijerl, Sluiter, & Frings-Dresen, 2006). Sinden et al., (2017) determined that in Canada, practitioners generally use strength-based tasks (low and mid-level lifts), push/ pull, carry, assessment of body positioning including walking, stair-climbing, crouching/ squatting and bending and upper extremity mobility including front reach, finger dexterity and grip strength. Sinden et al., (2017) also identified the use of biomechanical observations to make conclusions about physical capacity. However, when using biomechanical observations, it is important that there is a well-established definition as well as reliable and valid outcomes when establishing workability.

2.2.3 Biomechanical approach to assessing functional capacity evaluations

Along with being a popular assessment evaluation method (Sinden et al., 2017), Dempsey, (1998), identified that to define acceptable task demands relative to worker capacity, criteria based upon principles of biomechanics are necessary. The biomechanical approach focuses on the ability of the individual within the safe limits of his or her musculoskeletal and neuromuscular system (Gardener & McKenna, 1999). The concept of

using biomechanical definitions to subjectively appraise lifting mechanics has been explored previously. Gardener & McKenna (1999), looked at the reliability in distinguishing the point at where an individual moved from safe lifting style to an unsafe lifting style using biomechanical endpoint definitions. The study identified only moderate to good inter-rater reliability prior to defining a biomechanical endpoint definition but good to excellent inter-rater reliability post biomechanical endpoint definition (Gardener & McKenna, 1999). Endpoint definitions included upper limb or trunk extreme range of motion, poor control of the load including poor balance and load was not maintained close to the body throughout the lift (Gardener & McKenna, 1999). Although this study demonstrated the ability to have good inter-rater reliability after the raters were given end-point criteria, the criteria remains vague in nature and therefore may not always be interpreted similarly each time. When presented with an FCE, if biomechanical criteria to determine test end-points and cut-offs are used, specific criteria or instructions are needed to attain the understanding of a wide audience such as test administrators of all different backgrounds (i.e. physiotherapists, clinicians etc.,). One way to address this issue is to incorporate objective analysis into FCEs through video-based markerless motion capture analysis.

2.3 Reliability and validity of Functional Capacity Evaluations

It is important that FCEs are both reliable and valid so that the employer can be accurately informed about the state of the worker. Regarding FCE's there are different definitions when it comes to reliability (V Gouttebarge et al., 2004). *Test-retest reliability*

determines the consistency of measures or scores from one testing occasion to another and assumes that the test characteristics do not change over time and cannot be influenced by testing effects (with practice or carryover effects), rater bias, or time of test intervals (Innes & Straker, 1999a). Secondly, *rater reliability* is the consistency which all examiners or evaluators make the same judgments or measurements such as applying the tools or reading the instruments (Innes & Straker, 1999a). Lastly, *intra-rater* refers to the consistency of data recorded more than once or across different testing occasions by the same person and *inter-rater reliability* is the variability or difference between two or more examiners (Innes & Straker, 1999a). *Construct validity* is the ability of a test to measure a hypothetical construct such as being able to differentiate between clients who are able to lift safely and who are not (Innes & Straker, 1999b). *Predictive validity* measures variables at two different points in time (Innes & Straker, 1999b) establishing the ability to predict a client's work performance, based on their performance during the FCE battery. *Concurrent validity* looks at the similarity or correlation of two measures at the same time so they both reflect the incident of behaviour (Innes & Straker, 1999b). For context, this study is about concurrent validity.

Tuckwell, Straker, & Barrett, (2002) hypothesized that clinician skill level may be a significant factor in test-retest reliability as a high degree of skill and experience is required to observe and record data as well as to observe the differences in the quality of movement performance. Overall, studies on reliability generally demonstrate "moderate to excellent levels of reliability, particularly for manual/ material handling components" (Matheson,

Verna, Dreisinger, Leggett, & Mayer, 2014). As an example, the WorkWell FCE has previously demonstrated substantial reliability for the floor to waist lift (Innes & Straker, 1999a) and good to excellent test-retest reliability for manual materials handling tasks in low back pain patients (Brouwer et al., 2003) and in healthy patients for the static pushing and shuttle walk test (Reneman, Brouwer, Dijkstra, & Geertzen, 2004). Moreover, the EPIC lift capacity test, a six-stage progressive lifting protocol (Matheson et al., 2005) demonstrated acceptable intra-rater reliability and inter-rater reliability within both a laboratory setting and within the field as well as good reliability on a test-retest basis for each sub-test of the EPIC (Matheson et al., 2014).

Validity remains an important considering in the FCE space. Poor construct validity was demonstrated when comparing injured and uninjured group using the WorkWell FCE (Innes & Straker, 1999b). Also, it has been reported that the predictive validity of the WorkWell FCE to identify individuals who can safely return to work is limited (Matheson, Isernhagen, & Hart, 2002) and needs to be researched further. The ARCON, a commonly used FCE paradigm (Sinden et al., 2017), demonstrated promising construct validity as static lift, push and pull components were found to be significantly improved in a group of subjects with low back dysfunction tested before and after a work hardening program (Robert, Blide, Whorter, & Coursey, 1995). The Ergo Kit has demonstrated poor concurrent validity when comparing upper lifting strength and the waist-to-overhead lift (Ijmker, Gerrits, & Reneman,

2003). Evaluation of the Ergo Kit has also shown poor predictive validity on durable return to work (Gouttebarge, Wind, Paul Kuijer, Sluiter, & Frings-Dresen, 2009).

The findings from the above FCEs and their reported reliability and validity present contentious results. Reliability, while still controversial, reported as being acceptable, and validity was lacking for different FCEs and across different subtests. To improve upon results and make different FCEs consistent across evaluators and testing methods with the ability to predict workability, perhaps using more objective biomechanical outcomes that are commonly observed visually during FCEs may enable improvement. Using markerless motion capture and pose-estimation is a method that can be explored to produce biomechanical outcome measures that may assist in improving upon the reliability and validity of FCEs. First, the concurrent validity of a markerless motion capture approach needs to be explored.

2.4 Quantifying movement features of a lifting task within an occupational context

As mentioned previously, lifting is a commonly assessed task in FCEs, specific to manual materials handling jobs, for the purpose of determining lifting capacity as well as movement competency.

Often, *a priori* movement features are chosen and described to determine movement competency and test end points, or when the employee can no longer complete a task. In relation to manual materials handling jobs and FCE, Allison et al., (2018) has stated that biomechanical monitoring should include observations of movement patterns and balance.

Furthermore, Trippolini et al., (2014) completed a study where clinicians etc., rated individuals completing an FCE using qualitative measures of base of support, posture, and control (e.g. smooth movements, momentum etc.,). Coordination can signify smoothness of movements, where changes in coordination might be useful for inferring effort relative to maximum capacity (R. Burgess-Limerick, Abernethy, Neal, & Kippers, 1995). Range of motion and posture is also an important factor to consider when analyzing lifting for occupational performance (Chaffin & Page, 1994; Reneman et al., 2005; Sinden et al., 2017). As an example the WorkWell FCE describes an upright posture and natural stance to be related to a 'light' level of effort, a very solid base and increased counter balance to be 'maximal' level of effort and a 'moderate' effort level coincides with smooth movements (Reneman et al., 2005). While these terms are descriptive, they continue to be vague and therefore objectivity within the analysis can improve accuracy, reliability, and validity of different FCEs containing the same task. Therefore, it may be applicable to study the ability of a 2D pose-estimation software to accurately calculate the following variables for FCEs.

2.4.1 Balance and Base of Support

Balance and BOS have been mentioned as measure of importance for when individuals are completing manual materials handling (Allison et al., 2018, Trippolini et al., 2014). Regarding lifting, perhaps it is because the lifting of unexpected loads can lead to over or underestimation in the amount of effort needed to lift the box, thus increasing the likelihood of losing balance (Commissaris & Toussaint, 1997). Commissaris & Toussaint

(1997), studied participant balance when lifting an unknown load and discovered that in 92% of all lifting trials which subjects were induced to overestimating the box's weight, they lost balance.

Furthermore, Armstrong et al., (2020) looked at the biomechanical differences in high exposure and low exposure lifters and objectively identified that lifters with lower low back exposure when completing a lift had a wider base of support. A wider BOS leads to greater BOS limits specifically in the medial-lateral (ML) direction. Further, the Epic lift capacity guidelines advocate to maintain foot placement in a broad and stable stance during work tasks ("EPIC Lift Capacity Test (ELC)," n.d.).

A stable posture is described as the body's COG being within the BOS (Holbein & Redfern, 1997). Thus, to quantify balance during lifting, looking at the COG in relation to the base of support is promising (Commissaris & Toussaint, 1997). Without the ground reaction forces and center of pressure measurement from force plates, the functional stability limit (FSL) is a way to quantify the relationship between the centre of base of support (CBOS) and COG. FSLs are quantified as the distance an individual would displace their COG from the CBOS to the BOS limit in a direction expressed as the maximum possible distance in that specified direction (Holbein-Jenny, McDermott, Shaw, & Demchak, 2007). As an example, the BOS limit in the anterior-posterior (AP) direction is the distance between the CBOS and the mid-point between the left and right toes (Figure 2). Balance and stability are important

metrics and therefore, FSL's in the ML, as well as the AP directions may be an important quantifiable measure to consider during FCEs.

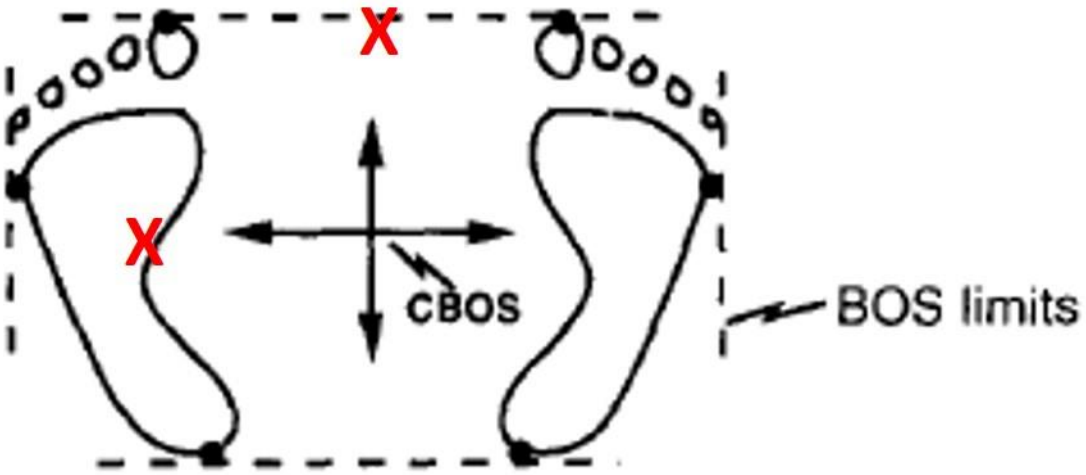


Figure 2: Figure adapted from Holbein and Redfern (1997) to illustrate the CBOS and BOS limits. The BOS limits specific to the anterior-posterior and medial-lateral directions are denoted by the red X.

2.4.2 Distance of the load relative to the body

The Epic lift capacity states that minimizing the horizontal distance between the load and the body during a lift limits a high-risk workstyle ((ELC), 2016). Along with directing individuals to "keep the load close to the body" (Allison et al., 2018, Smith, 1994), Armstrong et al., (2020) were able to differentiate high risk and low risk movers, in part, on the load proximity to the body. Low risk movers (i.e. decreased low back and shoulder exposure) tended to hold the load closer to the body during the lift (Armstrong et al., 2020). Reducing the distance of the load from the body reduces trunk inclination and the load

moment as increasing the horizontal distance influences the posture adopted to lift the load (Burgess-Limerick & Abernethy, 1998). Therefore, objectively measuring the distance of the load in relation to the body (e.g. COG) can prove useful during FCE performance.

2.4.3 Posture

When it comes to quantifying posture during lifting techniques, for a floor-to-waist or floor to knuckle height lift, the terminology stoop versus squat is often assessed. A stoop lift incorporates greater trunk flexion and less knee flexion than a squat lift technique when picking up a box. Armstrong et al., (2020) objectively quantified that lifters who exhibit a stoop strategy versus a squat strategy experienced higher low back loading. While in manual materials handling job design the stoop versus squat postures can vary depending on the individual's preference, load lifted, and task constraint (e.g. large or small box) it is still important to quantify during FCE.

Range of motion and posture can both be indicators of performance for manual materials handling tasks, such as lifting (Trippolini et al., 2014). It is important to consider biomechanical analysis during FCEs for identification of individuals that may be at higher risk of injury (Armstrong et al., 2020). It was discovered that individuals with a high exposure (to the low back and shoulders) movement pattern during lifting had greater trunk and shoulder angles than the lower exposure lifters (Armstrong et al., 2020). Therefore, having a more upright posture and limiting the shoulder angular displacement while lifting

could reduce low back and shoulder exposure suggesting trunk angle (specific to flexion) and shoulder angles (flexion and abduction) to be important indicators of lifting performance.

2.4.4 Coordination

Inter-joint or inter-segmental coordination can represent smoothness by quantifying if individuals are completing a lifting task in-phase (i.e. the joints are flexing and extending at the same time respectively) or out-of-phase. Thus, analyzing this interaction between joints or segments can assist in quantifying the role of each joint during a task (Burgess-Limerick, Abernethy, & Neal, 1993). One measure of inter-joint coordination is continuous relative phase (CRP). CRP is a signal generated that represents the difference in phase angles between two signals (Lamb & Stöckl, 2014). As an example, the phase angle of the knee flexion-extension and phase angle of the hip flexion-extension are both calculated and the difference between the two joint phase angles are analyzed. CRP can be used to analyze which segment or joint is moving first in the appropriate flexion/ extension direction identifying in-phase or out-of-phase lift (Robin Burgess-Limerick et al., 1993) An in-phase lift is identified when the segments or joints are moving together and an out-of-phase lift is when the segments or joints are not moving together. For this study, we will not necessarily claim in-phase or out-of-phase, but it is important to not the general purpose of the coordination measure. We can also use this measure to assess coordination regarding coordination variability between two joints or segments (Robertson, Caldwell, & Hamill,

2014, p. 306). Coordination variability would explain the extent to how much participants are demonstrating coordination or not, over the whole lift.

When looking at trunk and lower limb coordination, it has been suggested that more severe disability patients demonstrate more out-of-phase joint coordination during lifting compared to healthy individuals (Pranata et al., 2018). Further, it has been suggested that individuals with a history of low back pain demonstrate a more anti-phase lifting technique with respect to the pelvis-trunk compared to their high back pain counterparts (Seay, Sauer, Patel, & Roy, 2016). The authors suggest the in-phase lifting motion is analogous to a "guarded-gait" pattern and may give insight as to why individuals are at higher risk of re-injury (Seay et al., 2016). Lastly, inter-joint coordination measures of CRP can demonstrate significant differences between men and women in the kinematics of lifts (Lindbeck & Kjellberg, 2001). The previous literature suggests inter-joint coordination to be an important measure to consider for individuals completing lifting tasks.

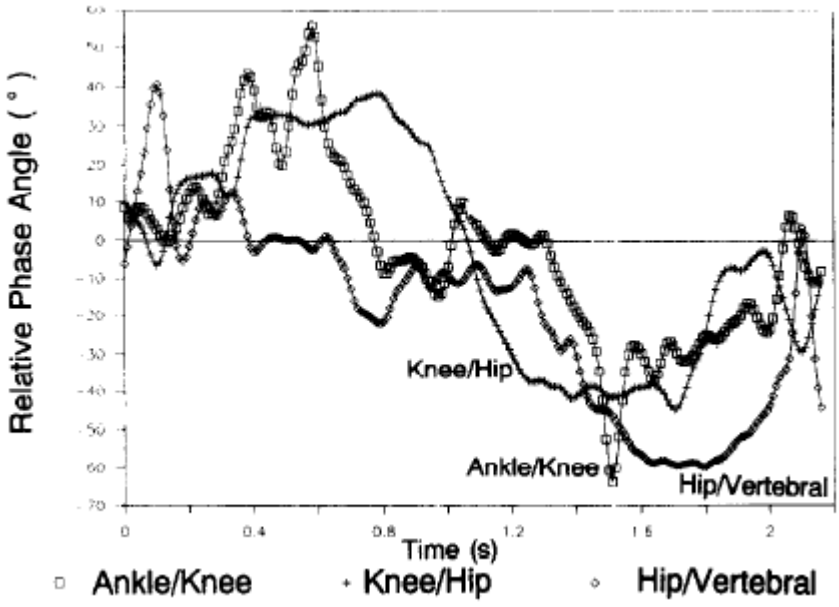


Figure 3: Example of relative phase angles as a function of time during one lifting trial. The proximal-to-distal configuration suggests the proximal joint leads the distal during the flexion phase (first half) and vice versa during the extension phase (second half) (Robin Burgess-Limerick et al., 1993)

2.5 2D pose-estimation

Pose-estimation has received much attention for its utility in surveillance, as well as human-computer interfaces (Luvizon, Picard, & Tabia, 2018). Oftentimes, specific algorithms behind pose-estimation software are kept secret for copyright purposes. Specific to the software used in this thesis, *wrnch* performs human motion capture via joint skeletal tracking, which tracks humans in a video by creating a skeleton overlay. While the "secret sauce" or algorithms behind *wrnch* is unknown to its user, we can still discuss the overall

methodology used that is publicly available. The software decodes data within the pixels in the video, where the computer then needs to know which pixels go together and what they represent ("Science and technology," n.d.). Capturing the human shape is performed with *human instance segmentation*, a set of machine learning algorithms optimized for the detection of human shapes ("Science and Technology", n.d.). Further, background segmentation, subtracting the background environment from the humans in the video, and the use of mesh data can all help with human segmentation and representation of the human body in 3D space and estimate its volume ("Science and Technology", n.d.).

There are other pose-estimation based software's used in research. One example is the Microsoft Kinect™. The Microsoft Kinect™ was originally developed to track the movements of a player interacting with a game during a video game, however, has been adapted to track human movements for research purposes (A. Pfister, West, Bronner, & Noah, 2014). The Kinect™ consists of an infrared light projector, an infrared camera, and a RGB video camera, which are all used to estimate depth data and distinguish shapes thus allowing the possibility of using 2D video to estimate 3D skeletal positions (A. Pfister et al., 2014). The Kinect™ has been used previously in research for postural control (Clark et al., 2012), dance gesture recognition (Raptis, Kirovski, & Hoppe, 2011), and gait analysis (Nambiar, Correia, & Soares, 2012; Stone & Skubic, 2011)S) as some examples. Overall, the Kinect™ demonstrated the ability to generally be able to track the participants movements, but systematic errors affected accuracy (Pfister et al., 2014).

Other pose-estimation based methods do exist that do not require the need of a depth sensor. Pose-estimation is a methodology to use 2D images or 2D video and estimate 3D trajectories. This process is complicated due to the ambiguous tasks presented. First, 2D location of human joints, or landmarks must be found in space which is all influenced by camera angles, clothing, body shape, or external and self-occlusions (Tome, Russell, & Agapito, 2017). Secondly, extracting 3D coordinates from the 2D landmarks poses a problem because the possible space of 3D poses of the human body is infinite (Tome et al., 2017). Therefore, some assumptions likely need to be input in the form of 3D geometric poses and temporal or structural constraints (Tome et al., 2017). These assumptions can be input as a form of inference over a combination of local observations on different body parts and the spatial dependencies between them (Cao et al., 2019). The human body has a geometric shape with spatial relationships between segments. Spatial dependency is then used as a way to estimate the distances between joints and segments in space (Qi, Bayramli, Ali, Zhang, & Lu, 2019). This way, any device that captures 2D video can be utilized to estimate 3D landmark trajectories through pose-estimation. Spatial dependencies are used in human pose-estimation to encode spatial relationships between adjacent parts following a kinematic chain (Cao et al., 2019). Secondly, non-tree models can be used which augment the structure with additional edges capturing occlusion, symmetry, and long-range relationships (Cao et al., 2019). Furthermore, pose-estimation methods have often used pictorial structure models, which optimize a configuration of parts as a function of local image evidence for a part and a

prior for the relative positioning of parts in the human chain (T. Pfister, Charles, & Zisserman, 2015). As an example, a software might use facial recognition algorithms to recognize a subject's face and then subsequently estimate the skeletal output by tracking down the kinematic chain. In pose-estimation algorithms, convolutional neural networks (CNN's) are utilized to improve upon the accuracy (Cao et al., 2019; A. Pfister et al., 2014). The idea behind using CNNs to improve accuracy in pose-estimation and skeletal tracking is that a large input source of videos is manually labelled to train the pose-estimation software. The trained model is then applied to new videos (a test set) by taking what it has learned through the CNN. The CNN can be a powerful framework to learn discriminative image features and then afterwards estimate 3D poses from them (Park, Hwang, & Kwak, 2016).

2.5.1 2D video-based analysis within biomechanics

Two-dimensional (2D) video-based analysis can be useful in a variety of contexts such as in sport (Hanley, Tucker, & Bissas, 2018), movement screening (Munro, Herrington, & Carolan, 2012) balance and gait (Clark et al., 2012), and characterizing postures during different tasks (Kingma et al., 1998; Norris & Olson, 2011; Schurr, Marshall, Resch, & Saliba, 2017; Sinden & MacDermid, 2016). Although the validation of 2D video analysis software has been previously looked at (Hanley et al., 2018; Norris & Olson, 2011; Sinden & MacDermid, 2016) it remains unutilized within the FCE space and with automated tracking (i.e. pose-estimation). This is surprising considering the high percentage of administrators that use biomechanical observations within their assessment (Sinden et al., 2017).

There are few studies that investigated the utility of 2D pose-estimation for lifting tasks (Kingma et al., 1998; Norris & Olson, 2011; Sinden & MacDermid, 2016). Two-dimensional sagittal plane analysis using manually tracked (i.e. digitized) Dartfish software of knee and hip flexion during lifting have shown promising results (Norris & Olson, 2011). Furthermore, firefighter trunk and knee postures in the sagittal plane could be assessed with reasonable error margins (9 degrees for trunk and 5 degrees for knee) of reliability of markerless video analysis (Sinden & MacDermid, 2016). However, these two studies still lack assessment of concurrent validity of a 2D markerless motion capture and pose-estimation software with a 3D motion capture-based analysis.

