MEASUREMENT OF LOWER EXTREMITY FRONTAL-PLANE ALIGNMENT AND KNEE OSTEOARTHRITIS SEVERITY USING PHOTOGRAPHIC AND RADIOGRAPHIC APPROACHES

by

Lisa Mary Sheehy

A thesis submitted to the School of Rehabilitation Therapy
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
(September, 2013)

Copyright © Lisa Mary Sheehy, 2013
Abstract

Osteoarthritis (OA) of the knee affects between 5.4% and 38% of older adults and this prevalence is increasing as the population ages and becomes more obese. As health costs rise, it is important to have accurate and cost-effective methods to assess knee OA and the risk for OA.

One risk factor for progression of knee OA is lower extremity (LE) frontal-plane malalignment. The first goal of this thesis was to assess the suitability of knee radiographs and LE photographs for the estimation of frontal-plane LE alignment. In the first study, several versions of the femoral shaft-tibial shaft (FS-TS) angle, assessed from knee radiographs, were compared to the hip-knee-ankle (HKA) angle, assessed from full-length radiographs. We concluded that the FS-TS angle is not a recommended substitute for the HKA angle, because the association between the two measures differs depending on alignment, OA severity and the method of determining the FS-TS angle.

In the second study, the hip-knee-ankle angle determined from a pelvis-to-ankle photograph (HKA-P) was assessed for its ability to estimate the HKA angle. The HKA-P angle was reliable and highly correlated to the HKA. It therefore shows promise as an accurate and cost-effective assessment tool for the estimation of LE alignment.

Commonly-used grading scales for the severity of knee OA seen on a radiograph emphasize just one feature of OA; therefore the second goal of this thesis was to assess the psychometric properties of the unicompartmental osteoarthritis grade (UCOAG), a composite scale which grades several features of OA in the tibiofemoral (TF) compartment.

In the third and fourth studies, the reliability, validity and sensitivity to change of the UCOAG scale was assessed and compared to two commonly-used scales (Kellgren-Lawrence and Osteoarthritis Research Society International joint space narrowing). The UCOAG scale showed moderate to excellent reliability. All three scales demonstrated comparable validity and
sensitivity to change. The UCOAG is therefore recommended for the assessment of OA severity and change over time.

This research provides evidence for the use of accurate and cost-effective measures to assess LE alignment using photographs, and TF OA severity using radiographs, for clinical assessment and research purposes.
Co-Authorship

Lisa Sheehy was the primary author of all chapters included in this thesis. For each chapter I designed the study, with the assistance of my supervisory committee (Linda McLean, Elsie Culham and T. Derek V. Cooke). I carried out the study procedures, performed the data analysis and wrote the initial manuscripts for each study. Assistance for data analysis and manuscript editing was given by my supervisory committee.

Chapter 3 was published in the journal Osteoarthritis & Cartilage in 2011. This study was carried out using the public database created by the Multicenter Osteoarthritis Study (MOST). The MOST study is sponsored by the National Institutes of Health (NIH) and the National Institutes of Aging in the United States. The MOST senior investigators along with their NIH grant numbers are: David Felson (Boston University) AG18820, James Torner (University of Iowa) AG18832, Cora Lewis (University of Alabama at Birmingham) AG18947 and Michael Nevitt (University of California at San Francisco) AG19069. My MOST sponsor was David Felson. Co-authors were David Felson (MOST; study design), YuQing Zhang (MOST; participant selection), Yuk-Miu Lam (statistical analysis), Neil Segal (MOST; editing of manuscript), John Lynch (MOST; editing of manuscript) and T. Derek V. Cooke.

For Chapter 4 there were no additional co-authors beyond the members of my supervisory committee. This chapter has not yet been submitted for publication.