Looking into the concurrent validity specifically between 2D markerless motion capture and pose-estimation with 3D motion capture-based approaches, there is less research. Concurrent validity has been studied in reference to different markerless motion capture systems. Harsted, Holsgaard-Larsen, Hestbæk, Boyle, & Lauridsen, (2019), used a 3D markerless motion capture system call “The Captury” to quantify the concurrent validity of lower extremity kinematics and jump characteristics in pre-school children. Briefly, “The Captury” system uses optical 3D motion capture system based on traditional commercial video cameras and based on a passive system (Harsted et al., 2019). It uses a background subtraction method and subsequently fits a template skeleton onto a determined silhouette (Harsted et al., 2019). The results of this study specific to knee flexion for squatting and jumping for the markerless system to estimate values greater than that of Vicon (Harsted et

al., 2019). Pose-estimation derived from *wrnch* was studied in regard to ergonomic tasks (McKinnon et al., 2020). Through several different tasks (e.g. “retrieve goods from freezer”), upper body joint and trunk angles were assessed between a pose-estimation and 3D motion capture (McKinnon et al., 2020). Their study showed fair (Cohen’s kappa 0.29 – 0.31 for reach distance and elbow angle) and moderate to substantial (Cohen’s kappa > 0.50 for trunk and shoulder angle) agreement for different outcome measures (McKinnon et al., 2020). There were further results demonstrating low agreement when analyzed using root mean squared errors (15-22 degrees for angles) (McKinnon et al., 2020). While these results may seem promising for different ergonomic tasks, lifting was not analyzed.

Therefore, one area where research is lacking is an assessment of the concurrent validity of 2D pose-estimation for calculating metrics of interested relevant to a FCE and manual materials handling work.

2.5.2 Approaches to concurrent validity

Specific to markerless motion capture and 3D motion capture, there are different analysis techniques to compare the approaches when measuring concurrent validity. As some examples, both correlations (e.g. Pearson product moment correlation) and agreement (e.g. Bland-Altman) have been used to assess the relationship and agreement, respectively, between a 2D and 3D kinematic analysis. Specific tools include root mean squared error (Hanley et al., 2018; Harsted et al., 2019; McKinnon et al., 2020), intraclass correlation coefficient (Harsted et al., 2019; A. Pfister et al., 2014; Verlekar et al., 2019), Pearson

product correlation coefficient (Schurr et al., 2017; Sorenson, Kernozek, Willson, Ragan, & Hove, 2015), linear regression (Alahmari, Herrington, & Jones, 2020; Sorenson et al., 2015), ANOVA (Hanley et al., 2018), paired t-tests (A. Pfister et al., 2014) and Bland-Altman analysis (Clark et al., 2012; Harsted et al., 2019; Maykut, Taylor-Haas, Paterno, DiCesare, & Ford, 2015; A. Pfister et al., 2014; Schurr et al., 2017).

Correlation (e.g. Pearson product moment correlation) and regression analysis are accepted means for assessing inter- and intra-rater reliability. Correlation shows how strongly pairs of variables are related, and linear regression is often calculated when a correlation exists (Giavarina, 2015). In an agreement study or methods comparison, a high correlation could simply mean the sample of the study is widespread (Giavarina, 2015). Thus, a good correlation can be different from a good or accepted agreement (Bland & Altman, 1995). Specific to the intraclass correlation coefficient, there is a lack of interpretability of high values and failure to provide information of the range over which the measures agree (McAlinden, Khadka, & Pesudovs, 2011). Therefore, it is discouraged in the assessment of concurrent validity through agreement analysis. Similarly, it has been pointed out that a regression analysis (r) demonstrates the strength between the two variables, again this is different from the agreement (Giavarina, 2015). Moreover, analyses such as paired t-tests or ANOVAs, look to measure if the mean difference between two or multiple methods or sets of data is 0. However, there is not any information to determine how well they agree for individuals (Bland & Altman, 1995). In turn, we are not able to draw the correct conclusions

using a t-test or ANOVA because if the two methods agree on average, the test will be non-significant regardless of the agreement of individuals (Bland & Altman, 1995).

The Bland-Altman analysis and plots can give a visual representation of the mean difference between the methods, relationships between the difference and mean of the 2 methods, and any systematic error (bias) represented in the dataset. Systematic error cannot be detected with a correlation coefficient (McAlinden et al., 2011). Therefore, considering the goal of this study is to assess concurrent validity through agreement, the Bland-Altman analysis is the most appropriate.

3. Research Questions and Hypotheses

The objective of this thesis was to determine the concurrent validity of outputs calculated using kinematic data from a 2D video-based markerless motion capture and pose-estimation software relative to kinematic data obtained using a laboratory-based 3D motion capture system for tasks specific to a functional capacity evaluation battery.

Research Question: Do kinematic measures including peak trunk, knee and shoulder flexion angle, peak shoulder abduction angle, peak functional stability limit in the anterior-posterior and medial-lateral, peak distance of the load relative to the body, and hip-knee flexion/ extension mean absolute relative phase angles calculated from kinematics obtained using 2D video-based pose-estimation software and a 3D motion capture analysis system demonstrate agreement?

Hypothesis 1A: Peak trunk flexion, peak knee flexion, peak shoulder flexion, peak distance of the load relative to the body, and MARP values will demonstrate agreement using a Bland-Altman analysis.

Hypothesis 1B: Peak shoulder abduction and FSL's will not demonstrate agreement.

4. Methods

4.1 Study Design

We applied a cross-sectional, within subject's design. Participants arrived to the lab for a single visit to perform a series of floor-to-waist height lifts. Lifting motions were recorded using a traditional 3D motion capture system (Vicon, Oxford, UK) and a video camera. Post collection, the video camera recordings were processed through a pose-estimation software (*wrnch*, Montreal, Quebec) to generate marker data in a 3D space. Marker data from the 3D motion capture system and markerless motion capture approach were each used to calculate dependent measures of interest. Bland-Altman plots were used to determine the agreement of the dependent measures calculated using marker data from the 3D motion capture system and markerless motion capture approach. In this case, agreement was calculated as a form of determining concurrent validity of the two approaches.

4.2 Participants

Twenty healthy participants completed the study (Table 1). Eligible participants were 18 years or older, free of injury and ready to partake in physical activity, as determined using the Get Active Questionnaire. There were no requirements regarding previous lifting experience or strength. The participant pool was composed of young University students primarily from the Department of Applied Health Sciences. The study was approved by the University of Waterloo ethics committee (ORE #41350). Prior to beginning the study, written informed consent was obtained from each participant. Participant height and mass were then

recorded taking note of the anterior superior iliac spine height for the purpose of adjusting the shelf height.

Table 1: Participant Demographics. MMH – manual materials handling experience. MMH experience in years refers to number of participants with the associated years of experience who said yes to MMH experience.

		Males (N=11)	Females (N=9)
Age (years)		21.6 ± 3.1	21 ± 1.2
Height (cm)		178.4 ± 11.0	163.1 ± 9.0
Weight (kg)		76.6 ± 12.3	58.5 ± 11.7
MMH Experience (# participants)		7	3
MMH experience (years)	≤ 1 year	5	1
	1 to 5 years	1	1
	≥ 5 years	1	1

4.3 Instrumentation

Primary instrumentation for this study included a 3D motion capture system (Vicon Nexus 2.6, (Vicon, Oxford, UK)) and a video camera (Sony HDR-CX240 Handycam). Participants' motion was recorded at 60 Hz using the 3D motion capture system and at 30 Hz using the video camera. The video camera was set up to record motion in the sagittal plane, relative to the lifting actions (Figure 6). The video camera was set up laterally to the shelf to

record the left side of the participant while completing the lift (Figure 5). Trials were eliminated from analysis if the video did not collect the entire motion of the participant completing the task or marker drop out during 3D motion capture impeded data quality. A total of 9 trials (3 participants) were excluded from analysis for the measures pertaining to the COG due to marker drop out that was needed in the calculation of COG.

Participant instrumentation included a whole-body marker set-up (Figure 4). The whole-body marker set up consisted of both clusters and individual markers. Clusters with 4 or 5 markers each were placed on body segments (bilaterally on the forearms, upper arms, shanks, thighs, as well as the pelvis and torso). Further, individual markers were placed on anatomical landmarks (anterior and posterior lateral head, 7^{th} cervical vertebra, suprasternal notch, 8^{th} thoracic vertebra, xiphoid process, and bilaterally on the acromion, lateral and medial epicondyle, ulnar and radial styloid, 5^{th} and 2^{nd} metacarpals, dorsum of the hand, iliac crest, anterior superior iliac spine, greater trochanter of the femur, lateral and medial femoral condyles, lateral and medial malleoli, calcaneus, base of 3^{rd} tarsal, and base of 1^{st} and 5^{th} tarsals) (Figure 4). Participants then completed a standing calibration trial and then subsequently calibration markers including 8^{th} thoracic vertebra, xiphoid process, bilaterally on the lateral and medial epicondyles, femoral condyles, malleoli, anterior superior iliac spine, greater trochanter, were removed for the remainder of the study.

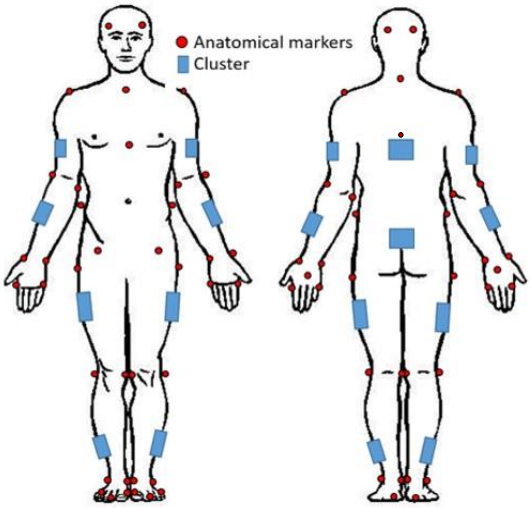


Figure 4: Anatomical landmarks (red dots) and clusters (blue squares) placed on the participant (Adapted from Remedios, Armstrong, Graham, & Fischer, 2020)**)**

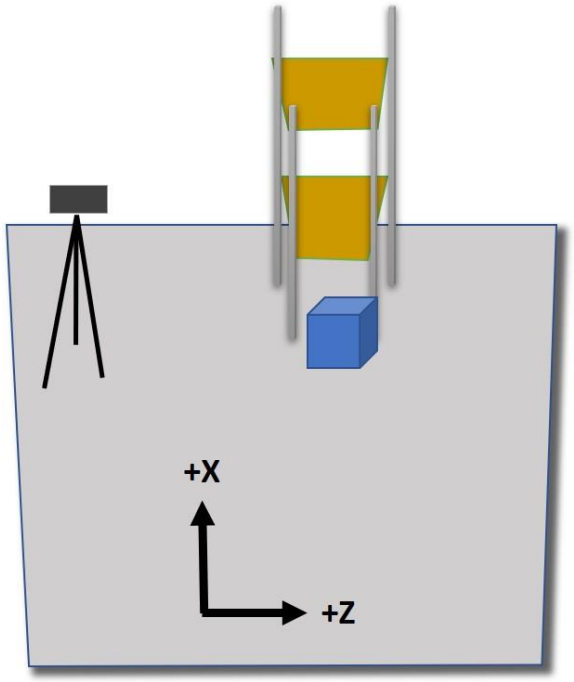


Figure 5: Visual representation of the laboratory set up. One lifting shelf was in the space with the first shelf set to the participants approximate mid-anterior superior iliac spine height. The origin of the space was placed posterior to the shelf and the movement of the participant defining the global coordinate system

4.4 Protocol

For the study, participants were required to lift and lower boxes at various loads for the floor-to-waist lifting task. Similar to best practice (Allison et al., 2018), the only instruction the participants were given was to keep the load close to the body during the entire lift. Participants completed a total of 7 lift and lowers at the floor-to-waist height level. Lifting and lowering test batteries are required as per the 'Best Practice Guideline' (Allison et al., 2018) and consistent with clinical practice at Functional Capacity Experts LLC, (Bossier

City, Louisiana). The maximum load lifted by any participant was determined subjectively up to 23 kg (Waters et al., 1993). The 7 loads consisted of a range of approximately 13.6 kg between the absolute maximum and minimum in increments of 2.27kg. The loads were implemented in a randomized order and participants were blinded to the loads in the box to eliminate bias. Further, 60 seconds of rest between lifts was required in attempt to reduce effects of fatigue. From the group of 7 lifts and lowers for each participant, 3 were retained for this analysis, including the participants heaviest lift (heavy condition), lightest lift (light load condition), and mean of the load's lifted (medium condition). Three lifts per participant represents a total sample size of n = 60 (20 participants x 3 lifts). A minimum of 59 samples satisfies the criteria for calculating agreement using Bland-Altman analysis with 2 samples (i.e. approaches) from each trial (Liao, 2010). Markers important for calculating whole-body COG were occluded in nine trials. Therefore, 60 trials were used when calculating the posture and coordination-based measures and 51 trials were included when calculating the COG to load and FSL outcome measures.

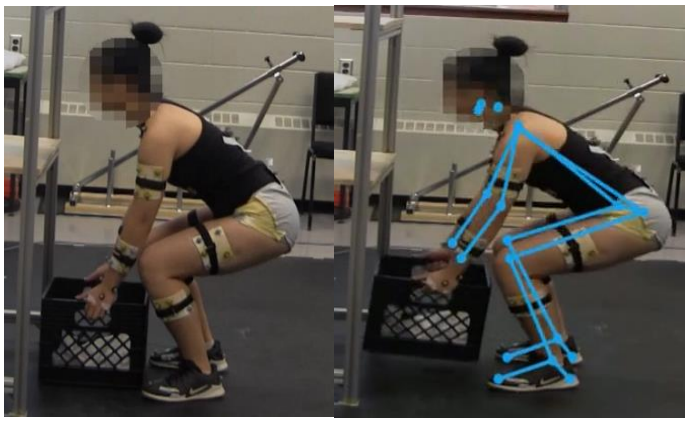


Figure 6: Figure on the left, demonstrating the raw video, along with the marker clusters attached to the participant. Figure on the left demonstrating the overlayed pose-estimation.

4.5 Outcome measures

Concurrent validity through agreement analysis was based on specific kinematic metrics calculated from each motion capture source. Table 2 lists the outcome measures that were calculated as well as the rationale for considering each measure (i.e. how it relates to the observational measured used in FCEs).

Table 2: Kinematic outcome measures calculated for floor-to-waist lift

Outcome	Significance
Peak trunk flexion angle	Posture
Peak knee flexion angle	Posture
Peak shoulder flexion and abduction	Posture
Centre of gravity (COG) relative to base of support a) Peak Functional Stability Limit (FSL %) anterior - posterior b) Peak functional stability limit (FSL %) medial-lateral	Balance
Peak distance between the location of the wrist to the COG in the sagittal plane COG related to wrist centre in the anterior-posterior direction relative to participant height (%)	Location of load relative to the body
Mean absolute relative phase angles **a)** Hip-knee flexion/ extension	Lifting coordination

4.6 Data processing and analysis

A flow chart of the data processing procedure is presented in Figure 7. This Figure represents the similarities and differences when processing and analyzing data from the markerless motion capture approach (2D) and 3D motion capture approach. Differences between the two approaches is evident in the models and global coordinate space for each approach, respectively. The model and coordinate system for the markerless motion capture (pose-estimation) approach is provided by *wrnch* and presented in Figure 9. This coordinate system differs from the 3D motion capture global coordinate system because values in all axes are restricted to 0 to 1 unit-normalized screen space coordinates. The pose-estimation model and 3D motion capture model present both similarities and differences. Each models' segment definitions are presented in Table 3 for comparison. These differences can affect the results of this study and therefore are important to consider upon interpretation of the outcome measures.

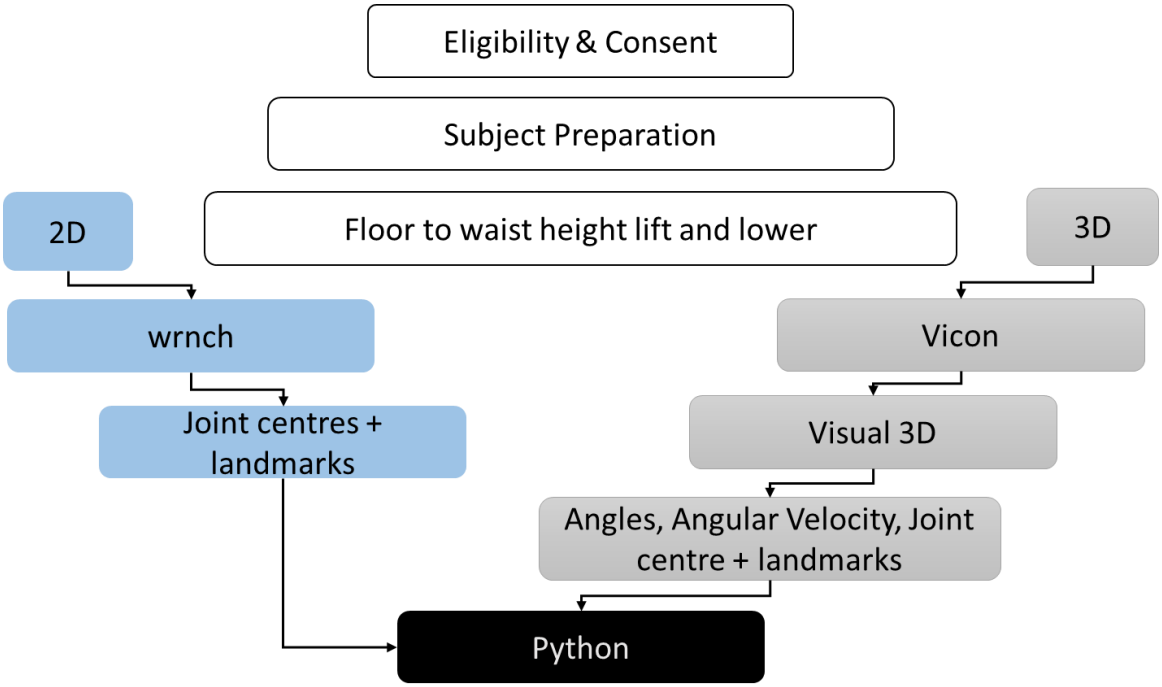


Figure 7: Overview of the data collection and processing demonstrating the differences between 2D and 3D analysis, where outcome measures were ultimately determined in Python

4.6.1 3D Marker data conditioning

All reflective markers were labelled, and gap filled in Vicon Nexus. Gaps were filled using cubic spline (gaps <_ 200 ms), pattern fill or rigid body fill (gaps > 200 ms) functions within the Vicon software and used appropriately for each gap length, consistent with best practice (Howarth & Callaghan, 2010). Vicon data were then exported to Visual3D (C-Motion Inc, Germantown, USA). In Visual3D, a rigid-linked skeleton model template based off the anatomical marker locations and clusters was created and applied first to the static

calibration trial of each participant to mold the individual anthropometrics and then subsequently to each lifting trial. For the model, landmarks, joint centers and segments are described below.

Defining Landmarks: Landmarks were created to assist in defining the pelvis segment, trunk segment and joint centers described below. A lower mid-torso landmark was created using the midpoint between the sacrum (pelvis cluster) and the mid-anterior superior iliac spine projected on a line. An upper mid-torso landmark was created as the midpoint between the 7th cervical vertebra and supra-sternal notch markers projected on a line. Additional landmarks were also created to model extremity segments. A mid-foot landmark was created between the 1st and 5th tarsal markers for the purpose of defining the foot segment. Left and right mid-head landmarks were created using the anterior and posterior lateral head markers for the left and right sides respectively to aid in defining the head segment.

Defining joint centers: Joint centers were created for the ankle, knee, elbow, and wrist according to ISB recommendations (Wu et al., 2002, 2005). These joint centers were all defined as the midpoint on a line projected between the corresponding medial and lateral anatomical markers. The midway point between the lateral and medial malleoli created the ankle joint center, the medial and lateral femoral condyles established the knee joint center, the elbow and wrist joint centers were estimated using the midpoint of the medial and lateral epicondyles and ulnar and radial styloid respectively.

For the shoulder joint center, the acromion marker as well as the upper mid-torso and lower mid-torso landmarks were used. The glenohumeral joint center was determined by using the acromion marker and projecting a point 5cm inferiorly along a line coincident with a vector pointing from the upper to lower mid-torso landmarks (Nussbaum & Zhang, 2000).

The hip joint centers were created once the pelvis segment was creating using a Coda pelvis model. The hip joint centers were defined based from the Bell 1989 and 1990 papers (Bell, Brand, & Pedersen, 1989; Bell, Pedersen, & Brand, 1990).

Defining body segments: A summary of the segment definitions are presented in Table 3. Details regarding the 3D motion capture model segment definitions are presented here. The pelvis segment, Coda model, used the sacrum center and the left and right anterior superior iliac spine markers to define the segment. The trunk segment was defined using the upper- and lower-mid torso landmarks as the distal and proximal endpoints, respectively. Body segments, bilaterally for the shank (knee joint center / ankle joint center), upper arm (shoulder joint center / elbow joint center), and forearm (elbow joint center / wrist joint center) were defined using proximal and distal joint centers, respectively. The foot segment was created using the ankle joint center as the proximal endpoint and mid-foot landmark as the distal endpoint. The radius of the proximal and distal ends of the shank, upper arm, forearm, and foot were defined as the distance between medial and lateral anatomical landmarks divided by 2 for each joint, respectively. The thigh segment was defined proximally with the hip joint center and distally with the knee joint center. The proximal thigh radius was defined as a

quarter of the distance between the left and right greater trochanter markers and distally as half the distance between the medial and lateral femoral condyles. The head segment was defined proximally using the upper mid-torso landmark and distally as the centroid of the anterior and posterior lateral head landmarks.

The trunk and pelvis segments were defined using a cylinder shape. The thigh, shank, foot, upper arm, and forearm segments will be defined as a cone shape. The hands were defined as a sphere shape.

Segment co-ordinate systems: Each segment corresponds to ISB recommendations with the Y-axis defined as superior-inferior, the X- axis as anterior/posterior and the Z-axis orthogonal to both Y and X axes and pointing positively to the right laterally.

Inverse-Kinematic Constraints: Inverse kinematic constraints for the skeletal model were set. The constraints were added between segments, creating a chain, to restrict the relative motion between the segments. The pelvis, setting the base, was rooted to the lab with 6 degrees of freedom (3 translational, 3 rotational). The trunk and thigh segments were rooted to the pelvis and had 3 degrees of freedom each (3 rotational). The shank segments, rooted to the thigh segments, had 1 degrees of freedom (rotational in the flexion/extension axis) and the foot was rooted to the shank with 3 degrees of freedom (3 rotational). For the upper body, the upper arms were linked to the trunk via a ball and socket joint with 3 degrees of freedom (3 rotational). The forearms were linked to the upper arms with 2 degrees of freedom (2 rotational).

Applying model to trials: Once the skeletal model described above was completed, further processing was required including applying the model to each motion trial and filtering the data. For each participant, the skeletal model described above was applied to the static calibration as well as the motion data for each participant, respectively. The motion data were filtered using a 4^{th} order, low pass Butterworth filter with an effective cut off 6 Hz (Winter, 2009) to remove high frequency noise from each signal.

4.6.2 Further 3D motion data conditioning

While most of the post-processing for the 3D motion capture approach was completed in Visual3D, the data needed to be further conditioned to better determine the concurrent validity between the methods. To match the sampling rate of the 2D analyzed data, the outputs obtained from Visual 3D were exported to Python (version 3.7) for further analysis. The discrete outcome measures will be described in detail below. However, the following were exported from Visual3D to Python; knee flexion-extension, hip flexion-extension, trunk flexion-extension, shoulder flexion-extension and abduction-addiction angles, knee and hip angular velocity (flexion-extension axis), as well as joint centers and anatomical landmarks used for calculating whole body COG and base of support. Once in Python, data obtained from the 3D motion capture system was down sampled to 30Hz. Secondly, trials were segmented by identifying start and end points for each task. Segmentation was accomplished by using the left wrist markers for the floor-to-waist lift, where local peaks and troughs were identified by determining absolute minimums and local maximums. For the posture (angular

displacements) and coordination (MARP) the lift start was defined as the local maximum prior to the descent of the wrist marker to pick up the box and the end was defined as the next local maximum wrist marker location following the lift of the box (see Figure 8A as an example). For the COG based outcomes measures (COG to load and FSL), the end of the trial was defined similarly, however, the start of the trial was defined as the peak (Figure 8B). This was done to ensure the COG based outcome measures were calculated for the extension phase of the lift only when the participants were holding the external load. Lastly, once the data was down sampled and segmented, data were interpolated to 101 points (100%) using a cubic spline.

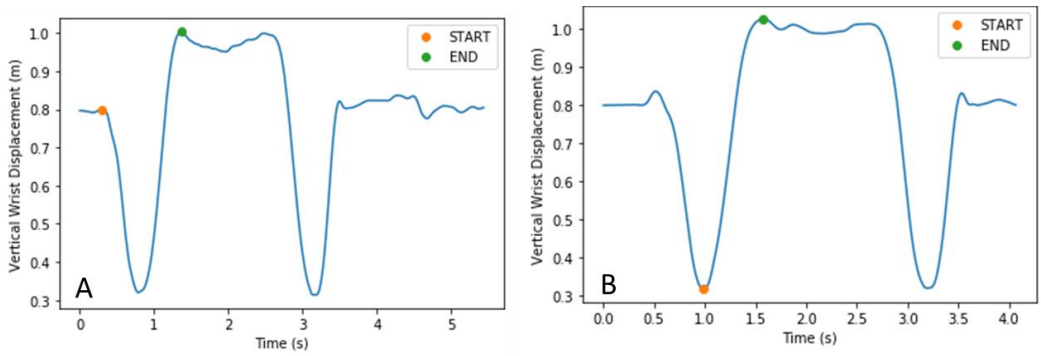


Figure 8: Vertical displacement of the left wrist joint coordinates to define start and end points of the lift. A - The start and end points define the participants flexion phase (descending to pick up the box) and extension phase (ascending lifting the box) and B – start and end points to define the participants extension phase only.

4.6.3 2D Video-based markerless motion capture data conditioning

2D video data of the participants completing the lifting trials in the sagittal plane were collected and subsequently processed through *wrnch*. *Wrnch* software applies a pose-estimation algorithm to predict 3D joint center coordinate positions in the X, Y, Z axes, based on the input 2D video data. However, a limitation in the *wrnch* outputs is that joint centre locations are only expressed in the global coordinate space and do not output segment or joint local coordinate systems. Further, we are aware of each segment endpoint in 3D space, however, there is a lack of information to define the axial rotation. Moreover, we are currently unable to compute local coordinate systems by developing segment-specific orientations. Despite these limitations, we chose a task that remains primarily in the sagittal plane and thus allows for continued analysis and identification of the concurrent validity relative to a 3D motion capture approach. In the global space for the 3D data, the X direction points laterally, the Y-axis points inferiorly and the Z direction points anteriorly in unit-normalized screen-space coordinates (*wrnch*.ai) (Figure 9).

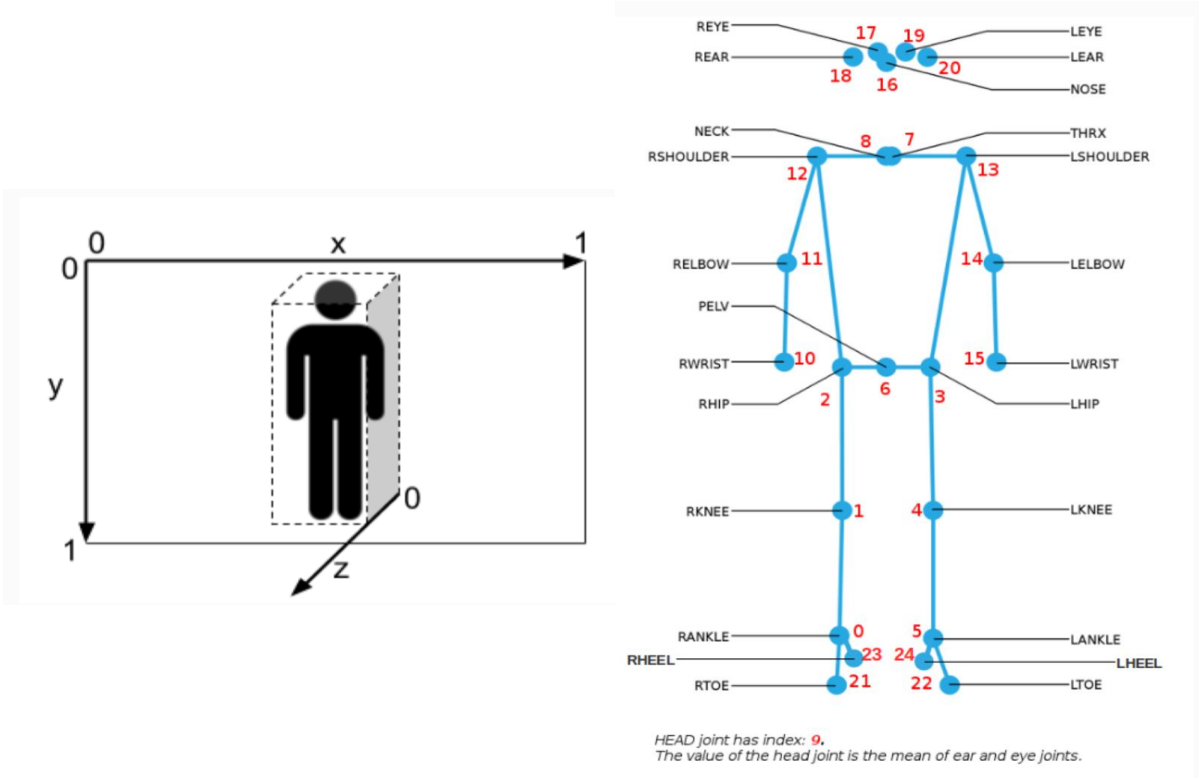


Figure 9: Image on the left demonstrates the raw 3D unit-normalized screen space coordinates outputted from *wrnch*. The image on the right demonstrates the 3D raw joint centers and landmarks outputted from *wrnch*. (*wrnch*.ai)

From the 2D lifting video trials, 3D raw joint centre coordinate data were exported from *wrnch* and opened in Python for data processing. Once imported into Python, data from the 2D pose-estimation software were gap filled (cubic spline interpolation) and then filtered using a 4th order low pass Butterworth filter with a 6 Hz cut-off (Winter, 2009) in the same way as the 3D motion capture data were processed. Trials were segmented in the same manner as the 3D motion data by identifying start and end points based on the wrist joint centre marker. However, due to differences in how coordinate systems were defined between

the 3D motion capture and *wrnch* pose-estimation approach, the following modification was required. Considering the directionality of the superior-inferior axis (0 to 1 superior to inferior) (Figure 10) an absolute maximum was determined for the peak wrist joint centre. Therefore, local minimums were found before and after the absolute peak to define the start and end points, respectively for the posture and coordination based measures (Figure 10). Similar to the 3D motion data, COG based outcome measures were calculated during the extension phase, when the participant was holding an external load. Therefore, the peak of the left wrist joint displacement was used as the start of the lift (Figure 10). Lastly, the data was then interpolated to 100% of the trial using a cubic spline interpolation method like the 3D motion capture outputs.

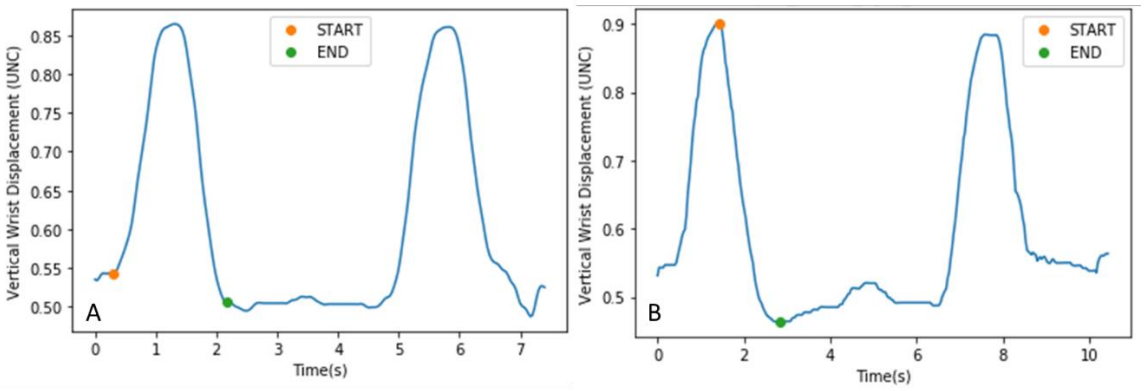


Figure 10: Vertical left wrist displacement used to identify start and end points of the lifting trial. Larger values on the y-axis denote a value closer to the ground. A – the start and end points used for the posture and coordination-based outcomes, and B – the start and end points used for the COG based outcomes.

Table 3: Definitions of model segments between the markerless motion capture pose-estimation model and the 3D motion capture-based model

Segment	Markerless motion capture-based pose-estimation model	3D motion capture-based model
Foot	Ankle joint center to toe anatomical landmark	Ankle joint center to the toe anatomical landmark
Shank	Knee joint center to ankle joint center	Knee joint center to ankle joint center
Thigh	Hip joint center to the knee joint center	Hip joint center to knee joint center
Pelvis	A point medial to the hip markers	CODA pelvis model, using the inter-ASIS distance and sacrum markers
Trunk	Pelvis marker to the mid-point between the "thorax" and "neck"	Upper and lower mid-torso markers
Upper arm	Shoulder joint center to elbow joint center	Shoulder joint center to elbow joint center
Forearm	Elbow joint center to wrist joint center	Elbow joint center to wrist joint center
Hand	See **Balance** below	Wrist joint center to the midpoint of markers placed on the 2nd and 5th metacarpals
Head	Centroid of the eyes and nose	Centroid of the anterior and posterior lateral markers (approximate height at the forehead)

4.6.4 Calculating outcome measures

Posture: Peak intersegmental knee flexion, shoulder flexion, shoulder abduction and trunk segment angles were calculated to represent posture. The posture measures derived from data obtained using both approaches were all calculated for the left side of the body. Differences were presented between the markerless motion capture approach and the 3D

motion capture approach. The above-mentioned posture measures were all calculated in Visual3D from marker data obtained using the motion-capture approach. A ZXY rotation sequence, where flexion-extension occurred about the Z-axis, abduction-adduction about the X-axis and internal-external rotation about the Y-axis was elicited for each joint/ segment angle. Since lifting is primarily a sagittal based movement, it was most appropriate to define the ZXY sequence for this study. Joint and segment angles for the markerless motion capture data were calculated using dot products (Equation 1). The dot product was used due to the limitations in defining 3D segmental orientations when using the *wrnch* outputs.

Knee flexion (-)/ extension (+) for both methods was calculated by defining the orientation of the shank relative to the thigh. The 3D motion capture method defined this angle in the thigh segment local coordinate system, whereas the markerless motion capture approach defined this angle in the global coordinate system. For the markerless motion capture approach the knee joint angle was calculated as the dot product of the vector defining the long axis of the shank and the vector defining the long axis of the thigh. However, to align the directions with the 3D motion data, 180 degrees was subtracted from the final value at each point in time.

Rotation about the shoulder joint for both approaches was referenced with the upper arm relative to the trunk segment. The markerless motion capture approach and 3D motion capture approach defined s*houlder flexion (+)* about the Z -axis and *shoulder abduction (+)* about the X-axis. The markerless motion capture approach for the shoulder angles were

calculated using the dot product of vectors of the upper arm and trunk for flexion and abduction.

Lastly, *trunk segment flexion (-)* and *extension (+)* were calculated with reference to the global coordinate space in both approaches. The trunk segment angle, of the 3D motion capture approach, was calculated with reference to the lab (Y – positive superior, X – positive anterior, Z – positive to the right) where flexion (-) occurred about the Z axis. The trunk segment angle was in relation to the global coordinate system, where flexion (+) was the angular displacement of the trunk segment from the vertical. A vector to represent the vertical was created such that a point vertically from the pelvis was subtracted from a point vertically from the pelvis at the top of the screen (Figure 11). The trunk angle at each point in time was multiplied by -1 to represent trunk flexion as negative and extension as positive, akin to the trunk angle derived from 3D motion capture. Peak knee flexion, peak trunk flexion, peak shoulder flexion and peak shoulder abduction were all calculated by determining the absolute maximum or minimum for each angle. For each angle described above, the peak value was calculated and reported for each lift. The peak was defined as the maximum (shoulder flexion, abduction) or minimum (knee flexion, trunk flexion) angle at any point during the lift.

Equation 1: Equation used to calculate intersegmental joint angles for the knee, hip, and shoulder as well as trunk segment angle at each time point (t). $\vec{a}(t)$ and $\vec{b}(t)$ are the two

segments rotating about the fulcrum of interest. For flexion $\vec{a}$ and $\vec{b}$ are in the X and Y axes and for abduction $\vec{a}$ and $\vec{b}$ are in the Y and Z axes.

$$\theta(t) = \cos^{-1}\frac{\vec{a}(t)\cdot\vec{b}(t)}{|\vec{a}(t)|\,|\vec{b}(t)|} \quad (1)$$

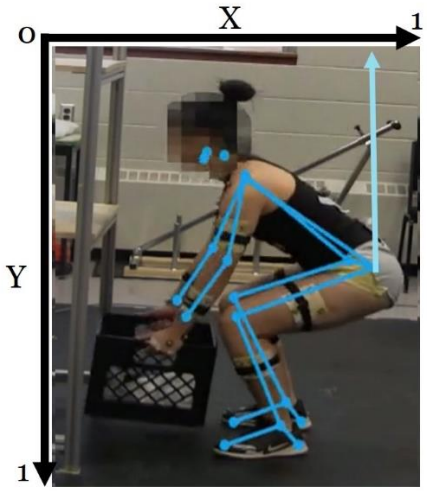


Figure 11: Visual representation of the trunk angle definition from the vertical (light blue line) for the 2D motion capture approach

Balance: To define a measure of balance, the COG was calculated for each lift in the global coordinate space for each method and related to the defined BOS. The 3D motion capture and markerless motion capture-based approach used the same methodology for calculating whole-body COG. For each approach, joint centers and other landmarks were exported from the 3D motion capture (Visual 3D) and from the 2D motion capture (*wrnch)* respectively and used to calculate whole-body COG. This method used the joint centers to

define segment endpoints. Commonly, whole-body COG is calculated by taking the COG of each segment in a specific axis, multiplying the segments' COG by the mass proportion for that segment and then summing them together to determine a whole-body COG (Robertson, 2014). However, for this study we first needed to consider the added mass due to the load applied at the hands (i.e. lifted box). Since the load was acting at the hands, we needed to first determine the coordinates of the hand center using both approaches. The hand center for the data obtained using the 3D motion capture was calculated in Visual 3D using the wrist joint center proximally and the midpoint of the 2nd and 5th metacarpals distally. The *wrnch* model did not output a hand center and therefore, the following calculations were completed to project the hand center from the wrist joint center.

Determining hand center coordinates from the wrnch outputs

The following was completed for the left and right sides of the body. First, the hand length was determined by multiplying the participants height by the length proportion (0.108) using the methods presented by Drillis & Contini, (1966) (Equation 2). However, since the hand length needed to be computed in unit-normalized screen spaced coordinates. The height used was calculated as the absolute maximum distance between the heel marker and the head marker during the trial at any point in time. This is an assumption that was made about participant height using the data output from *wrnch*.

Equation 2: Determining hand length from height (Drillis & Contini, 1966). Equation 2, where, $height = (\max(Head_{SI} - Heel_{SI}))$. SI – superior – inferior axis.

$$\text{Hand length} = 0.108 * \text{height} \quad (2)$$

Second, the length of the hand was multiplied by the hand segment center of mass location proportion from the proximal joint (wrist) (Dempster, 1955) (Equation 3). This length was later used to project from the wrist joint center.

Equation 3: Calculation to determine the hand center of mass distance from the wrist joint center. (Robertson et al., 2014, p. 65)

$$\text{Hand distance} = \text{Hand length} * 0.506 \quad (3)$$

The orientation of the hand was assumed to be in the same direction of the forearm long axis orientation; thus, the orientation of the long axis of the forearm was calculated. A forearm vector was defined as the elbow joint center subtracted from the wrist joint center in each axis (X, Y, Z). Next, the norm of the forearm vector was determined using Equation 4 for each point in time. The forearm vectors in each axis were then divided by the norm of the forearm vector to yield a unit vector describing the orientation of the forearm long axis vector at each point in time (Equation 5).

Equation 4: Equation used to find the norm of the forearm vector. X, Y and Z are the forearm vectors in the associated axis. FA = forearm.

$$\|\mathrm{FA}\| = \sqrt{\overrightarrow{\mathrm{FA}_\mathrm{X}}^{\,2} + \overrightarrow{\mathrm{FA}_\mathrm{Y}}^{\,2} + \overrightarrow{\mathrm{FA}_\mathrm{Z}}^{\,2}} \tag{4}$$

Equation 5: Unit vector calculation of the forearm in each X, Y and Z direction

$$\vec{\mathrm{U}} = \frac{\overrightarrow{\mathrm{FA}_\mathrm{X}}}{\|\mathrm{FA}\|}, \frac{\overrightarrow{\mathrm{FA}_\mathrm{Y}}}{\|\mathrm{FA}\|}, \frac{\overrightarrow{\mathrm{FA}_\mathrm{Z}}}{\|\mathrm{FA}\|} \tag{5}$$

Lastly, the unit vectors, $\vec{U}$, were then multiplied by the hand COG distance from the wrist determined in Equation 3 and then added to the wrist joint center coordinates (Equation 6).

Equation 6: Hand COG coordinates calculated by adding the 'hand distance' (Equation 3) multiplied by the forearm unit vector for each direction, to the wrist joint center in each axis.

$$\begin{aligned}
\mathrm{Hand}_{\mathrm{COG\,X}} &= \mathrm{Wrist}_\mathrm{X} + \left(\vec{\mathrm{U}}_\mathrm{x} * \text{Hand distance}\right) \\
\mathrm{Hand}_{\mathrm{COG\,Y}} &= \mathrm{Wrist}_\mathrm{Y} + \left(\vec{\mathrm{U}}_\mathrm{Y} * \text{Hand distance}\right) \\
\mathrm{Hand}_{\mathrm{COG\,Z}} &= \mathrm{Wrist}_\mathrm{Z} + \left(\vec{\mathrm{U}}_\mathrm{Z} * \text{Hand distance}\right)
\end{aligned} \tag{6}$$

Once the hand center from the *wrnch* outputs was calculated, the rest of the segments' COG coordinate locations were also calculated. Then segments' COG were calculated using a percentage of the segment length from the proximal end. These values were derived from

Dempster's approximations (Equation 7) (Robertson et al., 2014, p. 65) and included in Table 4.

Equation 7: Using the proximal end length of the center of mass segment length (R_{prox}) to determine each segments COG location (Robertson et al., 2014, p. 70) in the X (x_{cog}) and Z (z_{cog}) axes. x or $z_{prox\ or\ dist}$ are the proximal or distal joint centers.

$$x_{cog} = x_{prox} + R_{prox} * (x_{dist} - x_{prox}) \quad (7)$$

$$z_{cog} = z_{prox} + R_{prox} * (z_{dist} - z_{prox})$$

For whole body COG, only the X and Z COG were calculated since the COG was projected onto the ground in relation to the AP and ML directions, respectively. Following the location of the segments' COG in the X and Z direction, the mass proportions applied to each segment were adjusted to account for the added mass in the hands and applied to estimate the whole-body COG. This was done by taking the segments mass proportion (*P*), multiplying it by the participants mass and then dividing it by the whole system mass (body mass plus lifted load) (Equation 8). The external load mass proportion was calculated by dividing the mass of the load by the system mass.

Equation 8: Calculating the adjusted segment proportions to account for the entire system mass. $P_{adjusted}$ = the adjusted segment mass proportion, P_s = the segments original mass proportion, Mass = the participants mass, $Mass_{load}$ the load in the box

$$P_{adjusted} = \frac{(P_s * Mass)}{(Mass_{body} + Mass_{load})} \quad (8)$$

Once the adjusted mass proportions and segments' COG coordinate locations were determined, the segments' COG location in the X and Z axes were multiplied by the mass proportion for the associated segment on the left and right sides where appropriate. The mass proportion of the load was applied to the hand centers. Thus, the whole-body COG using outputs obtained from the two approaches was determined through Equation 9.