Chapters 5 and 6 were also based on the MOST database. My MOST sponsor was Michael Nevitt. These chapters have not yet been submitted for publication but when they are the co-authors will be the members of my supervisory committee as well as Jingbo Niu (MOST; participant selection and statistical analysis), John Lynch (MOST; study design), Neil Segal (MOST; study design), Jasvinder Singh (MOST; study design) and Michael Nevitt (MOST, study design).
Acknowledgements

I would like to extend heartfelt thanks to my supervisors, Linda Mclean and Elsie Culham, and to my advisor, T. Derek V. Cooke. Thank you, Linda, for mentoring me in the world of research and rehabilitation science. I am eternally grateful for your assistance with statistical analysis and writing, and your ability to help me see the big picture. Thank you, Elsie, for encouraging me to switch thesis topics, so that I might finish my PhD in a reasonable time-frame, and for agreeing to be my co-supervisor. Thank you too for your patience and help with organizing my writing so that it makes sense. Thank you Derek, for suggesting my thesis topics, for introducing me to the MOST database and for providing me with a new “orthopaedic” perspective.

I would like to thank the faculty and staff in the School of Rehabilitation Therapy, for their friendliness and encouragement to myself and all graduate students. In particular, I thank Debra Hamilton, Jean Jeffrey and Sharon David for their patience and assistance. Unfortunately I have not had the opportunity to have much contact recently with my fellow students; however I wish to thank Cindy Auchincloss, Charla Gray, M.J. O’Donovan and Kamary Coriolano for their support and for keeping me in touch with the student experience.

I wish to extend my gratitude to the people at MOST, especially Jean Hietpas, for their patience and understanding. I also wish to thank Orthopedic Imaging and Alignment Services (OAISYS) for the free use of their Surveyor™ software, and Chris Wale for the customizations required for my studies. I offer my appreciation to Margaret Bollen and Anneliese Kohar for performing the full-length radiographs needed for Chapter 4. I thank David Felson and the Physiotherapy Foundation of Canada for research grants which helped me to complete Chapters 3 and 4, respectively.

Finally, I extend sincere thanks to my family and friends. Thank you to my parents for not sighing too deeply when I announced that I was returning to school. Thank you to my
running friends (especially Emily Beedell) for keeping me fit and socialized. Thank you to
Joseph Federico and the gang at Physiotherapy on Kent, for allowing me to continue my clinical
work in a flexible and friendly environment.

And the biggest thank you goes to my wonderful husband, John, for supporting me in my
life-long dream of doing research. I love you so much, and thoroughly appreciate the extra child
care, dish-duty, patience and support that you have given me. I truly could not have done this
without you. To my “little guys”, Everett and Devyn, who have never known a Mom who hasn’t
been a student, I love you too. Thank you for helping me to learn the true meaning of “multi-
tasking”. We’ll see about the cat …
# Table of Contents

Abstract ........................................................................................................................................... ii

Co-Authorship ................................................................................................................................. iv

Acknowledgements ........................................................................................................................... v

Table of Contents ............................................................................................................................... vii

List of Figures .................................................................................................................................... xii

List of Tables ....................................................................................................................................... xiii

List of Abbreviations ........................................................................................................................ xiv

Chapter 1 Introduction .................................................................................................................... 1

1.1 Knee Osteoarthritis .................................................................................................................. 1

1.2 Measurement of Frontal-Plane Alignment .............................................................................. 5

1.3 Measurement of Radiographic Knee Osteoarthritis Severity ............................................... 6

1.4 Statement of Purpose ............................................................................................................... 7

1.5 Thesis Overview ...................................................................................................................... 7

Chapter 2 Literature Review ......................................................................................................... 10

2.1 Introduction .............................................................................................................................. 10

2.2 Tibiofemoral Frontal-Plane Alignment .................................................................................... 10

2.2.1 Frontal-Plane Alignment and Risk for Onset and Progression of Tibiofemoral
Osteoarthritis ................................................................................................................................. 11

2.2.2 Assessment of Frontal-Plane Alignment ............................................................................ 12

2.2.2.1 Imaging Methods ........................................................................................................... 12

2.2.2.2 Non-Imaging Methods .................................................................................................. 17

2.3 Measurement of Tibiofemoral Osteoarthritis Severity ......................................................... 24

2.3.1 Severity Measurements from Radiographs ...................................................................... 24

2.3.1.1 Methods Used to Acquire Suitable Radiographs ......................................................... 25

2.3.1.2 Global Scales .................................................................................................................. 27