Equation 9: Equation used to calculate whole-body COG in the X (AP) and Z (ML) directions. (Robertson et al., 2014, p. 70). P_{adj} is the adjusted mass proportion and X or Z_{cog} is the segments' COG in the associated axis.

$$X_{WBcog} = \sum_{s=1}^{S} p_{adj} * x_{cog} \quad (9)$$

$$Z_{WBcog} = \sum_{s=1}^{S} p_{adj} * z_{cog}$$

Table 4: Segment proximal and distal definition following Dempster (Robertson et al., 2014, p. 60 Table 3.2; Winter, 2009, p.86, Table 4.1) but adapted to accommodate for the limitations in the *wrnch* data outputs. The differences in defining the segment between the two approaches are noted in the table. SS – suprasternal notch, C7 – 7th cervical vertebrae

Segment	**Proximal**	**Distal**	**COM from proximal (R_{prox})**	**Mass Proportion**
Hand	*wrnch:* NA Visual 3D: wrist joint center	*wrnch:* NA Visual 3D: midpoint between 2nd and 5th metacarpals	0.506	0.0060
Forearm	Elbow joint center	Wrist joint center	0.430	0.0160
Upper arm	Shoulder joint center	Elbow joint center	0.436	0.0280
Foot	Ankle joint center	Toe	0.500	0.0145
Leg/ Shank	Knee joint center	Ankle joint center	0.433	0.0465
Thigh	Hip joint center	Knee joint center	0.433	0.1000
Head	Head center (Table 3)	*Wrnch:* Midpoint between "neck" and "thorax" Visual3D: Midpoint between C7 and SS	1.00	0.0810
Pelvis	Midpoint of the right and left hip centers	*Wrnch:* pelvis Visual3D: pelvis center from pelvis segment (Table 3)	*0.*105	0.1420
Trunk	Midpoint of neck and thorax	Pelvis center (Table 3)	0.630	0.3550
				Total = 1.00

Base of Support and Functional Stability Limits

The BOS was defined using the toe and heel anatomical landmarks from the 3D motion capture data and pose-estimation marker data in the global space for each approach respectively (Figure 12). The relationship for the COG to BOS was calculate for the AP and ML directions. Thus, for both approaches the FSL was calculated in the same manner. The FSL was defined as a distance between the centre of BOS (CBOS) to the COG in the appropriate direction expressed as a percent of the distance from the BOS limit in that direction (Equation 10) (Holbein-Jenny et al., 2007). Where the BOS limit in the AP direction was defined as the distance between the midpoint of the left and right toe markers and the CBOS. The BOS limit in the ML direction was the distance between the midpoint of the left toe and left heel marker and CBOS (Holbein-Jenny et al., 2007). The CBOS was calculated as the centroid between the two toe and heel markers, and therefore was assumed to be equidistant in each direction. While interpretation of directionality of the FSL is not evaluated for this study it could still have value on interpretation of the concurrent validity for the approaches. Therefore, if the FSL was defined as positive (> 0%) the interpretation is the COG in the AP direction is located anterior to the CBOS (i.e. towards the toes). Further, if the FSL was negative (< 0%), this would be interpreted as the COG being posterior to the CBOS (i.e. towards the heels or the red and blue 'x' in Figure 12). Similarly, in the ML direction, a positive FSL (> 0%) would mean the COG is to the left of the CBOS (i.e. towards the left foot or the red 'o' and red 'x' in Figure 12) and vice versa for a negative FSL (< 0%). For each direction, the maximum and minimum peak FSL were determined. The final peak FSL

was determined as the absolute maximum or minimum with the directionality preserved. This peak was determined as the absolute deviation of the COG from the CBOS at any point in time during the lift.

Equation 10: Functional stability limit calculation in the anterior-posterior direction (Holbein-Jenny et al., 2007)

$$FSL_{AP} = \frac{(CBOS_{AP} - COG_{AP})}{BOS\ limit_{AP}} * 100 \quad (10)$$

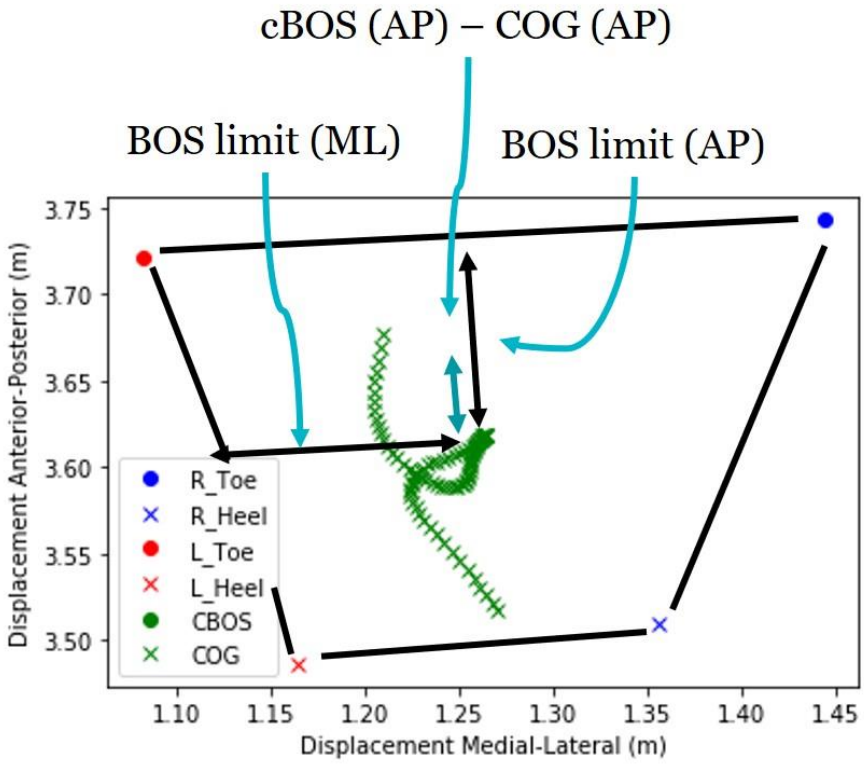


Figure 12: BOS definitions denoted by the black solid lines connecting the toe and heel markers. Demonstrating the BOS limits in the medial-lateral and anterior-posterior

directions and further demonstrating the calculation of the CBOS and COG in the anterior posterior direction.

Distance of the load from the body: High risk movers tend to hold the load further from their body during lifting (Armstrong et al., 2020), demonstrating the importance of quantifying this metric during an FCE. Both data capture approaches utilized the same methodology for this measure. To define the load distance from the participants body, the COG was expressed in relation to the left wrist joint centre in the AP direction. Since the markerless motion data are represented in unit-normalized screen space coordinates, both the 3D motion capture and 2D markerless motion capture data and pose-estimation were normalized to percent height for this outcome measure. The distance between the head marker and left heel marker was calculated frame-by-frame, where the maximum distance was retained to define height. The COG and wrist difference were then divided by the height measure for normalization across trials and methods (Equation 11).

Equation 11: Determining the location of the wrist marker in the anterior-posterior direction relative to the COG in the anterior-posterior direction as a percentage of the height of the individual. AP = anterior-posterior, SI = superior-inferior, WJC = wrist joint

$$COG\ to\ wrist = \frac{(WJC_{AP} - COG_{AP})}{(\max(Head_{SI} - Heel_{SI}))} * 100 \quad (11)$$

Lifting coordination: Subjectively, clinicians look for coordination and smoothness in movements (Allison et al., 2018). To measure this phenomenon, mean absolute relative phase angles (MARP) of the hip-knee were calculated for the flexion (lowering to grasp the box) and extension (raising the box up to the shelf) phases of the lift, respectively. A velocity-based procedure (Limerick-Burgess et al., 1993) was used to calculate MARP. The following steps were completed to calculate MARP angles (Figures demonstrating how the data were manipulated in each step as necessary to calculate MARP are described in Appendix B (p.121). All steps were completed for both approaches and only minor differences are presented between the two approaches for the coordination measures. These differences will be described as necessary when explaining the methodology below.

First, the angular displacement and angular velocity (central difference method) from Visual 3D were extracted for analysis from the 3D motion capture-based method. Similar to the postural measures the angular displacement and velocity calculated from the data obtained using both approaches are in reference to the left side of the body. However, the joint angular displacements and velocities were calculated differently for the markerless motion capture and pose-estimation based data. The left hip joint angle was calculated using the dot product (Equation 1), with two vectors defined by the thigh segment and trunk segment (Table 3). Knee flexion was calculated previously and described above. For the markerless motion capture based data, angular velocity was calculated using the central difference method (Equation 12) for the knee and hip flexion and extension angles.

Equation 12: Joint angular velocity of the 2D joint angular displacement of the knee and hip using the central differentiation technique, (t) is each point in time and θ is the angular displacement.

$$\omega = \frac{\theta(t+1) - \theta(t-1)}{2(\Delta t)} \quad (12)$$

The following description to calculate MARP from both data capture methods are continued below. Both approaches used the below steps without any differences in the calculations. Second, the knee and hip angular displacements and angular velocities were normalized to eliminate magnitude differences between the two joints (Galgon & Shewokis, 2016). The normalization methods can be seen in Equation 13 and Equation 14. This technique limits the flexion/ extension angular displacements and velocity values to between -1 and +1 depending on the magnitude of the angle. Angular displacement was normalized such that the minimum of the original values equals -1 and the maximum value equals +1 (Lamb & Stöckl, 2014). By doing this, the 0 value will be the midway value between the maximum and minimums (Lamb & Stöckl, 2014). However, a different normalization equation is used for angular velocity to preserve the 0 meaning for angular velocity (Lamb & Stöckl, 2014; (Burgess-Limerick et al., 1993).

Equation 13: Equation to normalize angular displacement. θ_{norm} is the normalized angular displacement, θ(t) is the angular displacement at time point (t) (Lamb & Stöckl, 2014)

$$\theta_{norm} = 2 * \left(\frac{\theta(t) - \min(\theta)}{\max(\theta) - \min(\theta)} \right) - 1 \quad (13)$$

Equation 14: Normalization method of angular velocity. ω_{norm} is the normalized angular velocity. ω(t) is the angular velocity at time point (t) (Lamb & Stöckl, 2014).

$$\omega_{norm} = \frac{\omega(t)}{\max|(\omega)|} \quad (14)$$

Following normalization for each of the knee and hip angular displacements and velocities, the relative phase angle for each was calculated, and subsequently subtracted from one another. The phase angle of the signal at each time point (t) was then calculated based on the normalized phase plane portraits as determined above (Equation 15) (Lamb & Stöckl, 2014). Then to calculate a CRP (Hip – Knee), the relative phase of the distal joint (ϕ_{knee}) was subtracted from the relative phase of the proximal joint (ϕ_{hip}) (Albert, Wrigley, & McLean, 2008; Galgon & Shewokis, 2016; Lamb & Stöckl, 2014) (Equation 16). Lastly, once the continuous relative phase of the hip-knee was calculated, the mean value of the flexion and extension phase were calculated, respectively (Equation 17). The flexion and extension phases of the lift were separated based on the minimum peak wrist displacement. The mean

value was calculated instead of the peak to capture a holistic view of coordination for each phase, respectively.

Equation 15: Phase angle, ϕ, at each time point (t) calculated by taking the inverse tangent of the normalized angular velocity divided by the normalized angular displacement (Albert et al., 2008)

$$\phi(t) = tan^{-1}\left(\frac{\omega_{norm}}{\theta_{norm}}\right) \tag{15}$$

Equation 16: Continuous relative phase angle (CRP) with a distal – proximal relationship. The inverse tangent at each point in time is taken from the relative phase of the hip minus the relative phase of the knee (Lamb & Stöckl, 2014; Galgon & Shewoski, 2016)

$$CRP\ (t) = \phi_{hip(t)} - \phi_{knee(t)} \tag{16}$$

$$CRP\ (t)$$

$$= tan^{-1}\left(\frac{\omega_{hip,norm}(t) * \theta_{knee,norm}(t) - \omega_{knee,norm}(t) * \theta_{hip,norm}(t)}{\theta_{hip,norm}(t) * \theta_{knee,norm}(t) + \omega_{hip,norm}(t) * \omega_{knee,norm}(t)}\right)$$

Equation 17: Mean absolute relative phase (MARP) angles for the flexion and extension phases were calculated separately. N is the number of time points in the flexion or extension phases (Galgon & Shewoski, 2016).

$$MARP = \frac{\sum CRP\ (t)}{N} \tag{17}$$

4.7 Statistical Analysis

Bland-Altman plots (Bland & Altman, 1986) were used to quantify agreement between the two methods. Bland-Altman (B-A) plots require that the difference between the two measures (outcome measure generated using 3D motion capture minus outcome measure generated using video-based markerless motion capture and pose-estimation) were calculated as well as the mean of the two measures ((outcome measure generated using 3D motion capture + outcome measure generated using 2D pose-estimation) / 2). The mean of the difference, $\bar{d}$, and standard deviations (*sd)* of the differences were calculated to determine if the outcome measure generated using 2D pose-estimation method under- or over-estimated the outcome measure generated using the 3D motion capture data. Limits of agreement (LOA) for each outcome measure comparison were calculated to assume 95% of the differences would lie within the LOA (Equation 18) (Bland & Altman, 1986). The 95% LOA aids in interpreting how well two methods of measurement agree, where the narrower the range of the LOA the better the agreement (Myles & Cui, 2007).

Bland-Altman analyses were interpreted using three metrics. First, the magnitude of the mean difference, $\bar{d}$, was reviewed to determine if there was consistent systematic error between the two techniques. Consistent systematic error occurs when one approach consistently over- or underestimates values when compared to the second approach. Secondly, the LOA and LOA bandwidth (*sd* * $\pm$1.96) were used to interpret the range in the differences between the two methods. Lastly, increasing or decreasing trends in the dataset

allowed for comparison between the methods. To analyze trends, a line of best fit was overlaid on the Bland-Altman plot data quantify emerging trends (Giavarina, 2015). These trends described if the differences were associated with the mean of the two methods. As an example, a positive linear trend on the Bland-Altman plots (upwards to the right) can potentially be interpreted as an increase in the differences as an increase in the mean between the methods.

Further, Giavarina, (2015) point out that it would be opportune to calculate the confidence interval of the mean difference to see how precise the measurements are. Specifically, the 95% confidence interval of the mean difference illustrates the magnitude of the systematic difference by describing possible error in the estimates due to a sampling error (Giavarina, 2015). In the absence of pre-determined values defining 'agreement', plotting the 95% confidence interval (Equation 20) of the mean difference allows conclusion of if there are systematic differences or not. We can assess systematic differences if the confidence intervals of the mean difference do not overlap on the plot with a value of 0 on the y-axis (Giavarina, 2015). This is measured by multiplying the standard error of the mean difference (Equation 19) by the t-value (t distribution with n-1 degrees of freedom) (Equation 14) (Giavarina, 2015).

The 95% confidence intervals of the mean difference are calculated using the standard error of the differences. This is interpreted as any confidence interval values that are not surrounding a difference of 0 units between the approaches demonstrates systematic error or

differences. The upper and lower LOA, representing 95% of where the differences would lie, demonstrate the span of calculated differences between the two approaches studied.

Equation 18: Limits of agreement using a 95% confidence interval. LOA (upper) and LOA (lower) represent the upper and lower LOA respectively (McAlinden et al., 2011)

$$LOA_{upper} = \bar{d} + (1.96 * sd) \quad (18)$$

$$LOA_{lower} = \bar{d} - (1.96 * sd)$$

Equation 19: Standard error of the mean difference calculation (Giavarina, 2015). Se is the standard error, n, is the number of trials and sd, is the standard deviation of the differences.

$$se = \sqrt{\frac{sd^2}{n}} \quad (19)$$

Equation 20: 95% confidence interval of the mean difference. The standard error, se, is multiplied by t-distribution value for n- 1 degrees of freedom and added or subtracted from the mean difference ($\bar{d}$). With n -1 = 59, ~t-value = 2.000 (95% confidence interval, two-tailed α = 0.05 (Carlton & Devore, 2017)

$$confidence\ intervals = \bar{d} +/ -(se * t) \quad (20)$$

The B-A analysis assumes the difference in the dataset $\bar{d}$, follows a normal distribution (Bland & Altman, 1995). Therefore, Shapiro-Wilks analysis was first performed on the differences $\bar{d}$, for each value respectively to accept or reject this assumption. Where there was not a normal distribution ($p < 0.05$), outliers were identified by determining data points that were 1.5 * interquartile range (IQR = 3rd quartile – 1st quartile) below or above the 1st and 3rd quartile range, respectively. While outliers in the dataset can skew the distribution, removing outliers was still avoided for this agreement study. However, it was determined that removing outliers did not affect the interpretation of the agreement. The test for normality using Shapiro-Wilks were completed in SPSS (Version 26.0, IBM Corporations, Armonk, NY) and the Bland-Altman analysis was completed in Microsoft® Excel® (Microsoft 365).

5. Results

5.1 Normality

Using the Shapiro-Wilks test for normality the difference values between the two methods for the peak knee flexion, peak shoulder abduction, peak COG to load measure, FSL AP, the MARP flexion phase and MARP extension phase all satisfied the normality assumption. However, peak trunk flexion, peak shoulder flexion and the FSL ML calculated differences did not satisfy the normality assumption. The number of outliers determined for the measures that did not satisfy a normal distribution are presented in Table 5. Outliers were determined by finding the values outside 1.5 * interquartile range from the 1st or 3rd quartile, respectively. Upon further testing, outliers were removed (as suggested) such that the differences assumed a normal distribution. However, the removal of outliers for the peak trunk flexion and peak shoulder flexion to ensure a normal distribution did not affect the interpretation of the agreement between the two methods. As a result, the Bland-Altman analysis results were based on all data, despite instances where data did not meet the normality assumption, as those instances did not affect the overall outcome or interpretation of results.

Table 5: Outcome measures with a $p < 0.05$ using the Shapiro-Wilks test for normality and the number of outliers determined.

Outcome measure	# Outliers
Peak trunk flexion	3
Peak shoulder flexion	7

FSL ML	10

5.2 Summary of results

The mean and standard deviations, maximum values and minimum values are represented in Table 6 for the outcomes obtained from the markerless motion capture and pose estimation as well as from the outcomes obtained using the marker data from the 3D motion capture based approach.

Table 6: Mean ± standard deviations, minimum values and maximum values from the outcome measures calculated using the data obtained from the two different approaches. 2D = outcome measures from 2D markerless motion capture and pose estimation. 3D = outcome measures calculated from data obtained using the 3D motion capture approach.

Outcome	**Mean ± standard deviation**		**Minimum Value***		**Maximum Value****	
	2D	3D	2D	3D	2D	3D
Peak knee flexion (°)	-79.0 ± 22.56	-100.7 ± 18.9	-127.0	-146.9	-38.9	-66.1
Peak trunk flexion (°)	-42.9 ± 13.3	-65.0 ± 9.8	-83.3	-87.3	-21.9	-44.3
Peak shoulder flexion (°)	48.6 ± 12.6	58.3 ± 8.6	30.1	40.0	83.5	71.4

Peak shoulder abduction (°)	22.3 ± 7.1	26.8 ± 8.2	11.5	7.6	43.7	50.6
MARP (flexion phase) (°/frame)	-4.5 ± 8.0	-6.7 ± 9.4	-21.6	-30.1	13.1	16.3
MARP (extension phase) (°/frame)	15.7 ± 11.2	17.6 ± 11.5	-4.7	-1.0	48.8	47.9
Peak COG to load distance (% height)	7.3 ± 1.6	22.3 ± 3.6	3.8	13.9	13.5	33.9
Peak FSL AP (% AP BOS limit)	94.5 ± 36.4	-86 ± 28.8	33.8	20.6	165.6	-139.2
Peak FSL ML (% left BOS limit)	See Figure 13					

* Minimum value for the knee flexion and trunk flexion represents a larger flexion angle. Minimum value for shoulder flexion and abduction represents a smaller flexion or abduction value. Minimum value is the absolute minimum (-) or (+), demonstrating the smallest deviation of COG from CBOS in either direction.

** Maximum values for FSL represent the absolute maximum peak FSL in either direction, therefore it is the maximum deviation of the COG from the CBOS in either direction.

The mean differences, $\bar{d}$, standard deviations of the differences, *sd*, the upper and lower LOA using a 95% confidence interval, and the LOA bandwidth are presented in Table 7 below.

Table 7: Mean difference, standard deviation of the mean difference, the upper and lower LOA and LOA bandwidth to support the Bland-Altman analysis

	Mean difference (mean difference), $\bar{d}$	**± Standard deviation of $\bar{d}$, *sd***	**Upper LOA ($\bar{d} + 1.96 * sd$)**	**Lower LOA ($\bar{d} - 1.96 * sd$)**	**LOA Bandwidth (*sd*1.96*)**
Peak Knee Flexion (deg)	-21.7	10.5	-1.1	-42.3	± 20.6
Peak Trunk Flexion (deg)	-22.1	7.3	-7.8	-36.4	± 14.3
Peak Shoulder Flexion (deg)	9.7	9.8	28.9	-9.3	± 19.1
Peak Shoulder Abduction (deg)	4.5	10.7	25.5	-16.6	± 21.0
MARP Flexion phase (%)	-2.3	9.6	16.5	-21.1	± 18.8
MARP Extension phase (%)	1.9	10.5	22.47	-18.6	± 20.6

Peak FSL AP (%)	-180.6	38.5	-105.1	-256.1	± 75.5
Peak FSL ML (%)	See Figure 13				

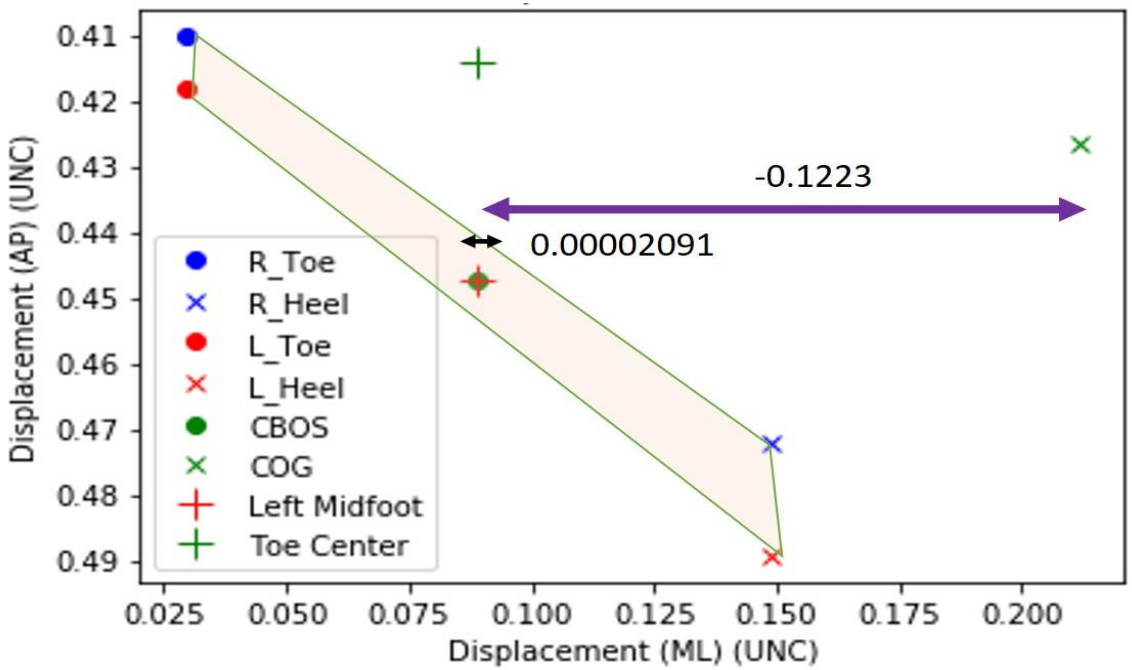


Figure 13: This figure explains why the FSL ML were omitted. In some cases, the BOS limit (small black arrow) was very close to 0 due to the toe and heel estimations provided by *wrnch*. When the BOS limit was negligible, and the COG was estimated to

be beyond the BOS (shaded area), this led to values that grossly skewed the dataset and lead to uninterpretable results. Purple arrow = the numerator when defining the FSL ML (CBOS – COG) (Equation 10).