2.3.1.3 Composite Scales .......................................................................................................... 32

2.3.1.4 Individual Osteoarthritis Feature Scales and Measurements .................................... 34

2.3.2 Severity Measurements from Magnetic Resonance Images .......................................... 40

2.3.2.1 Ordinal Scales ................................................................................................................ 41

2.3.2.2 Continuous Measurements ............................................................................................ 43

2.4 Concluding Remarks .............................................................................................................. 44
List of Figures

Figure 2-1: Varus knee illustrating the mechanical and anatomic axes and angles. .................. 13
Figure 3-1: Diagram of a full-length lower limb radiograph with a varus alignment. ............ 48
Figure 3-2: Mean offsets (with 95% confidence intervals) between the hip-knee-ankle (HKA) angle and the different methods of determining the femoral shaft-tibial shaft (FS-TS) angles, for each alignment group. ........................................................................................................ 59
Figure 4-1: Participant set-up for the radiograph and first photograph. ......................... 73
Figure 4-2: Calibrated template, used to position the participants’ feet accurately and to measure lower extremity rotation. ........................................................................................................ 74
Figure 4-3: Determination of the hip-knee-ankle (HKA) angle with a full-length lower extremity radiograph. ........................................................................................................ 78
Figure 4-4: Determination of the hip-knee-ankle angle with a full-length lower extremity photograph (HKA-P). ........................................................................................................ 82
Figure 4-5: Bland-Altman plot of intra-rater reliability for Reader 1. ............................... 92
Figure 4-6: Bland-Altman plot for the hip-knee-ankle (HKA) angle and the hip-knee-ankle angle measured from a photograph (HKA-P) (concurrent validity). ......................................................... 93
Figure 6-1: Radiographic grade plotted against the WORMS composite score for 72 knees with a range of osteoarthritis severity. ................................................................. 143
List of Tables

Table 3-1: Demographic data, with mean and standard deviation, for each alignment group......56
Table 3-2: Means and 95% confidence intervals (CI) for lower limb angles and (HKA - FS-TS) offsets, divided by sex...............................................................57
Table 3-3: Pearson correlations (r) between the hip-knee-ankle (HKA) angle and the different methods of measuring the femoral shaft-tibial shaft (FS-TS) angle..............................................61
Table 4-1: Demographic characteristics of the participant sample.................................................................86
Table 4-2: Pearson’s correlations (r) and Bland-Altman biases between estimated hip-knee-ankle (HKA) angles calculated with different estimates of the knee and ankle joint centres and the proximal femoral point, and the actual HKA.............................................................................88
Table 4-3: Hip-knee-ankle angle assessed from a photograph (HKA-P) and hip-knee-ankle (HKA) angle results assessed from a radiograph for the right knee.........................................................90
Table 4-4: Intraclass correlation coefficient (ICC_{2,1}) and Bland-Altman analysis results for intra-rater, inter-rater and test-retest reliability for the hip-knee-ankle angle assessed from a photograph (right knee)..........................................................................................................................91
Table 5-1: Description of participant samples..................................................................................................115
Table 5-2: Unicompartmental osteoarthritis grades (UCOAG) for intra-rater, inter-rater and test-retest reliability. .................................................................................................................................116
Table 5-3: Measurements of intra-rater, inter-rater and test-retest reliability for the unicompartmental osteoarthritis grade (UCOAG).........................................................................................118
Table 6-1: Description of participant samples [mean (standard deviation)]..................................................140
Table 6-2: KL, OARSI JSN and UCOAG grades and WORMS composite scores for concurrent validity and sensitivity to change.............................................................................................................................141
Table 6-3: Spearman’s rank correlation coefficients (r) for concurrent validity of several methods of radiographic knee osteoarthritis assessment.................................................................144
Table 6-4: Spearman’s rank correlation coefficients (r) for sensitivity to change over 30 months of several methods of radiographic knee osteoarthritis assessment........................................145
Table 7-1: Frequency of radiographs with medial and lateral tibiofemoral compartments designated as “most-affected”, by the Multicenter Osteoarthritis Study (MOST) and by the readers.........................................................................................................................168
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>Risk of committing a Type I error</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ASIS</td>
<td>Anterior superior iliac spine</td>
</tr>
<tr>
<td>β</td>
<td>Risk of committing a Type II error</td>
</tr>
<tr>
<td>BLOKS</td>
<td>Boston Leeds Osteoarthritis Knee Score</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index (kg/m²)</td>
</tr>
<tr>
<td>CHECK</td>
<td>Cohort Hip and Cohort Knee study</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>cm</td>
<td>Centimetres</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>DMOAD</td>
<td>Disease-modifying osteoarthritis drug</td>
</tr>
<tr>
<td>FS-TS</td>
<td>Femoral shaft – tibial shaft (angle, °)</td>
</tr>
<tr>
<td>HKA</td>
<td>Hip-knee-ankle (angle, °)</td>
</tr>
<tr>
<td>HKA-P</td>
<td>Hip-knee-ankle angle estimated from a photograph (°)</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>JSN</td>
<td>Joint space narrowing (ordinal scale, from 0-3 or 0-4)</td>
</tr>
<tr>
<td>JSW</td>
<td>Joint space width (continuous scale, often measured in millimetres)</td>
</tr>
<tr>
<td>KIDA</td>
<td>Knee images digital analysis</td>
</tr>
<tr>
<td>KL</td>
<td>Kellgren Lawrence</td>
</tr>
<tr>
<td>KOACAD</td>
<td>Knee OA computer-aided diagnosis</td>
</tr>
<tr>
<td>LDLDA</td>
<td>Logically derived line drawing atlas</td>
</tr>
<tr>
<td>LE</td>
<td>Lower extremity</td>
</tr>
<tr>
<td>MDC</td>
<td>Minimal detectable change</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>mm</td>
<td>Millimetres</td>
</tr>
<tr>
<td>MOAKS</td>
<td>MRI Osteoarthritis Knee Score</td>
</tr>
<tr>
<td>MOST</td>
<td>Multicenter Osteoarthritis Study</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance image/imaging</td>
</tr>
<tr>
<td>MTP</td>
<td>Metatarsophalangeal</td>
</tr>
<tr>
<td>OA</td>
<td>Osteoarthritis</td>
</tr>
<tr>
<td>OAISYS</td>
<td>Orthopedic Alignment and Imaging Systems, Inc.</td>
</tr>
<tr>
<td>OARSI</td>
<td>Osteoarthritis Research Society International</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PA</td>
<td>Posteroanterior</td>
</tr>
<tr>
<td>SPS</td>
<td>Superior pubic symphysis</td>
</tr>
<tr>
<td>SRM</td>
<td>Standardized response mean</td>
</tr>
<tr>
<td>r</td>
<td>Correlation coefficient</td>
</tr>
<tr>
<td>ROAD</td>
<td>Research on OA against disability</td>
</tr>
<tr>
<td>TF</td>
<td>Tibiofemoral</td>
</tr>
<tr>
<td>UCOAG</td>
<td>Unicompartmental osteoarthritis grade</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual analogue scale</td>
</tr>
<tr>
<td>WOMAC</td>
<td>Western Ontario and McMaster Universities Index</td>
</tr>
<tr>
<td>WORMS</td>
<td>Whole-organ magnetic resonance imaging score</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>Chi-square</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Knee Osteoarthritis

Osteoarthritis (OA) is a progressive joint disease hallmarked by cartilage and bone breakdown. It is a significant cause of pain and disability in our aging population. In knee OA, excessive or prolonged force or instability leads to fibrillation and thinning of the articular cartilage. Associated with cartilage changes, the periarticular bone remodels, causing osteophytes. While osteophytes are often thought of as contributing to the progression of OA, they may actually help to stabilize the joint. Erosion of the subchondral bone occurs as the cartilage continues to wear. Deeper into the bone structure, areas of sclerosis and cysts form.

It has been acknowledged recently that other tissues are also affected in knee OA. Ligaments and menisci degenerate, sometimes even before cartilage damage can be appreciated on a radiograph. This can lead to increased cartilage wear and joint instability, creating a cycle of destruction. The synovium becomes mildly inflamed and the joint capsule thickens. These whole joint changes ultimately cause pain, deformity and disability in many people.