5.3 Postural measures

The Bland-Altman plots for the posture measures are presented in Figure 14. Markerless motion capture and pose-estimation data using *wrnch* underestimated the peak knee and peak trunk flexion angles by greater than 20 degrees, on average. The pose-estimation outcomes underestimate shoulder flexion peak angles, on average, but only about 5 degrees of underestimation in the shoulder abduction axis. Considering the LOA, specifically the LOA agreement bandwidth, the peak angles were approximately $\pm$ 20 degrees on either side of the mean difference measure for the knee and shoulder angles. However, the trunk angle demonstrated a LOA bandwidth that was smaller, approximately $\pm$ 14 degrees. Peak knee flexion and peak trunk flexion demonstrate a slight trend demonstrating that when the mean peak angle becomes smaller (closer to 0) the difference becomes larger (large negative difference). The peak shoulder flexion angle does not demonstrate much of a trend, however, a larger mean value between the measures tends towards a more negative difference between the two values. Shoulder abduction angles did not demonstrate a trend in the dataset between the two approaches.

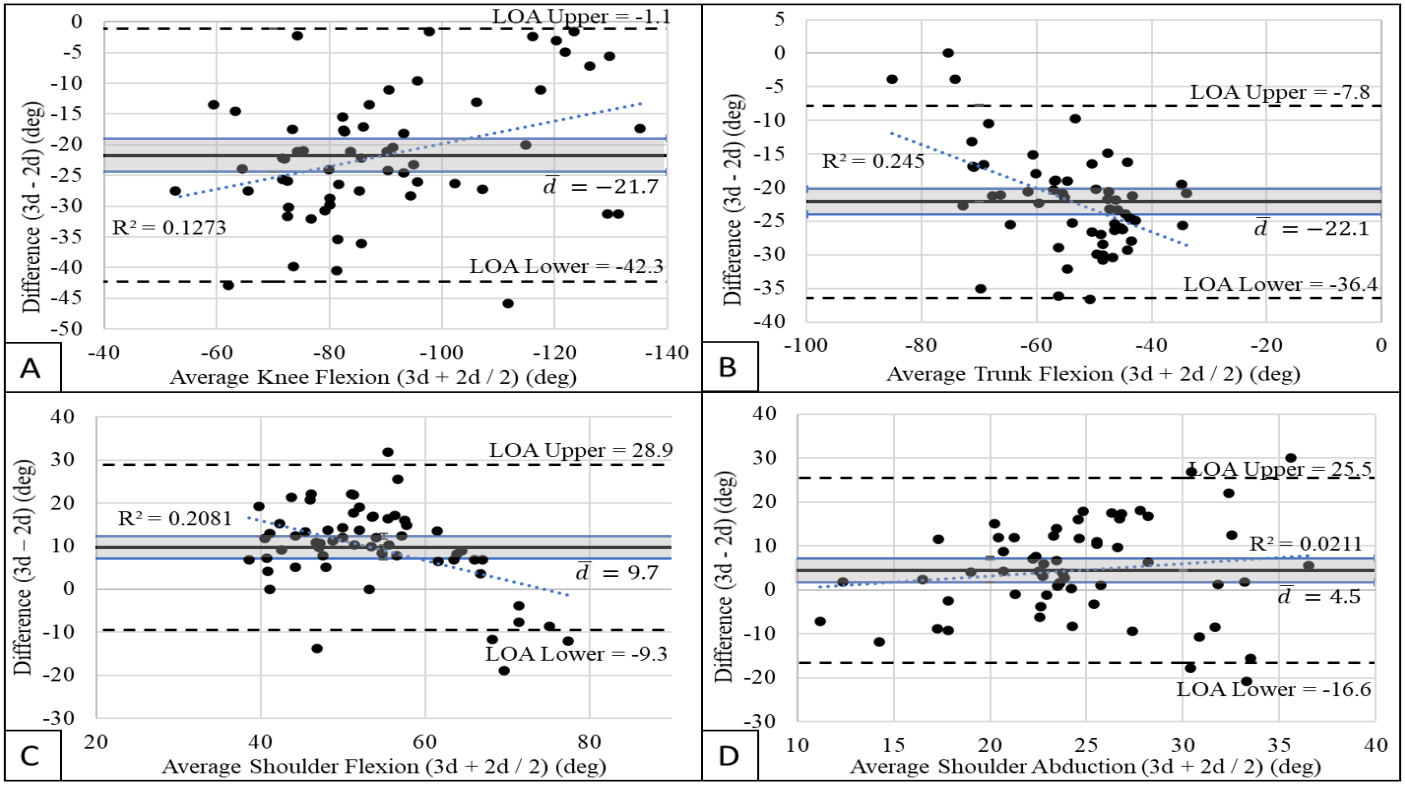
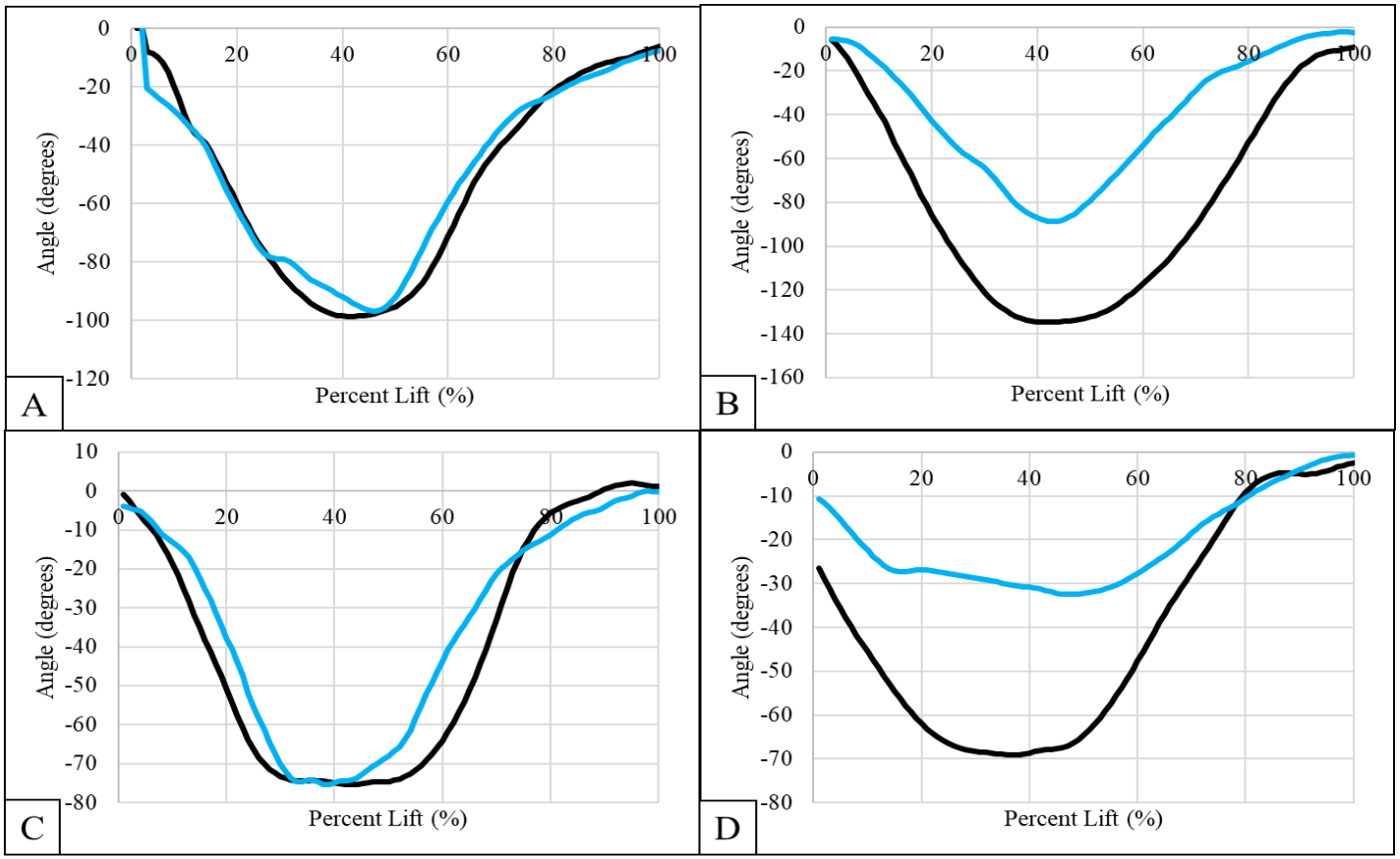


Figure 14: Bland - Altman plot of the two measures calculated from the two approaches, respectively. Solid black line = mean difference, Dashed lines = upper and lower LOA with 95% CI. A – peak knee flexion; B – peak trunk flexion; C- peak shoulder flexion; D – peak shoulder abduction. The grey shaded area outlined by blue lines represents the 95% confidence interval of the mean difference (Equation 20). The blue dotted line is the linear regression line (R^2).

To give a representation of the extremes that were generated from the two approaches, Figure 15 demonstrates temporal waveforms representing the smallest peak difference identified in the dataset versus the largest peak difference identified.



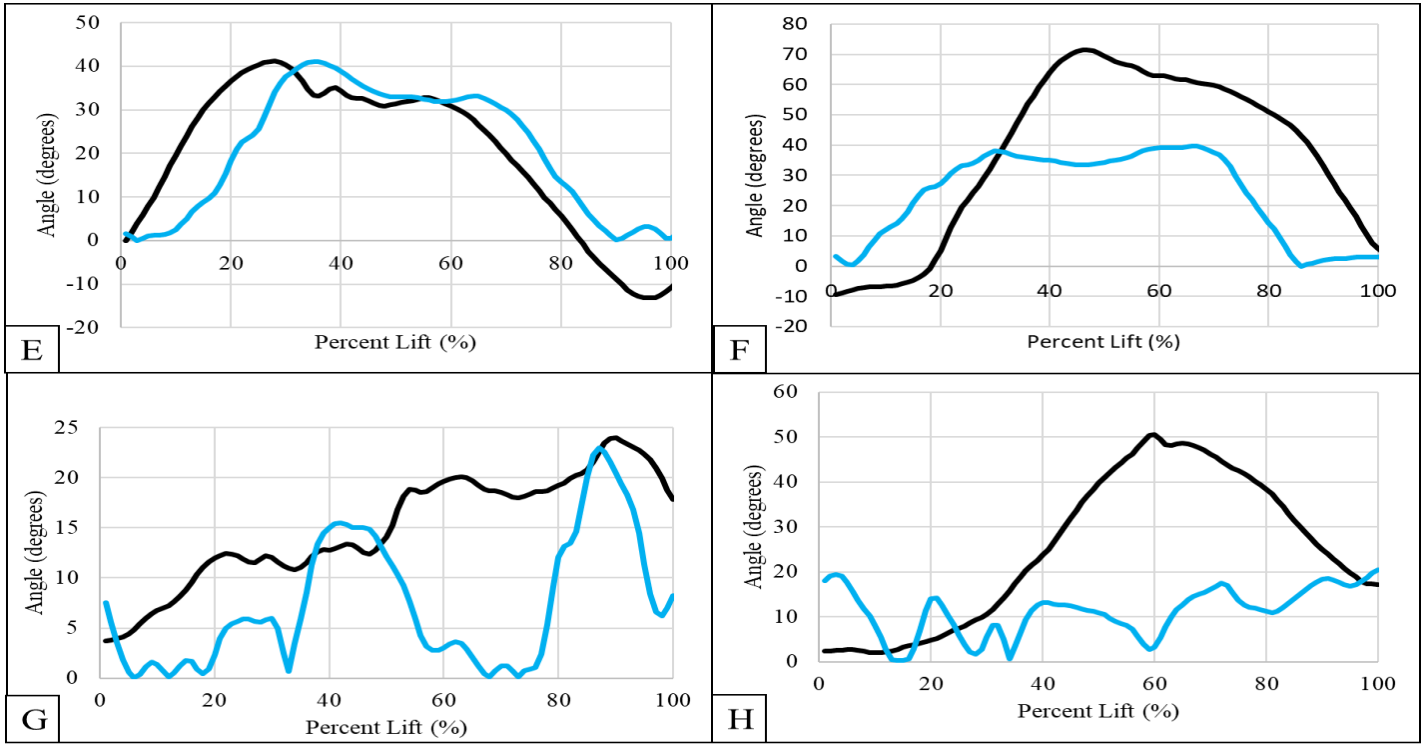


Figure 15: Figures demonstrating the smallest peak difference between the two approaches and associated waveforms (left side) and the largest peak difference between the two approaches and the associated waveforms (right side) A-B = knee flexion, C-D = trunk flexion, E-F = shoulder flexion, G-H = shoulder abduction. Black = 3D motion capture approach, Blue = 2D markerless motion capture + pose-estimation approach.

5.4 Balance Measures

The outcome measures relating to balance include FSL in the AP and ML directions. The FSL determined through data obtained from the markerless motion capture and pose-estimation demonstrated an anterior COG relative to the CBOS 98% of the time. The values obtained from the 3D motion capture data had a peak FSL where the COG was posterior to the CBOS 100% of the time (See Figure 17). The values for the FSL in the AP direction did not demonstrate a trend in the dataset.

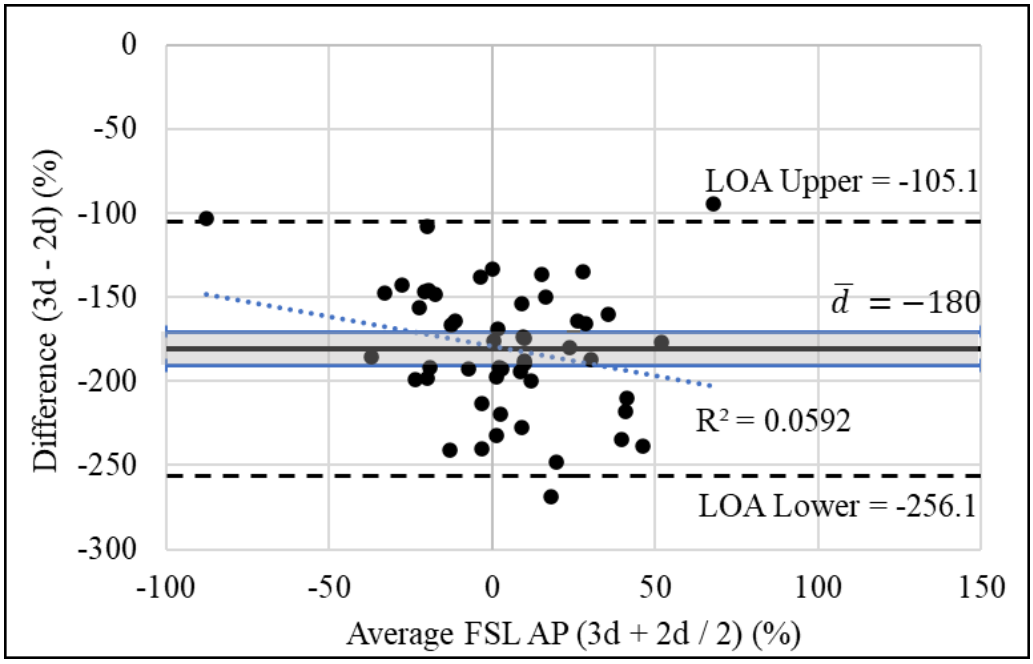


Figure 16: Bland - Altman plot of the two measures calculated from the two approaches, respectively. Solid black line = mean difference, Dashed lines = upper and lower LOA with 95% CI. The grey shaded area outlined by blue lines represents the 95% confidence interval of the mean difference (Equation 20). The blue dotted line is the linear regression line (R^2).

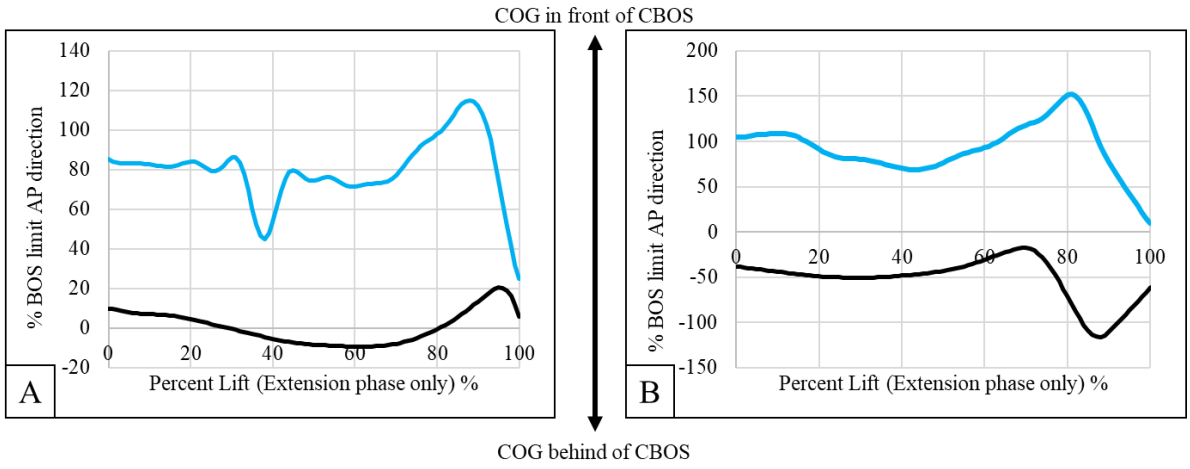


Figure 17: Figure representing the smallest (A) peak difference in the FSL and largest (A) peak difference in the FSL AP and the associated waveforms over the entire lift calculated from the 2D markerless motion capture + pose-estimation (blue) and 3D motion capture method (black). As the values become larger (> 0), the COG creeps towards the front of the BOS (towards toes) and as the values become smaller (<0), the COG moves posterior to the CBOS and towards the heels.

5.5 Distance of the load to the COG

The Bland-Altman plot for the peak distance of the load to the COG between the 2D pose-estimation and 3D motion capture analysis is presented in Figure 18. The mean difference is around 15 % height, thus presenting underestimation of the 2D markerless motion data and pose-estimation in comparison the 3D motion capture-based outputs. LOA bandwidth remained in the negative axis at approximately $\pm$ 5 % height. This measure demonstrated a trend. When the average of the two measures increases, the difference between the two measures also increases.

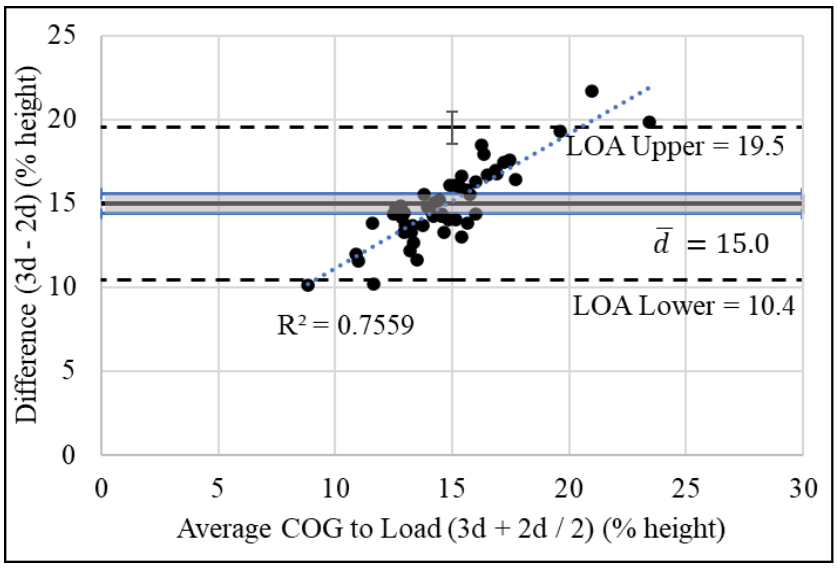


Figure 18: Bland - Altman plot of the two measures calculated from the two approaches, respectively. Solid black line = mean difference, Dashed lines = upper and lower LOA with 95% CI. The grey shaded area outlined by blue lines represents the 95% confidence interval of the mean difference (Equation 20). The blue dotted line is the linear regression line (R^2).

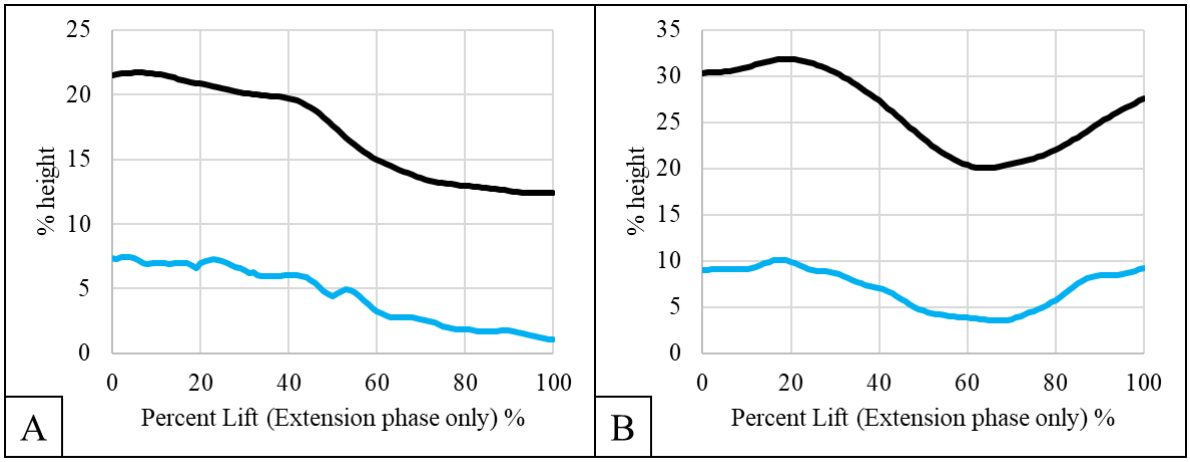


Figure 19: COG to load time series waveform during the extension phase of the lift only. A - the waveform of the COG to load measure representing the smallest peak difference and B - the waveform of the COG to load measure representing the largest peak difference between the two approaches

5.6 Lifting Coordination

Lifting coordination values represented using MARP for flexion (bending to pick up the box) and extension (lifting the box to the shelf from the ground) between the two methods analyzed are presented in Figure 20. The mean differences did not show a pattern of over- or underestimation in the dataset with values averaging close to 0 degrees. However, the LOA bandwidth spanned approximately $\pm$ 20 degrees of over- and underestimation of values calculated from the markerless motion capture and pose-estimation. The MARP in the flexion phase demonstrated a small positive trend indicating the mean differences tend towards values >0 as the average values between the two also tend to >0. The MARP of the extension phase did not demonstrate any trends in the dataset. Figure 21 demonstrates examples that

represent the smallest and largest differences found in the dataset when calculating MARP using the data obtained from the markerless motion capture and pose estimation approach versus the 3D approaches. The Figures are representative of the flexion (A-B) and extension (C-D) phases of the lift.