The impact of OA is substantial, as it affects 13% of Canadians. By 2040, 71% of all Canadians over 71 years of age are expected to have some type of OA. Beyond the pain and functional impairment this will cause, it will likely create an extensive burden on health care and society.

The knee is the most common joint to be affected by OA. Although no Canadian data are available for knee OA specifically, estimates of the prevalence of knee OA in older adults...
world-wide vary considerably. Prevalence ranges from 5.4% in Italy to 38% in Korea. These numbers show the rate at which the population is affected by knee OA, and suggest that a significant portion of older adults, at least one in twenty, and up to one in three, may be dealing with knee pain, stiffness and related disability. Murphy et al. analysing results from the Johnson County study, concluded that the lifetime risk of symptomatic knee OA for individuals living in a rural or suburban part of North Carolina, United States was 44.7%, meaning that almost half of the population of this part of America would have this painful condition by the time they were 85 years old.

The great variability in prevalence estimates might be due to differences in the sampled populations. The specific recruitment strategies and entry criteria were different for each study. In particular, the age ranges of recruited participants and the age categories used for analyses varied widely. The variability in prevalence might also partly be due to differences in the definition of knee OA. Felson et al. reported that self-report of OA and knee pain gave much higher prevalence rates than a radiographic definition of knee OA. However, because many individuals with radiographic evidence of OA are asymptomatic, a knee OA definition which includes both radiographic and clinical criteria is most representative. All of the prevalence studies noted above included symptomatic, radiographic knee OA as their prevalence criteria, although the actual radiographic criteria and symptom assessments varied. Finally, genetic and lifestyle factors prominent in the population might also play a role in the differing prevalence rates across continents.

The main risk factors for knee OA are increasing age, obesity and prior knee injury. Knee OA incidence increases significantly with age. For example, the odds ratio (OR) for having knee OA at age 65 and older, compared to individuals less than 35 years old, is 28.4.
Obesity is also a well-documented risk factor for knee OA\textsuperscript{17-20}. Men and women with a body mass index (BMI) in the obese range (30 kg/m\textsuperscript{3} to 35 kg/m\textsuperscript{3}) have ORs of 4.0 and 3.5 respectively compared to individuals with a BMI in the normal range (18.5 kg/m\textsuperscript{3} to 25 kg/m\textsuperscript{3})\textsuperscript{20}. Traumatic knee injury is a considerable risk factor, especially anterior cruciate ligament tears, meniscal damage and tibial plateau fractures\textsuperscript{5,21-24}. Even surgical repair does not fully mitigate the risk of developing subsequent knee OA\textsuperscript{19,22,25,26}.

Several less-important but relevant risk factors include ethnicity and heredity. There is a significant risk of knee OA if first degree relatives have it\textsuperscript{5,20}. Another unmodifiable risk factor is female sex,\textsuperscript{16} although the reasons for this are unclear. Repetitive movements like squatting and kneeling, which may be associated with certain occupations, are associated with an increased risk of knee OA\textsuperscript{5,20,27,28}. A lifestyle of high levels of physical activity might increase the risk of OA, although this is controversial and might be more associated with an increase in the likelihood of knee injury\textsuperscript{19,29-34}. Varus or valgus deformity may be a risk of knee OA incidence\textsuperscript{19,35} and is definitely a risk for the progression of existing disease\textsuperscript{36}. Muscle weakness, decreased proprioception and ligament laxity are also considered risk factors\textsuperscript{5,19,37}. Surprisingly, heavy smoking offers a strongly protective effect with respect to knee OA\textsuperscript{18,20}.

The monetary costs associated with OA are considerable. While there are few Canadian reports that focus specifically on knee OA, a recent Ontario study compared 1474 individuals with all types of OA to 4422 matched [for age, sex and residence (rural or urban)] individuals regarding costs for physician services, hospitalizations and outpatient costs\textsuperscript{38}. The cost for those with OA was $2233 per year compared to $1033 for the controls. An Ontario study found that individuals with all types of OA incurred an average of $768 in drug costs annually\textsuperscript{39}. The majority of this amount ($569) was spent on arthritis drugs and other prescription drugs, with
lesser amounts spent on gastroprotective drugs ($88) and complementary medicine products ($51)\textsuperscript{39}. Total joint arthroplasty, a common treatment for late-stage knee OA, was estimated to cost greater than $21 000 in pre-surgical, surgical and post-surgical costs per procedure in Ontario in 2007\textsuperscript{40}. Despite this price, knee replacement surgery was deemed cost-effective over time, as ongoing costs decrease considerably after surgery\textsuperscript{40,41}. Sixty percent of individuals with disabling knee or hip OA in a 2005 Ontario study reported personal costs associated with their condition, with an average annual cost of $12 200, primarily due to lost income for the patient and their informal caregivers and the need for assistance with chores around the home\textsuperscript{42}. These costs did not include prescription and non-prescription drugs, or non-drug treatments such as physiotherapy, so the actual costs might be much higher.

Beyond the monetary costs to the individual and health care system, there are psychological and social burdens of knee OA, caused by pain and disability. On a population basis, mobility limitations progress slowly, but on an individual basis some people’s function declines quite quickly\textsuperscript{43}. Even early in the course of the disease, individuals with knee OA may limit their activities\textsuperscript{44}. Pain is highly related to physical functioning and limitation of activities but loss of knee flexion and quadriceps strength also has negative effects on function\textsuperscript{45}. Depression, social isolation and poor quality of life are also associated with knee OA\textsuperscript{46-50}. Low to moderate satisfaction with physical leisure activities, travel, hobbies and social events, which are rated as highly important in a person’s life, has been reported because of knee OA and the pain and intrusiveness associated with it\textsuperscript{51}. This is particularly true for younger individuals, who might perceive the intrusion of OA in their lives as greater than for older individuals\textsuperscript{51}. 

4
1.2 Measurement of Frontal-Plane Alignment

Because malalignment of the lower extremity (LE) in the frontal plane is a risk factor for the onset and progression of knee OA, it must be assessed and monitored. The presence of varus or valgus alignment may suggest the need for early intervention, for example, orthotics, braces or surgical correction (tibial osteotomy). An accurate measurement of alignment is also essential for proper placement of the implant during knee arthroplasty surgery. Proper placement resulting in restoration of neutral alignment ensures more even load distribution and prevention of premature wear and loosening of the implanted joints. Frontal-plane alignment is determined by calculating the angle from the centre of the femoral head to the centre of the knee to the centre of the ankle [hip-knee-ankle (HKA) angle], as seen on a full-length LE radiograph. However, because of perceived difficulties with obtaining these radiographs, alignment is often measured as the angle from a point at the mid-shaft of the femur to the centre of the knee to a point at the mid-shaft of the tibia [femoral shaft-tibial shaft (FS-TS) angle], as seen on a knee radiograph. While this is a common practice, it is unclear whether the FS-TS angle is a close estimation of the HKA angle, and suitable for use in research and surgical planning. Research is needed to fully determine the relationship between these two angles.

There is also a need to measure frontal-plane alignment for screening populations for malalignment as a risk factor for knee OA and for monitoring individuals with malalignment in a clinical setting. Since radiography, and the resultant exposure to ionizing radiation, is not reasonable in these instances, clinical methods have been developed, such as goniometry and the distance from the knee to a plumb line dangled from a haemostat held between the legs. Some of these methods are poorly correlated to the HKA angle, while others are highly
correlated but with a large difference between the resulting measurement and the HKA angle \(^6\). One method that holds promise is to measure the HKA angle from a full-length LE photograph \(^7\). This simple, fast and inexpensive method needs to be further tested on individuals of different ages, with a variety of body types, in order to assess its suitability for screening those at risk for knee OA. The best method to estimate the joint centres on a photograph must also be elucidated.