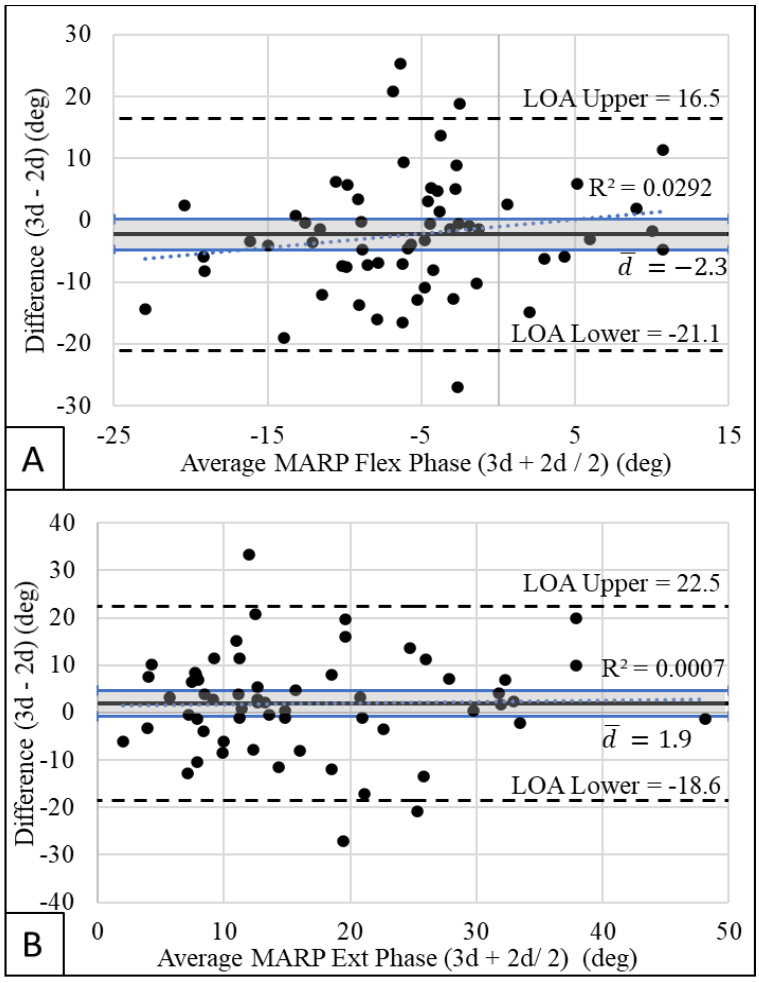


Figure 20: Bland - Altman plot of the two measures calculated from the two approaches, respectively. Solid black line = mean difference, Dashed lines = upper and lower LOA with 95% CI. MARP values of the flexion (A) and extension (B) phases of the lift. The

grey shaded area outlined by blue lines represents the 95% confidence interval of the mean difference (Equation 20). The blue dotted line is the linear regression line (R^2).

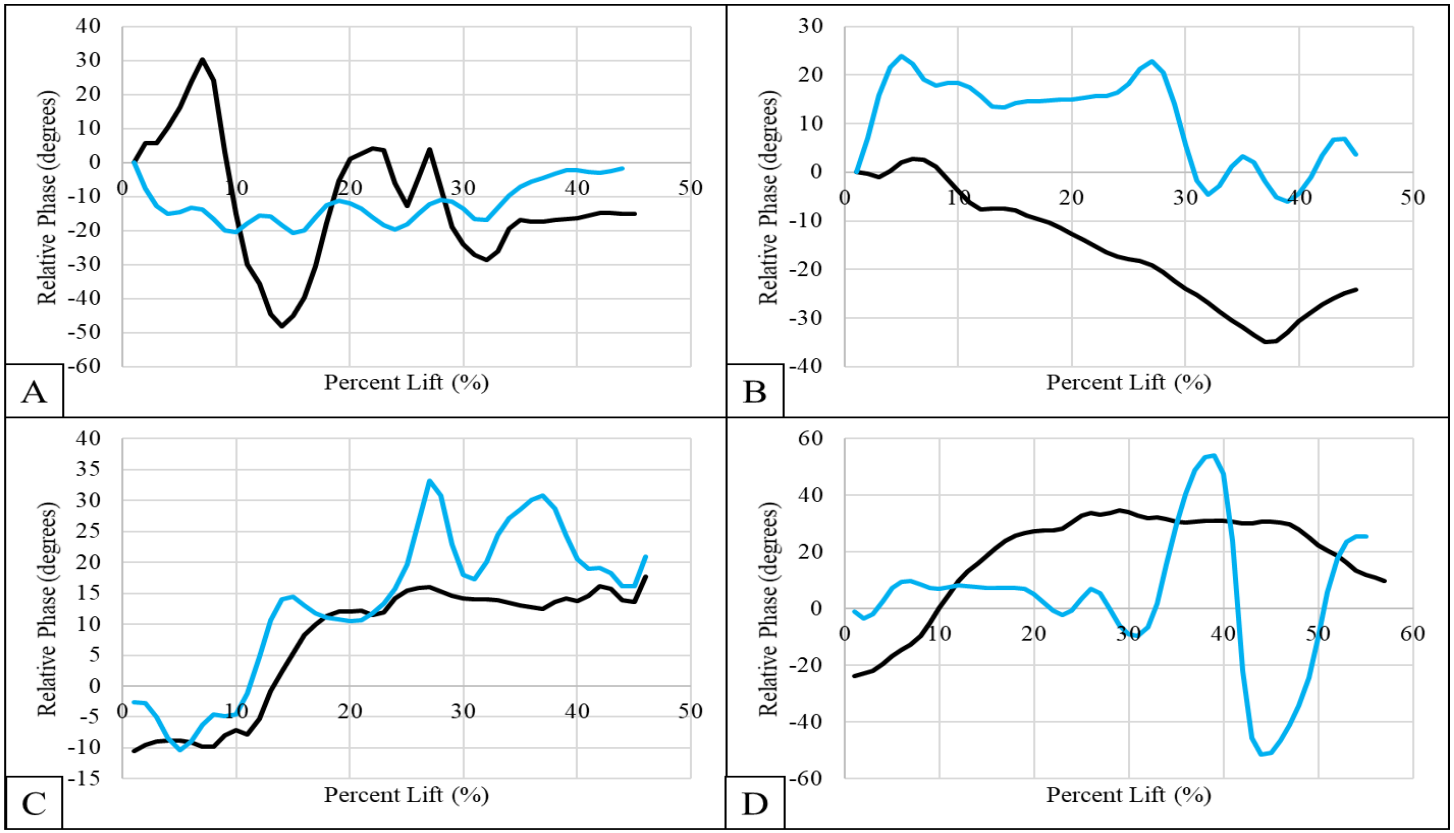


Figure 21: Examples of the smallest MARP difference calculated between the two approaches (left side) and largest difference when calculating MARP using the two approaches (right side) for the flexion phase (A - B) and extension phase (C - D) of the lift. Black representing the MARP from the 3D approach, and blue represents the MARP of the 2D markerless approach.

6. Discussion

6.1 Interpreting Agreeability

6.1.1 Posture and coordination outcome measures

This study investigated the agreement of FCE relevant outcome measures calculated using kinematics obtained from markerless motion capture data and 3D motion capture-based data. The Bland-Altman method, designed to evaluate the agreement between two methods, was used in this study to quantify concurrent validity of the two approaches. However, it is first important to determine what difference between the two measures is acceptable. Focusing on an interpretation of the absolute measures of agreement using LOA, when considering joint angle data, a LOA bandwidth of greater or less than $\pm$ 9.8 degrees has been suggested to represent poor agreement (Keogh et al., 2019). Similarly, Mjøsund et al., (2017) considered LOA to be acceptable if upper and lower LOA values were less than 5 degrees in either direction (LOA bandwidth = 10 degrees). While they suggested that these limits were not necessarily based on empirical evidence, it is argued that they seem reasonable for clinical assessment of lumbar movement reflected in the "American Medical Association's guide to measuring spinal range of motion" (Mjøsund et al., 2017; Nitschke, Nattrass, Disler, Chou, & Ooi, 1999). Therefore, regarding clinically relevant outcomes, the measurements need to be accurate enough to depict these differences.

For this study, we hypothesized that postural angles in the sagittal plane and MARP values as a measure of coordination were expected to demonstrate agreement between the

two approaches. However, considering the generalized interpretation of LOA bandwidth (~ ± 10 degrees), our data do not agree and thus do not support hypothesis 1A. Shoulder abduction angle was expected to demonstrate poor agreement between the approaches. Utilizing our LOA bandwidth interpretation, the data support hypothesis 1B. On average, there was a LOA bandwidth for the posture outcome measures (peak angles) of approximately ±20 degrees. This is more than double the acceptable value of ± 9.8 degrees (Keogh et al., 2019) for clinically relevant postural assessment. Overall, the rejection of Hypothesis 1A in regard to the angular outcome measures is inferred from an underestimation in of angles when using kinematics obtained from 2D video and a pose estimation algorithm compared to those calculated using the 3D motion capture method. Frontal plane postural measures (shoulder abduction) did not demonstrate consistent systematic error thus allowing a different interpretation from the sagittal based postural angles.

For the coordination measures, MARP of the flexion and extension phases of the lift, the same criteria for agreement was applied as above for the postural measures. Due to the results and LOA bandwidth for the coordination measures, data do not support Hypothesis 1A when considering the MARP outcome measure.

So, we now ask, why? Why did the two methods disagree? There are different sources of error that may help to explain the lack of agreement observed. Automated tracking methods, while novel, also have errors or downfalls within the methodology that can lead to systematic error.

First, we can attribute systematic errors present between the two approaches to the methodology behind the estimation of joint centers. As an example, joint center estimation using 3D motion capture of the knee joint is completed by finding the midpoint between the medial and lateral femoral condyles on the femoral mechanical axis (Grood & Suntay, 1983). Further, Bell et al., (1989, 1990) demonstrated the prediction of the hip joint center through mathematical assumption using the superficial anatomical markers. Therefore, in 3D motion capture analysis we can predict joint centers, internal to the surface landmarks. Whereas 2D pose-estimation, likely estimates a projection of these joint centers on the surface of the participant in space. Misidentification of joint centers can influence segment definitions and then further intersegmental joint angles. However, we can only speculate that the automated tracking of the joint centers is a limitation causing systematic error within the dataset presented, as we do not know the exact approach used to train the pose estimation algorithm to identify joint centers.

It is possible that the pose estimation algorithm was trained by using data that was first manually labeled (i.e., joint centers were digitized by hand). Previous studies, using manual labelling for 2D motion data for the purpose of calculating angular kinematics of the lower extremity (e.g. Alahmari et al., 2020; Maykut et al., 2015; Ortiz et al., 2016; Schurr et al., 2017) provided differing results in the agreement between 2D and 3D motion capture methodologies. One source of error presented is the utilization of manual labelling still images. From these studies we can infer potential agreement and disagreement between

methodologies, and that sources of error are a potential result of where the joints are labelled in space and in relation to one another.

When training a model to track joint locations and skeletal landmarks, the inputs and corrections that are manually applied can affect the pose-estimation results. As an example, if a model or algorithm is trained on gait videos, then it is likely the model will not fit as well or present as accurate results on a subject performing a lifting task. In addition, one hypothesis for how pose-estimation determines joint centers is through identification of facial features first to detect the head and then working distally to label the joints and landmarks down the chain. As an example, our lab investigated the difference in *wrnch* joint outputs when using a input where the face was blurred versus unblurred. There were visually different pelvis vertical displacements when running the blurred facial video of the lift and the non-blurred facial video of the same lift through *wrnch* (Figure 22). Although we did not use blurred videos as inputs for this study, the video was recorded in the sagittal plane thus limiting the facial recognition potential. Therefore, perhaps in pose-estimation algorithms, facial recognition is required to be able to track anatomical landmarks for a variety of different tasks to be successful and generalizable.

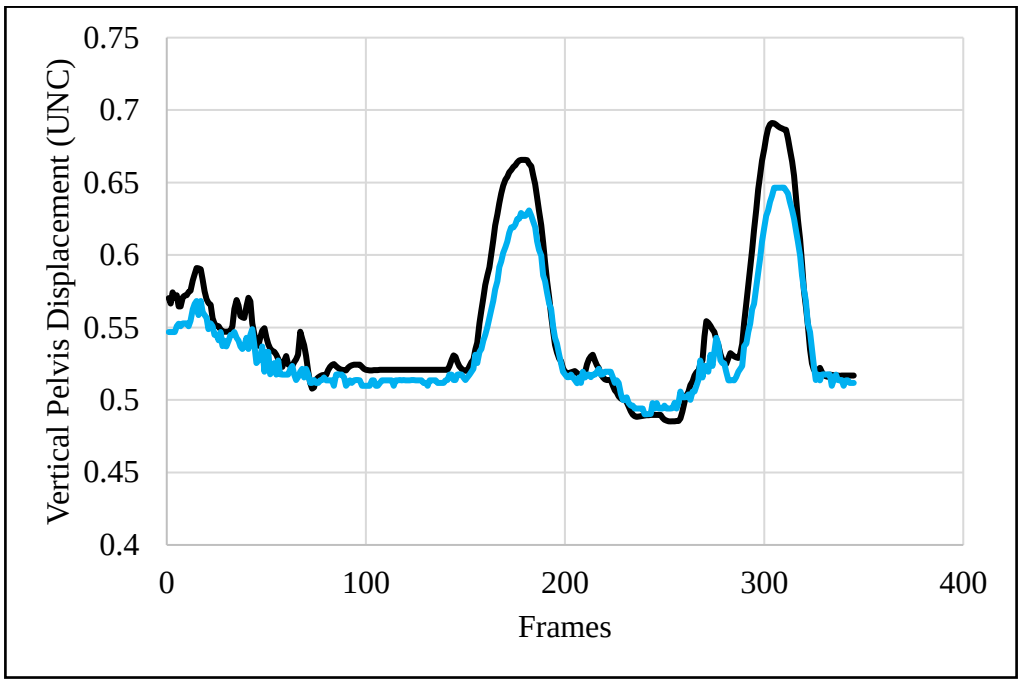


Figure 22: Vertical displacement of the pelvis during a lift and a lower from floor to waist shelf height using a face-blurred (blue) video and non-blurred face (black) video of the same recording as inputs into and processed through *wrnch*. UNC = unit screen spaced coordinates.

6.1.2 COG to load difference measure

It was hypothesized that 2D video and pose estimation based peak COG to the load would agree with the measure as calculated using 3D motion capture data. Similar to the angular measures, it is important to define what constitutes agreement. However, since this measure is not presented in the same units as the angular measures (% height versus degrees) we cannot use the same agreement criteria. Therefore, pointing to Figure 18 we notice a grey rectangular box spanning both sides of the mean difference. Giavarina, (2015) has previously

stated that we can add a measure to the mean difference to look at systematic error or potentially error due to sampling. Referring back to the statistical analysis of this study (Section 4.7), it was mentioned that a 95% confidence interval around the mean difference can interpret systematic error. If this confidence interval (represented by gray shaded area on the plots) does not cross or overlap with a difference of 0 (y-axis), we can attribute systematic error between the two approaches (Giavarina, 2015). Therefore, using this criterion, we reject Hypothesis 1A.

With a mean difference of 15 % height and subsequently a positive LOA, there is a systematic error that demonstrates a systematic bias where the 2D pose-estimation method underestimates this measure compared to the 3D motion capture method. We can hypothesize that this systematic bias can be due to the incorrect estimation of the location of the wrist marker in the sagittal plane, the difference in estimated height between the two approaches, or the calculated COG. The COG from data obtained using the 3D approach has a more posterior COG than from the COG calculated from *wrnch* outputs. Thus, we can interpret the smaller COG to load measure is in part due to this difference.

Further, the trend in the data set suggests that as the load of the box moves further away from the body on average, the difference between the two approaches also increases. Further, the LOA bandwidth is < 5 % height, this suggests that although there is systematic error presented, there is not a lot of variance in the differences. This lack of variability between the two measures for this dataset could be due to the fact that participants at the time

of the study were instructed to hold the load close to the body, thus limiting the amount of variability present in the data set.

6.1.3 Balance outcome measures

The FSL in the AP and ML directions from the video-based motion capture and pose-estimation analysis demonstrated values that were vastly different from the values presented from the 3D motion capture analysis. Again, we cannot directly compare the values to the other measures, however, we can infer from the Bland-Altman plots that the FSL measures did not demonstrate agreement between the two approaches, which supports Hypothesis 1B when considering both the FSL AP and ML.

Upon further investigation into the lack of agreement between the two methods regarding the FSL, it was found that the left and right heel markers overlapped in respect to the ML direction (i.e. the heel markers did not demonstrate separation in the ML direction, see Figure 23B). Additionally, based on Figure 9, it would be expected that the left foot markers are closer to 1 and the right foot markers are closer to 0 on the Z-axis. However, the outputs in *wrnch* demonstrated the opposite (Figure 23B, D, F, H). Therefore, we can infer that when in the sagittal plane specifically, the Z-axis, pointing orthogonal to the X and Y axes and “out of the screen”, elicits discrepancies in the data. Looking at Figure 23 we can see a lack of ability for *wrnch* to appropriately track the right foot, specifically the right toe marker (denoted by the blue circles in Figure 22). This could be due to the inability of the right toe to be tracked as it is further away from the camera, and because of potential

occlusion due to the box. Therefore, this discrepancy in tracking the right foot markers can lead to the lack of agreement specific to the balance measures. Furthermore, the COG in the AP direction from the data obtained using the 3D motion capture was found to be posterior to the CBOS, likely due to the addition of the loads applied at the hands. However, the COG calculated from the 2D markerless motion capture and pose-estimation data demonstrated a COG anterior to the CBOS. These differences could be due to the assumptions made when deriving segments of from the *wrnch* data for calculating whole-body COG.

Perhaps another source of systematic error between the two methods was the placement of the video camera (McKinnon et al., 2020), namely "parallax" error. The video camera was placed in the sagittal plane of the participant and at a height that ensured capture of the entire body of the participant completing the task. Perhaps the practicality of recording the whole-body motion limits the accuracy of the data points specific to the feet markers which may be seen at an alternate angle and thus field of view, as opposed to the trunk or pelvis markers. This theory rises from the fact that the FSL, which is highly dependent on a strong definition of BOS and subsequently the CBOS, were the least agreed upon measure between the two methods.

Pose-estimation techniques, sometimes present different approaches for their skeletal tracking algorithm (Section 2.5 above). Further speculation for the lack of agreement between the 2D and 3D methods can possibly be attributed to the lack of facial features present in the videos because of the sagittal view of the video camera. Less precision in skeletal motion

could occur if the algorithm relies on facial recognition as a starting point to find the person in space (See Section 866.1.1). If this is the case, perhaps this most largely effects the interpretation of the feet markers as the pose-estimation algorithm might work its way down the chain (from facial features to shoulder etc., to the heel and toe markers).

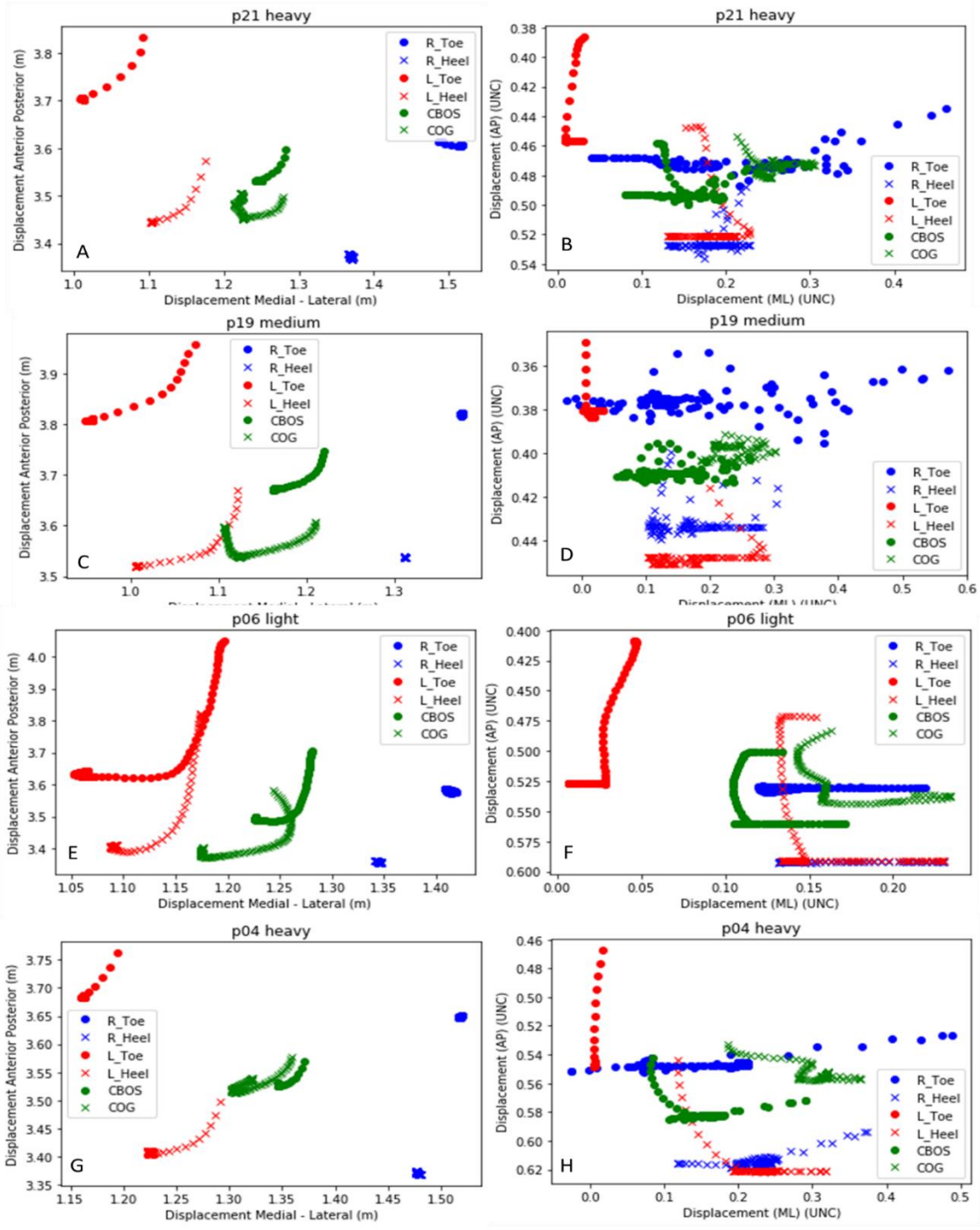

p21 heavy
Displacement Anterior Posterior (m)
Displacement Medial - Lateral (m)
R_Toe
R_Heel
L_Toe
L_Heel
CBOS
COG
A
p21 heavy
Displacement (AP) (UNC)
Displacement (ML) (UNC)
B
p19 medium
C
p19 medium
D
p06 light
E
p06 light
F
p04 heavy
G
p04 heavy
H

Figure 23: Examples of differences comparing the BOS, CBOS, and COG from the data obtained using the 3D motion capture (left side) and data obtained using 2D pose-estimation (right side). Specifically, the figures on the right side demonstrate larger variability in the toe and heel definitions.