1.3 Measurement of Radiographic Knee Osteoarthritis Severity

Diagnosis of knee OA is based on symptoms of pain and stiffness, and the presence of OA changes on a knee radiograph. Radiographs are also used to monitor change and for treatment planning. Grading scales are applied to knee radiographs to rate the severity of OA. Current scales vary from poor to excellent in their reliability \(^72\)\(^-\)\(^74\), poor to moderate in their sensitivity to change \(^75\)\(^\)\(^7\)\(^6\), and negligible to moderate in their relationship to other knee OA features (pain, alignment, function) \(^77\)\(^-\)\(^79\). Many scales emphasize a single feature of knee OA, which may limit their usefulness for different presentations of the disease \(^8\)\(^0\), \(^8\)\(^1\). To address these issues, an original scale, the unicompartmental osteoarthritis grade (UCOAG), was introduced in 1999 by Cooke et al. \(^8\)\(^2\). It includes several features of knee OA and was designed to have high reliability and sensitivity to change, and to be associated with changes in alignment and deformity \(^8\)\(^2\). Inter-rater reliability has been assessed on anteroposterior knee radiographs, taken with the knee in full extension \(^8\)\(^2\), however complete psychometric testing has not been performed. Because knee radiographs for research purposes now are typically performed with the knee in flexion, it is important that the psychometric properties of the UCOAG grades obtained from radiographs taken with the knee in flexion be determined.
1.4 Statement of Purpose

The overall objective of this thesis is to evaluate tools which may be used to assess for knee OA risk and to monitor the severity and progression of the disease, for clinical decision making and research purposes. The first goal is to investigate the suitability of knee radiographs and photographs for estimating frontal-plane LE alignment. The second goal of the thesis is to determine the psychometric properties of the UCOAG obtained from posteroanterior fixed-flexion radiographs with respect to its ability to measure the severity and progression of knee OA.

1.5 Thesis Overview

Chapter 2 consists of a review of the literature in two areas. The first part of the chapter is an in-depth review of the methods currently used to measure frontal-plane alignment at the knee, including imaging and clinical methods. The second part of Chapter 2 is an in-depth review of the methods used to assess the severity of knee OA as seen on a knee radiograph and on a magnetic resonance image (MRI). Only literature presented in English is included in the review.

Chapters 3 and 4 present two studies on the measurement of frontal-plane LE alignment. Chapter 3 addresses the use of knee radiographs as a substitute for full-length LE radiographs in the measurement of frontal-plane alignment. The relationship between the HKA angle as measured from a full-length LE radiograph and the FS-TS angle as measured from a knee radiograph is uncertain. Therefore, the purpose of this study was to investigate this relationship in individuals with, or at high risk of knee OA. More specifically, we wished to study how the two angles are related relative to the direction and magnitude of LE deformity and whether the length of the axes used for the calculation of the FS-TS angle influences its relationship to the
HKA angle. This will inform researchers whether a knee radiograph is suitable for their investigations, or whether a full-length radiograph is required.

Chapter 4 presents a relatively novel method of measuring frontal-plane alignment. A pelvis-to-ankle photograph was used to estimate the HKA angle. Points were placed on the photograph at the pelvis, knee and ankle. Several methods of placing these points were investigated to find the combination best able to estimate the HKA angle measured from a radiograph. The intra- and inter-rater reliability of this method was tested, along with its correlation to the HKA angle. Such photographs are expected to be useful for screening and monitoring purposes where ionizing radiation is not desired and will be of use to physiotherapists and other clinicians who do not have easy access to radiographs.

Chapters 5 and 6 investigate the psychometric properties of the UCOAG for the assessment of the severity of tibiofemoral OA from a frontal-plane knee radiograph\(^\text{82}\). Initial reliability testing was performed using anteroposterior knee radiographs\(^\text{82}\). More complete intra-rater, inter-rater and test-retest reliability testing is presented in Chapter 5, using posteroanterior fixed-flexion radiographs, in individuals with or at risk of knee OA. Chapter 6 presents the results of validity and sensitivity-to-change analyses of the UCOAG obtained from posteroanterior fixed-flexion radiographs from the same database. The first goal of this study was to determine if knee OA severity assessed with the UCOAG and two existing scales is valid, compared to knee OA severity as seen on an MRI. The second goal was to assess the sensitivity to change over 30 months for the three scales. The results presented in Chapters 5 and 6 will determine whether the UCOAG has sufficient reliability, validity and sensitivity to change to support its use as a scale to grade tibiofemoral OA severity from a radiograph for research and clinical purposes.
Chapter 7 is a final discussion of the overall findings from the four studies. Recommendations for clinical and research applications are presented, along with recommendations for future research.
Chapter 2