6.2 Comparison to previous literature

This study used a novel markerless motion capture and pose estimation approach to determine outcome measures relevant to an FCE assessment from whole body trajectory data in 3D. To the best of the authors knowledge, there is only 1 study that has previously used *wrnch* as a 2D markerless motion capture system in comparison to 3D motion capture. McKinnon et al., (2020), calculated trunk flexion, shoulder elevation angle, and elbow flexion angle as well as reach envelopes using data obtained from 2D video recording and pose-estimation and 3D motion capture (Raptor-12HS, Motion Analysis, Rohnert Park, CA, USA). While their study did not use Bland-Altman plots to explore agreement, they concluded reach envelopes and angular displacements calculated from the video-based approach, showed low root mean squared errors compared to the motion capture (McKinnon et al., 2020). In the case of greater errors, camera angle (off-plane) was attributed to the errors (McKinnon et al., 2020).

In terms of automated tracking of human movement using 2D video-based solutions, the Kinect™ has been popularly used. Previous studies have demonstrated peak knee flexion using Kinect™ to have mean differences $\pm$ standard deviations of -19 $\pm$ 10 degrees (Pfister et al., 2014), as well as peak trunk displacement with LOA ranging from -8.7 to 0.20 degrees

(Clark et al., 2012). The Kinect™ was proven unacceptable for clinical measurement analysis during gait (Mentiplay et al., 2015; A. Pfister et al., 2014). In contrast, Clark et al., (2012), did demonstrate agreement. Perhaps, the depth sensor of Kinect™ plus the extensive calibration presented by Clark et al., (2012) lead to more desirable results. Previous studies mention that a potential detrimental factor to the differences when using Kinect, is the inability to assess internal / external joint rotations of the limbs as it does not possess the ability to accurately determine an orthogonal axis to the primary longitudinal axis (Pfister et al., 2014; Clark et al., 2012). This factor is similarly presented in this study and the outputs provided by *wrnch*. However, Mentiplay et al., (2015), computed 3D joint angles using Euler angles as opposed to a 2D cosine method, and did not find agreement when measuring knee, hip, and ankle angles during gait. This suggests that perhaps along with the inability to assess internal / external joint rotations of the limbs using 2D-video based methodologies, the skeletal joint tracking does not track joint or landmarks locations with high enough accuracy.

Another 2D video automated tracking method presented is “The Captury”, which demonstrated a variety of mean difference and LOA, with the most extreme case for knee varus (LOA = -33 to 19 degrees), when analyzing the concurrent validity of a marker-based (Vicon, UK) and markerless-based (The Captury) approaches for jump characteristics (Harsted et al., 2018). Similar to this study, the 2D markerless-based system underestimated joint angles relative to jumping in comparison the marker-based approach (Harsted et al., 2019).

Furthermore, looking into 2D-video based methodologies for calculating kinematics using manual tracking or labelling of still images, concurrent validity remains controversial. Using Kinovea software, a manual labelling and tracking software, there has been mean differences of trunk, hip, knee and ankle flexion of 1.68 degrees, 2.60 degrees, 0.74 degrees, and 3.1 degrees, respectively (Schurr et al., 2017). However, the LOA ranged, in the most extreme case, from -54.45 degrees to 57.81 degrees of trunk flexion (Schurr et al., 2017). Moreover, when using Dartfish Motion Analysis Software (Dartfish, Fribourg, Switzerland), studies have demonstrated task variability when it comes to determining the utility of 2D video analysis. Using Dartfish technology to measure frontal plane kinematics have shown poor (ICC = 0.39, r = 0.194) ((Maykut et al., 2015; Ortiz et al., 2016) to moderate (r = 0.60) (Maykut et al., 2015) to strong (ICC = 0.96) (Ortiz et al., 2016) correlations between the 2D methodology and 3D motion capture. This variability in results could be due to the tasks analyzed, running versus drop jumps, as well as the measures calculated, e.g. frontal plane projection angle versus knee and hip abduction. Nevertheless, these results within the one study suggest Dartfish technology for assessing 2D kinematics in comparison to 3D motion capture remains questionable.

Overall, single leg tasks and gait were the most popular to study and assess the concurrent validity in previous research. Furthermore, the methodology and video-based approaches, except for the Kinect™ sensor, often required manual labelling and determination of joint angles in a single plane at pre-determined time frames. This highlights

the novelty of using *wrnch* pose-estimation for this analysis to automatically track joint centers in 3D. Despite these methodological differences, the studies presented controversial results in whether the specific 2D system is clinically acceptable. Values when completing Bland-Altman plots of previous research demonstrated mean differences ranging from 0 to -20 degrees with LOA also exhibiting variability within the results. Similar to this study, errors were mostly attributed to the camera placement and view as well as the inability for joint or landmark trajectories to be represented in a local coordinate system or in a plane/ axis orthogonal to the plane of view. This presents the need to further understand sources of error and potentially determine a way to overcome these errors before utilizing 2D analysis in the field or clinical setting for whole-body motion tracking analysis.

6.3 Methodological Limitations

It is important to acknowledge limitations of this work that may impact the interpretation and generalizability of the results. First, a sample size of 20 participants required each lift be treated individually without averaging participant measures. However, this could reduce the natural variability of measures that may be presented if there were 60 participants. Also, approximately 50% of the participants had some manual materials handling experience and many participants were recruited from the department of applied health sciences. This sample could influence the variability presented as it is hypothesized most participants have a prior knowledge of lifting. However, since individual lifts were

compared for agreement as opposed to sample means, this sample population was acceptable for the study.

Next, biomechanical analyses were estimated for the floor-to-waist height lifting task only. Analysis of a single task limits the generalizability of the 2D markerless motion capture method (2D video and pose estimation) to be applied to various other tasks. Some examples that would also be considered in an FCE are, pulling, pushing, and carrying. However, considering the lack of agreement between the markerless motion capture and pose-estimation and 3D motion capture-based methods, perhaps more work on the joint centre identification and tracking algorithm needs to be done prior to analysis of other tasks.

Furthermore, limiting the camera view to the sagittal plane, while practical for a primarily sagittal task, did not allow exploration of the 2D video plus pose-estimation performance in different planes. However, because the task chosen for this study was a floor-to-waist lifting task commonly used in FCE's and in manual materials handling work, a frontal view of the participant would have to be analyzed with the video camera posterior to the participant due to the limitations of the shelves in space.

Another limitation presented is in the inability of the pose-estimation software to define 3D segmental orientations. While *wrnch* does have a feature that presents segmental orientations and the ability to derive segmental and perhaps intersegmental rotation of the body segments (in quaternions), we decided to forgo that analysis for the time being. This is because first, an analysis of the 3D joint locations and landmarks exported from *wrnch*,

needed to be analyzed prior to further investigation into the deeper layers of the algorithm and software. The raw 3D joint center and landmark outputs that are used for this thesis are all presented on the long axis of the segments. Therefore, using the 3D joint center data presented by *wrnch*'s pose-estimation model, we cannot compute joint or segment angles in 3D space using a Cardan sequence and then subsequently Euler angle calculations. Specifically, angles were calculated differently between the two approaches as markerless motion data was a projection based on the dot product and 3D motion capture a Cardan rotation. Similarly, Schurr et al., (2017), attributed a partially supported concurrent validity of 2D and 3D analysis to the lack of 3rd planar motion not accounted for in a 2D assessment.

Lastly, one limitation is the "black-box" when using *wrnch* for analysis. While we can make educated assumptions on how the algorithm works, we do not know for certain the errors that are presented. Therefore, if we want to attempt to improve upon these data outputs, we then need to make assumptions and improve upon the data set.

6.4 Future Directions and potential clinical applications

This study presented with an overarching objective to increase the quantification of subjectively appraised FCE measures. We proposed a methodology exploring the utility of markerless motion capture and pose-estimation to quantify these measures. However, in comparison to a lab-based 3D motion capture system, the outcomes derived from the markerless-based motion capture did not demonstrate agreement. One way to attempt to improve upon the outcome measures from the 2D data set is to train an artificial neural

network model using the 3D motion capture data collected. It has been reported that convolutional neural networks have significantly boosted the accuracy on body pose-estimation (Cao et al., 2019). Therefore, use of a CNN as a tool to improve the accuracy of pose-estimation for FCEs is a desired future direction. Further, it will be important to validate the 2D pose-estimation with different tasks, both symmetrical and asymmetrical with different views of the camera (e.g. frontal or off-axis). Lastly, we will explore the use of the outputted kinematic data in the form of quaternions to better describe the segment orientations and rotations in space. If we are able to decompose the quaternions into constituent Cardan angles, such that joint angles are defined more consistently with the definitions used when calculated using the lab-based 3D motion capture system, then perhaps it is less important where the camera is located in space. This would be beneficial in a clinical setting where there may be obstructions in the space and so the video camera can only be placed in specific locations.

Bringing it full circle, this thesis explored the validation in 2D video and pose-estimation relative to 3D motion capture analysis with outputs that were deemed to be relevant for FCE analysis. While we first needed to explore the concurrent validity of the 2D pose-estimation software, we further need to explore if the outcome measures chosen for this thesis are also reliable and valid indicators of performance.

7. Conclusion

This thesis explored the agreement between a 2D video marklerless motion capture and pose-estimation approach and 3D motion capture approach for gather data necessary to conduct a biomechanical analysis of FCE performance. Concurrent validity was determined between these two methods through an agreement analysis using Bland-Altman plots (Bland & Altman, 1986). The dependent variables calculated from the two data collection approaches were specific to biomechanical measures that are often used subjectively to evaluate worker capacity. Based on the results of the study, we conclude that overall, there was a lack of agreement between the two methods of analysis. It was found that values represented in the sagittal plane (e.g. peak angles) demonstrated an underestimation in the peak value when comparing the 2D pose-estimation method to the 3D motion capture method. However, the shoulder abduction angle (frontal plane) found to incorporate values where the 2D pose-estimation method both over- and underestimation compared to the 3D motion capture analysis. FSL's did not demonstrate values that were in agreement between the two methods likely due to the ability of the pose-estimation to track the toe and heel markers and differentiate the right and left feet. Lastly, the MARP measures representing inter-joint coordination between the hip and knee did not demonstrate agreement nor did they demonstrate over or underestimation systematic error between the two methods.

The novelty of this study was presented in both the methodology by calculating measures that were not only simplistic kinematic outcomes that can be related in a variety of contexts, (e.g. joint angles), but also measures that are relevant to an FCE setting (e.g. FSL). This thesis

further discusses important considerations and limitations to the use of pose-estimation for the purpose of calculating biomechanical metrics of interest pertaining to FCE performance. Overall, the lack of agreement between the two methods infers the need to further investigate the causes of the systematic error to improve the agreement.

Bibliography

(ELC), E. lift capacity test. (2016). Epic Rehab.

Alahmari, A., Herrington, L., & Jones, R. (2020). Concurrent validity of two-dimensional video analysis of lower-extremity frontal plane of movement during multidirectional single-leg landing. *Physical Therapy in Sport*, *42*, 40–45. https://doi.org/10.1016/j.ptsp.2019.12.009

Albert, W. J., Wrigley, A. T., & McLean, R. B. (2008). Are males and females similarly consistent in their respective lifting patterns? *Theoretical Issues in Ergonomics Science*, *9*(4), 347–358. https://doi.org/10.1080/14639220701507398

Allison, S., Galper, J., Hoyle, D., & Mecham, J. (2018). *Current concepts in functional capacity evaluation: A Best Practices Guideline* (pp. 1–43). pp. 1–43.

Armstrong, D. P., Budarick, A. R., Pegg, C. E. E., Graham, R. B., & Fischer, S. L. (2020). Feature Detection and Biomechanical Analysis to Objectively Identify High Exposure Movement Strategies When Performing the EPIC Lift Capacity test. *Journal of Occupational Rehabilitation*. https://doi.org/10.1007/s10926-020-09890-2

Armstrong, D. P., Pegg, C. E. E., & Fischer, S. L. (2019). Is deep squat movement strategy related to floor-to-waist height lifting strategy: implications for physical employment testing. *Ergonomics*. https://doi.org/10.1080/00140139.2019.1689303

Armstrong, D. P., Ross, G. B., Graham, R. B., & Fischer, S. L. (2019). *Considering movement competency within physical employment standards*. *63*, 603–613. https://doi.org/10.3233/WOR-192955

Bell, A. L., Brand, R. A., & Pedersen, D. R. (1989). Prediction of hip joint centre location from

external landmarks. *Human Movement Science*, *8*(1), 3–16. https://doi.org/10.1016/0167-9457(89)90020-1

Bell, A. L., Pedersen, D. R., & Brand, R. A. (1990). A comparison of the accuracy of several hip center location prediction methods. *Journal of Biomechanics*, *23*(6), 617–621. https://doi.org/10.1016/0021-9290(90)90054-7

Bigos, S. J., Battie, M. C., Fisher, L. D., Hansson, T. H., Spengler, D. M., & Nachemson, A. L. (1992). Prospective evaluation of preemployment screening methods. *Spine*, *17*(8), 922–926.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310.

Bland, J. M., & Altman, D. G. (1995). Comparing two methods of clinical measurement: A personal history. *International Journal of Epidemiology*, *24*. https://doi.org/10.1093/ije/24.supplement_1.s7

Brouwer, S., Reneman, M, F., Dijkstra, P, U., Groothoff, J. W., Schellekens, J. M. ., & Goeken, L. N. . (2003). Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Patients with Chronic Low Bck Pain. *Journal of Occupational Rehabilitation*, *13*(4), 207–218. https://doi.org/10.1023/A

Burgess-Limerick, R., Abernethy, B., Neal, R. J., & Kippers, V. (1995). Self-selected manual lifting technique: Functional consequences of the interjoint coordination. *Human Factors*, *37*(2), 395–411. https://doi.org/10.1518/001872095779064537

Burgess-Limerick, Robin, Abernethy, B., & Neal, R. J. (1993). Relative phase quantifies interjoint coordination. *Journal of Biomechanics*, *26*(1), 91–94. https://doi.org/10.1016/0021-

9290(93)90617-N

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S.-E., & Sheikh, Y. A. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *XXX*(XXX), 1–1. https://doi.org/10.1109/tpami.2019.2929257

Carlton, M. A., & Devore, J. . (2017). Appendix A: Statistical Tables. In *Probability with Applications in Engineering, Science, and Technology* (pp. 597–607). https://doi.org/10.1002/9780470496862.app1

Chaffin, D B, & Page, G. B. (1994). Postural effects on biomechanical and psychophysical weight-lifting limits. *ERGONOMICS*, *37*(4), 663–676. https://doi.org/10.1080/00140139408963681

Chaffin, Don B. (1974). Human Strength Capability and Low-Back Pain. *Journal of Occupational Medicine*, *16*(4), 248–254.

Chaffin, Don B, Herrin, G. D., & Keyserling, W. M. (1978). Preemployment Strength Testing: An Updated Position. *Journal of Occupational Medicine*, *20*(6), 403–408.

Chaffin, Don B, & Park, K. S. (1973). A Longitudinal Study of Low-Back Pain as Associated with Occupational Weight Lifting Factors. *American Industrial Hygiene Association Journal*, *34*(12), 513–525.

Clark, R. A., Pua, Y. H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. *Gait and Posture*, *36*(3), 372–377. https://doi.org/10.1016/j.gaitpost.2012.03.033

Commissaris, D. A. C. M., & Toussaint, H. M. (1997). Load knowledge affects low-back loading and control of balance in lifting tasks. *Ergonomics*, *40*(5), 559–575. https://doi.org/10.1080/001401397188035

Dempsey, P. G. (1998). A critical review of biomechanical, epidemiological, physiological and psychophysical criteria for designing manual materials handling tasks. *Ergonomics*, *41*(1), 73–88. https://doi.org/10.1080/001401398187332

Dempster, W. T. (1955). Space requirements of the seated operator, geometrical, kinematic, and mechanical aspects of the body with special reference to the limbs. Michigan State Univ East Lansing.

Drillis, R., & Contini, R. (1966). Body Segmetn Parameters. New York, New York: Office of Vocational Rehabilitation, (Report No.: No. 1166-03.).

EPIC Lift Capacity Test (ELC). (n.d.). Retrieved January 14, 2020, from https://www.epicrehab.com/products/products-background/epic-lift-capacity-test-elc/

Galgon, A. K., & Shewokis, P. A. (2016). Using Mean Absolute Relative Phase , Deviation Phase and Point-Estimation Relative Phase to Measure Postural Coordination in a Serial Reaching Task. *Journal of Sports Science and Medicine*, *15*, 131–141.

Gardener, L., & McKenna, K. (1999). Reliability of occupational therapists in determining safe, maximal lifting capacity. *Australian Occupational Therapy Journal*, *46*(3), 110–119. https://doi.org/10.1046/j.1440-1630.1999.00184.x

Genovese, E., & Isernhagen, S. J. (2009). Approach to Requesting a Functional Capacity Evaluation. In E. Genovese & J. S. Galper (Eds.), *Guide to the Evaluation of Functional Ability: How to*

Request, Interpret, and Apply Functional Capacity Evaluations (pp. 19–40). American Medical Association.

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Gouttebarge, V, Wind, H., Kuijer, P. P. F. M., & Frings-Dresen, M. H. W. (2004). Reliability and validity of Functional Capacity Evaluation methods: A systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. *International Archives of Occupational and Environmental Health*, *77*(8), 527–537. https://doi.org/10.1007/s00420-004-0549-7

Gouttebarge, Vi, Wind, H., Kuijer, P., Sluiter, J., & Frings-Dresen, M. (2009). Construct Validity of Functional Capacity Evaluation Lifting Tests in Construction Workers on Sick Leave as a Result of Musculoskeletal Disorders. *Archives of Physical Medicine and Rehabilitation*, *90*, 302–308. https://doi.org/10.1016/j.apmr.2008.07.020

Grood, E. S., & Suntay, W. J. (1983). A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee. *Journal of Biomechanical Engineering*, *105*(2), 136–144. https://doi.org/10.1115/1.3138397

Gross, D. P., & Battie, M. C. (2005). Factors Influencing Results of Functional Capacity Evaluations in Workers' Compensation Claimants With Low Back Pain. *Physical Therapy*, *85*(4), 315–322. Retrieved from https://academic.oup.com/ptj/article-abstract/85/4/315/2805001

Hanley, B., Tucker, C. B., & Bissas, A. (2018). Differences between motion capture and video analysis systems in calculating knee angles in elite-standard race walking. *Journal of Sports*

Sciences, *36*(11), 1250–1255. https://doi.org/10.1080/02640414.2017.1372928

Harbin, G., & Olson, J. (2005). Post-offer, pre-placement testing in industry. *American Journal of Industrial Medicine*, *47*(4), 296–307. https://doi.org/10.1002/ajim.20150

Harsted, S., Holsgaard-Larsen, A., Hestbæk, L., Boyle, E., & Lauridsen, H. H. (2019). Concurrent validity of lower extremity kinematics and jump characteristics captured in pre-school children by a markerless 3D motion capture system. *Chiropractic and Manual Therapies*, *27*(1), 1–16. https://doi.org/10.1186/s12998-019-0261-z

Hart, D. I., Stanley Berlin, P., Brager, P. E., Caruso, M., Hejduk, loseph F., Howar, lulie M., … Wah, M. D. (1994). Development of Clinical Standards in Industrial Rehabilitation. *Journal of Orthopaedic & Sports Physical Therapy*, *19*(5), 232–241. Retrieved from www.jospt.org

Hart, D. L., Isernhagen, S. J., & Matheson, L. N. (1993). Guidelines for Functional Capacity Evaluation of People With Medical conditions' for the Use of Functional Measure-ments: Reference Manual for Functional Capacity Evaluations, Virginia: The Task Force on Objective Functional Measurements. *Journal of Orthopaedic & Sports Physical Therapy*, *18*(6), 682–686. Retrieved from www.jospt.org

Holbein-Jenny, M. A., McDermott, K., Shaw, C., & Demchak, J. (2007). Validity of functional stability limits as a measure of balance in adults aged 23-73 years. *Ergonomics*, *50*(5), 631–646. https://doi.org/10.1080/00140130601154814

Holbein, M. A., & Redfern, M. S. (1997). Functional stability limits while holding loads in various positions. *International Journal of Industrial Ergonomics*, *19*(5), 387–395. https://doi.org/10.1016/S0169-8141(96)00023-6

Howarth, S. J., & Callaghan, J. P. (2010). Quantitative assessment of the accuracy for three interpolation techniques in kinematic analysis of human movement Quantitative assessment of the accuracy for three interpolation techniques in kinematic analysis of human movement. *Computer Methods in Biomechanics and Biomedical Engineering*, *13*(6), 847–855. https://doi.org/10.1080/10255841003664701

Ijmker, S., Gerrits, E. H. J., & Reneman, M. F. (2003). Upper Lifting Performance of Healthy Young Adults in Functional Capacity Evaluations: A Comparison of Two Protocols. In *Journal of Occupational Rehabilitation* (Vol. 13).

Innes, E. (2006). Reliability and Validity of Functional Capacity Evaluations: An Update. *International Journal of Disability Management*, *1*(1), 135–148. https://doi.org/10.1375/jdmr.1.1.135

Innes, E., & Straker, L. (1999a). Reliability of work-related assessments. *Work*, *13*(2), 107–124.

Innes, E., & Straker, L. (1999b). Validity of work-related assessments. *Work*, *13*, 125–152.

Isernhagen, S. (2009). Introduction to Functional Capacity Evaluation. In E. Genovese & J. . Galper (Eds.), *Guide to the Evaluation of Functional Ability: How to Request, Interpret, and Apply Functional Capacity Evaluations* (pp. 1–18). American Medical Association.

Keogh, J. W. L., Cox, A., Anderson, S., Liew, B., Olsen, A., Schram, B., & Furness, J. (2019). Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: A systematic review. *PLoS ONE*, *14*(5), 1–24. https://doi.org/10.1371/journal.pone.0215806

King, P. M., Tuckwell, N., & Bawett, T. E. (1998). A Critical Review of Functional Capacity

Evaluations. *Physical Therapy*, *78*(8), 852–866. Retrieved from https://academic.oup.com/ptj/article-abstract/78/8/852/2633317

Kingma, I., De Looze, M. P., Van Dieën, J. H., Toussaint, H. M., Adams, M. A., & Baten, C. T. M. (1998). When is a lifting movement too asymmetric to identify lowback loading by 2-D analysis? *Ergonomics*, *41*(10), 1453–1461. https://doi.org/10.1080/001401398186207

Lamb, P. F., & Stöckl, M. (2014). On the use of continuous relative phase: Review of current approaches and outline for a new standard. *Clinical Biomechanics*, *29*(5), 484–493. https://doi.org/10.1016/j.clinbiomech.2014.03.008

Liao, J. J. Z. (2010). Sample size calculation for an agreement study. *Pharmaceutical Statistics*, *132*, 125–132. https://doi.org/10.1002/pst

Lindbeck, L., & Kjellberg, K. (2001). Gender differences in lifting technique. *Ergonomics*, *44*(2), 202–214. https://doi.org/10.1080/00140130120142

Luvizon, D. C., Picard, D., & Tabia, H. (2018). 2D / 3D Pose Estimation and Action Recognition using Multitask Deep Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5137–5146.

Matheson, L. N., Isernhagen, S. J., & Hart, D. L. (2002). Relationships Among Lifting Ability, Grip Force, and Return to Work. *Physical Therapy*, *82*(3), 249–256. Retrieved from https://academic.oup.com/ptj/article-abstract/82/3/249/2837009

Matheson, L. N., Verna, J., Dreisinger, T. E., Leggett, S., & Mayer, J. (2014). Age and gender normative data for lift capacity. *Work*, *49*, 257–269. https://doi.org/10.3233/WOR-131671

Maykut, J. N., Taylor-Haas, J. A., Paterno, M. V, DiCesare, C. A., & Ford, K. R. (2015). Concurrent validity and reliability of 2d kinematic analysis of frontal plane motion during running. *International Journal of Sports Physical Therapy*, *10*(2), 136–146. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25883862%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4387721

McAlinden, C., Khadka, J., & Pesudovs, K. (2011). Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic and Physiological Optics*, *31*(4), 330–338. https://doi.org/10.1111/j.1475-1313.2011.00851.x

McKinnon, C. D., Sonne, M. W., & Keir, P. J. (2020). Assessment of Joint Angle and Reach Envelope Demands Using a Video- - Based Physical Demands Description Tool. *Human Factors*, 1–11. https://doi.org/10.1177/0018720820951349

Mentiplay, B. F., Perraton, L. G., Bower, K. J., Pua, Y. H., McGaw, R., Heywood, S., & Clark, R. A. (2015). Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables. *Journal of Biomechanics*, *48*(10), 2166–2170. https://doi.org/10.1016/j.jbiomech.2015.05.021

Mjøsund, H. L., Boyle, E., Kjaer, P., Mieritz, R. M., Skallgård, T., & Kent, P. (2017). Clinically acceptable agreement between the ViMove wireless motion sensor system and the Vicon motion capture system when measuring lumbar region inclination motion in the sagittal and coronal planes. *BMC Musculoskeletal Disorders*, *18*(1), 1–9. https://doi.org/10.1186/s12891-017-1489-1

Munro, A., Herrington, L., & Carolan, M. (2012). Reliability of 2-dimensional video assessment of

frontal-plane dynamic knee valgus during common athletic screening tasks. *Journal of Sport Rehabilitation*, *21*(1), 7–11. https://doi.org/10.1123/jsr.21.1.7

Myles, P. S., & Cui, J. (2007). I. Using the Bland-Altman method to measure agreement with repeated measures. *British Journal of Anaesthesia*, *99*(3), 309–311. https://doi.org/10.1093/bja/aem214

Nambiar, A. M., Correia, P., & Soares, L. D. (2012). Frontal gait recognition combining 2D and 3D data. *MM and Sec'12 - Proceedings of the 14th ACM Multimedia and Security Workshop*, 145–150. https://doi.org/10.1145/2361407.2361432

Nitschke, J., Nattrass, C., Disler, P., Chou, M., & Ooi, K. (1999). Reliability of the American Medical Association Guides' Model for Measuring Spinal Range of Motion. *Spine*, *24*(3), 262–268.

Norris, B. S., & Olson, S. L. (2011). Concurrent validity and reliability of two-dimensional video analysis of hip and knee joint motion during mechanical lifting. *Physiotherapy Theory and Practice*, *27*(7), 521–530. https://doi.org/10.3109/09593985.2010.533745

Nussbaum, M. A., & Zhang, X. (2000). Heuristics for locating upper extremity joint centres from a reduced set of surface markers. *Human Movement Science*, *19*(5), 797–816. https://doi.org/10.1016/S0167-9457(00)00020-8

Ortiz, A., Rosario-Canales, M., Rodríguez, A., Seda, A., Figueroa, C., & Venegas-Ríos, H. (2016). Reliability and concurrent validity between two-dimensional and three-dimensional evaluations of knee valgus during drop jumps. *Open Access Journal of Sports Medicine*, 65. https://doi.org/10.2147/oajsm.s100242

Park, S., Hwang, J., & Kwak, N. (2016). 3D human pose estimation using convolutional neural

networks with 2D pose information. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9915 LNCS*, 156–169. https://doi.org/10.1007/978-3-319-49409-8_15

Pfister, A., West, A. M., Bronner, S., & Noah, J. A. (2014). Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of Medical Engineering and Technology*, *38*(5), 274–280. https://doi.org/10.3109/03091902.2014.909540

Pfister, T., Charles, J., & Zisserman, A. (2015). Flowing ConvNets for Human Pose Estimation in Videos. *Proceedings of the IEEE International Conference on Computer Vision*, 1913–1921. https://doi.org/10.5540/tema.2013.014.03.0319

Pranata, A., Perraton, L., Clark, R., Mentiplay, B., Fortin, K., Long, B., … Bryant, A. L. (2018). Trunk and lower limb coordination during lifting in people with and without chronic low back pain Trunk and lower limb coordination during lifting in people with and without chronic low back pain. *Journal of Biomechanics*. https://doi.org/10.1016/j.jbiomech.2018.02.016

Qi, T., Bayramli, B., Ali, U., Zhang, Q., & Lu, H. (2019). *Spatial Shortcut Network for Human Pose Estimation*. (arXiv preprint arXiv:1904.03141). Retrieved from http://arxiv.org/abs/1904.03141

Raptis, M., Kirovski, D., & Hoppe, H. (2011). Real-time classification of dance gestures from skeleton animation. *Proceedings - SCA 2011: ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, *1*, 147–156. https://doi.org/10.1145/2019406.2019426

Remedios, S. M., Armstrong, D. P., Graham, R. B., & Fischer, S. L. (2020). Exploring the Application of Pattern Recognition and Machine Learning for Identifying Movement Phenotypes During Deep Squat and Hurdle Step Movements. *Frontiers in Bioengineering and*

Biotechnology, *8*(April), 1–15. https://doi.org/10.3389/fbioe.2020.00364

Reneman, M. F., Brouwer, S., Dijkstra, P. U., & Geertzen, J. H. (2004). Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. *Journal of Occupational Rehabilitation*. https://doi.org/10.1023/B

Reneman, M. F., Fokkens, A. S., Dijkstra, P. U., Geertzen, J. H. B., & Groothoff, J. W. (2005). Testing lifting capacity: validity of determining effort level by means of observation. *Spine*, *30*(2), 40–46. https://doi.org/10.1097/01.brs.0000150500.80699.89

Robert, J. J., Blide, R. W., Whorter, K. M., & Coursey, C. (1995). The effects of a work hardening program on cardiovascular fitness and muscular strength. *Spine*, Vol. 20, pp. 1187–1193. https://doi.org/10.1097/00007632-199505150-00014

Robertson, D. G. E., Caldwell, G. E., & Hamill, J. (2014). *Research Methods in Biomechanics, Second Edition.*

Schurr, S. A., Marshall, A. N., Resch, J. E., & Saliba, S. A. (2017). Two-Dimensional Video Analysis Is Comparable To 3D Motion Capture in Lower Extremity Movement Assessment. *International Journal of Sports Physical Therapy*, *12*(2), 163–172. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28515970%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5380858

Science and technology. (n.d.). Retrieved from https://wrnch.ai/technology/

Seay, J. F., Sauer, S. G., Patel, T., & Roy, T. C. (2016). A history of low back pain affects pelvis and trunk coordination during a sustained manual materials handling task. *Journal of Sport and Health Science*, *5*(1), 52–60. https://doi.org/10.1016/j.jshs.2016.01.011

Sinden, K. E., & MacDermid, J. C. (2016). Evaluating the Reliability of a Marker-Less, Digital Video Analysis Approach to Characterize Fire-fighter Trunk and Knee Postures During a Lift Task: A Proof-of-Concept Study. *Journal of Ergonomics*, *06*(01). https://doi.org/10.4172/2165-7556.1000145

Sinden, K. E., McGillivary, T. L., Chapman, E., & Fischer, S. L. (2017). Survey of kinesiologists' functional capacity evaluation practice in Canada. *Work*, *56*(4), 571–580. https://doi.org/10.3233/WOR-172519

Smith, R. L. (1994). Therapists' Ability to Identify Safe Maximum Lifting in Low Back Pain Patients During Functional Capacity Evaluation. *Journal of Orthopaedic & Sports Physical Therapy*, *19*(5), 277–281. Retrieved from www.jospt.org

Soer, R., van der Schans, C. P., Groothoff, J. W., Geertzen, J. H. B., & Reneman, M. F. (2008). Towards Consensus in Operational Definitions in Functional Capacity Evaluation: a Delphi Survey. *Journal of Occupational Rehabilitation*, *18*(4), 389–400. https://doi.org/10.1007/s10926-008-9155-y

Sorenson, B., Kernozek, T. W., Willson, J. D., Ragan, R., & Hove, J. (2015). Two- and Three-Dimensional relationships between Knee and Hip kinematic motion analysis: Single-Leg Drop-Jump landings. *Journal of Sport Rehabilitation*, *24*(4), 363–372. https://doi.org/10.1123/jsr.2014-0206

Stone, E. E., & Skubic, M. (2011). Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6491–6494.

Tengland, P. A. (2011). The concept of work ability. *Journal of Occupational Rehabilitation*, *21*, 275–285. https://doi.org/10.1007/s10926-010-9269-x

Tome, D., Russell, C., & Agapito, L. (2017). Lifting from the deep: Convolutional 3D pose estimation from a single image. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 5689–5698. https://doi.org/10.1109/CVPR.2017.603

Trafimow, J., & Aruin, A. S. (2018). The Use of Negative Acceleration as Accessory Force during Lifting. *Advances in Orthopedics*, 1–4.

Trippolini, M. A., Dijkstra, P. U., Jansen, B., Oesch, P., Geertzen, J. H. B., & Reneman, M. F. (2014). Reliability of Clinician Rated Physical Effort Determination During Functional Capacity Evaluation in Patients with Chronic Musculoskeletal Pain. *Journal of Occupational Medicine and Toxicology*, *24*, 361–369. https://doi.org/10.1007/s10926-013-9470-9

Tuckwell, N. L., Straker, L., & Barrett, T. E. (2002). Test-retest reliability on nine tasks of the Physical Work Performance Evaluation. In *Work* (Vol. 19). Retrieved from IOS Press website: https://content.iospress.com/download/work/wor00260?id=work%2Fwor00260

Verlekar, T. T., De Vroey, H., Claeys, K., Hallez, H., Soares, L. D., & Correia, P. L. (2019). Estimation and validation of temporal gait features using a markerless 2D video system. *Computer Methods and Programs in Biomedicine*, *175*, 45–51. https://doi.org/10.1016/j.cmpb.2019.04.002

Wind, H., Gouttebarge, V., Kuijerl, P. P. F. M., Sluiter, J. K., & Frings-Dresen, M. H. W. (2006). The utility of Functional Capacity Evaluation: the opinion of physicians and other experts in the field of return to work and disability claims. *International Archives of Occupational and*

Environmental Health, *79*, 528–534. https://doi.org/10.1007/s00420-005-0081-4

Winter, D. A. (2009). *Biomechanics and motor control of human movement*. John Wiley & Sons.

Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., … Stokes, I. (2002). ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine. *Journal of Biomechanics*, *35*(4), 543–548. https://doi.org/10.1016/S0021-9290(01)00222-6

Wu, G., van der Helm, F. C. T., (DirkJan) Veeger, H. E. J., Makhsous, M., Van Roy, P., Anglin, C., … Buchholz, B. (2005). ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand. *Journal of Biomechanics*, *38*(5), 981–992. https://doi.org/10.1016/J.JBIOMECH.2004.05.042

Appendices

Appendix A: 3D calibration pose and segment orientation definitions

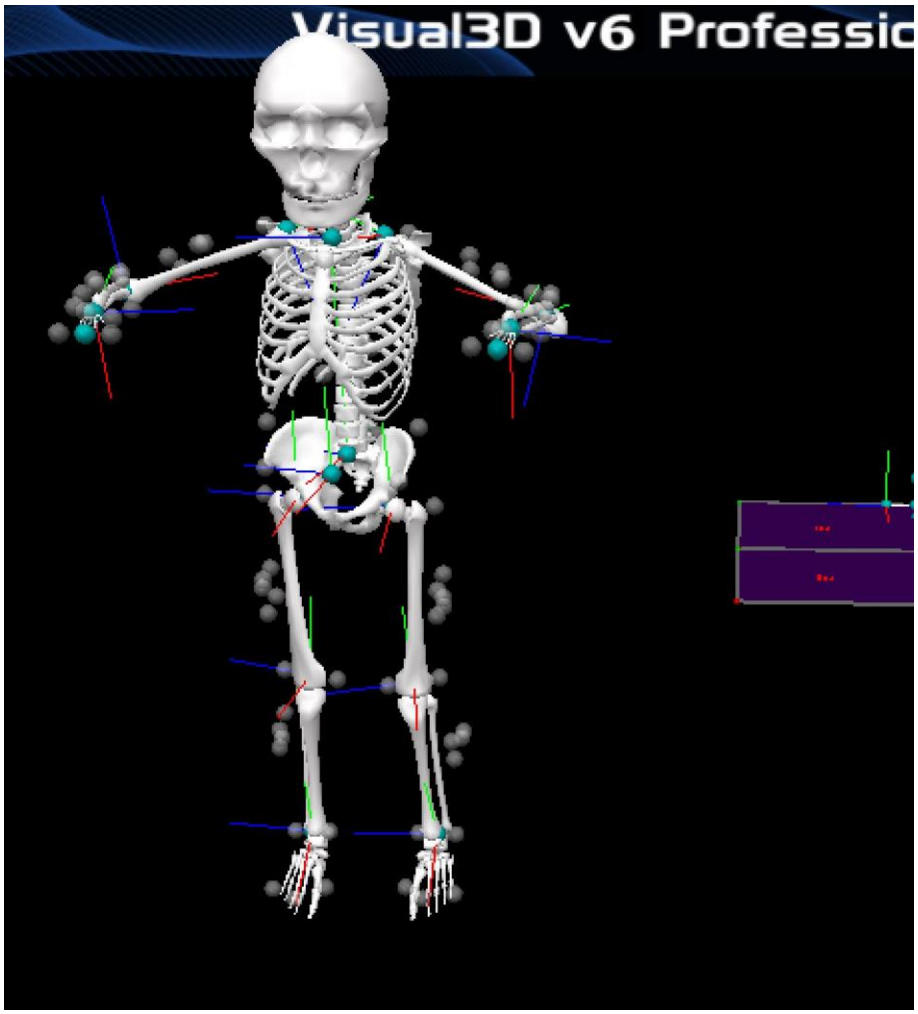


Figure A-1: Visual representation of the 3D model used in Visual 3D. Y-axis pointing superior-inferior, X-axis pointing anterior-posterior and Z-axis pointing medial-lateral

Appendix B: Mean Absolute Relative Phase Angle Visuals

Step 1: Calculate hip and knee angular displacement and hip and knee angular velocity

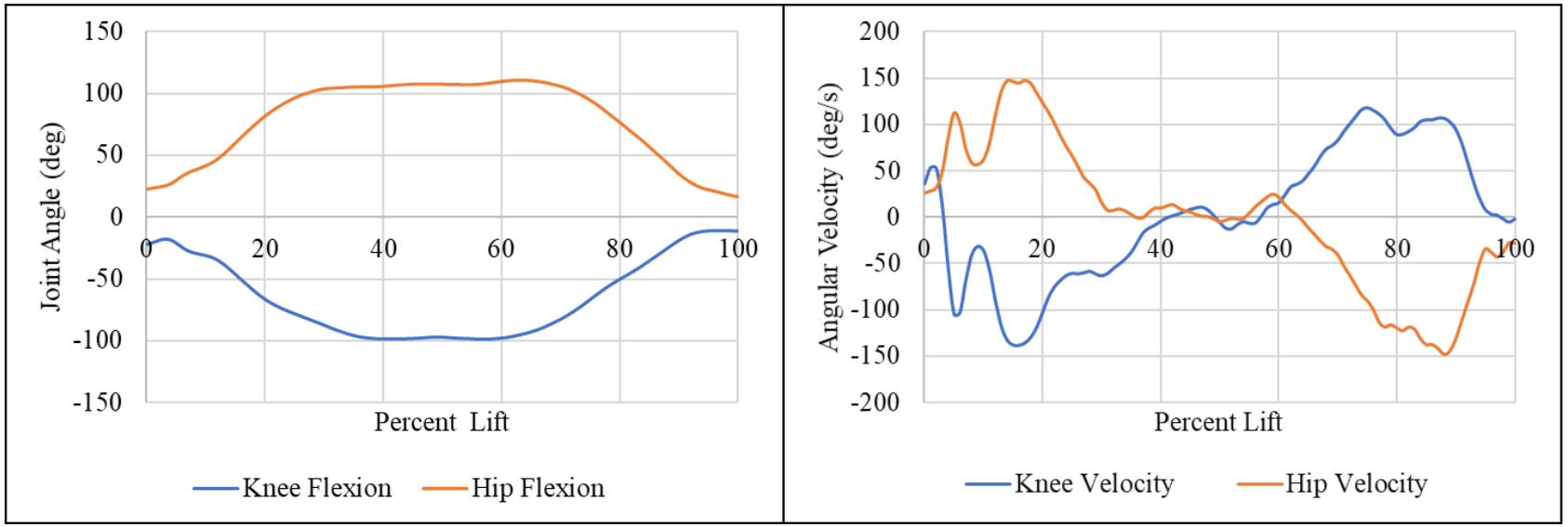


Figure B-1: Knee and hip angular displacement (left) and knee and hip angular velocity (right)

Step 2: Normalize angular displacement and angular velocity of the knee and hip to (-1 to +1) and

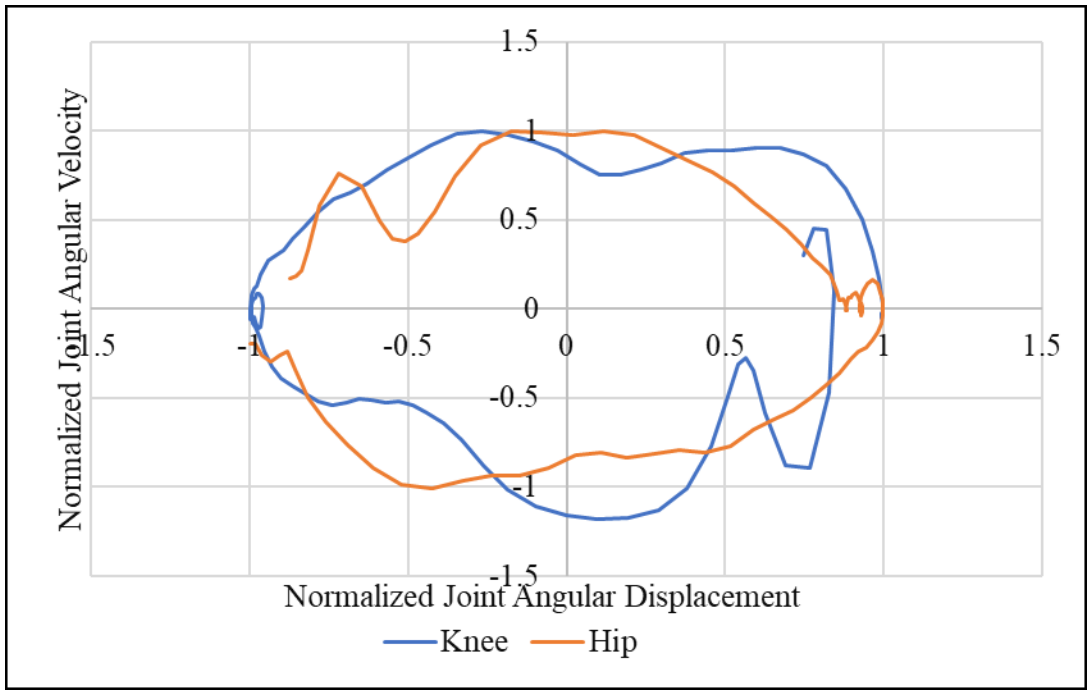


Figure B-2: Normalized hip and angle joint angular displacement versus joint angular velocity.

Step 3: Determine relative phase curves of the hip – knee normalized phase angular displacements and angular velocities