Literature Review

2.1 Introduction

The first part of this literature review will address the measurement of tibiofemoral (TF) frontal-plane alignment. Accurate and responsive measurement of alignment is important to assess risk of onset of knee OA and to monitor its progression. The measurement of knee OA severity and progression from radiographs and magnetic resonance images (MRI) will be discussed in the second part of this chapter.

2.2 Tibiofemoral Frontal-Plane Alignment

Being knock-kneed (genu valgum, or valgus) or bow-legged (genu varum, or varus) is a risk factor for the development and progression of TF OA\textsuperscript{35, 52, 53, 83-86}. Therefore it is important to assess lower extremity (LE) frontal-plane alignment to evaluate risk for OA and to monitor its progression. While a full-length LE radiograph is considered the criterion standard to measure frontal-plane alignment, other methods such as knee radiographs and photographs are used as well. What follows is a review of the literature exploring the association of frontal-plane alignment with TF OA and a review of methods of measurement. OA is also common in the patellofemoral compartment of the knee; however this review will focus solely on the TF compartments.
2.2.1 Frontal-Plane Alignment and Risk for Onset and Progression of Tibiofemoral Osteoarthritis

Alignment has been suggested as one of the genetic factors associated with knee OA development [87]. Varus alignment, the most common frontal-plane malalignment, leads to increased loading in the medial TF compartment [52]. There is some evidence that frontal-plane malalignment is a risk factor for the onset of knee OA. One United States-based osteoarthritis study showed non-significant odds ratios (OR) (1.07 to 1.10, p > 0.05) for varus alignment on the incidence of medial TF OA over nine years [88]. However, others have shown small to moderate risks (OR of 1.49 and 2.06, p < 0.05) associated with varus alignment on the incidence of knee OA [53, 83]. The influence of valgus alignment on incident knee OA is not as strong. Many have failed to find an increased risk of incident knee OA with valgus alignment [53, 83, 88]. However, a recent study by Felson et al. [35] documented an OR of 2.5 (p = 0.04) for incidence of lateral TF OA in a participant sample with four degrees or greater of valgus deformity. The combination of malalignment and being overweight or obese increases the risk of incident OA beyond the risk caused by malalignment alone in individuals of normal weight [83]. There is evidence of a reduced risk of incident medial TF compartment OA in individuals with valgus deformity (OR 0.84 per degree valgus, p < 0.05) [89].

It is well-accepted that varus alignment is associated with progression of OA in the medial TF compartment, with statistically significant ORs for progression of OA ranging from 2.90 to 10.96 (p < 0.05) [52, 53, 83-86]. In these studies progression of OA was monitored over 18 months to six and a half years, using sequential radiographs or MRI. Similarly, valgus alignment, which increases loading in the lateral TF compartment [52], is highly associated with progression of
OA in the lateral TF compartment (OR 3.42 to 10.44, p < 0.05)\textsuperscript{35, 52, 84-86}. The odds of progression increase with the degree of malalignment\textsuperscript{35, 52, 84, 85, 90}.

Valgus alignment also decreases loading in the medial TF compartment. Valgus alignment is therefore associated with a lower odds of progression of OA in the medial compartment (OR 0.34 to 0.88, p < 0.05)\textsuperscript{53, 86, 89}. Furthermore, there is reduced medial TF compartment cartilage loss in individuals with OA with neutral and valgus alignment, compared to those with varus alignment\textsuperscript{91}. The opposite is also true; varus alignment is associated with a reduction of cartilage loss in the lateral TF compartment, OR 0.12 (p < 0.05)\textsuperscript{86, 91}.

\textbf{2.2.2 Assessment of Frontal-Plane Alignment}

\textbf{2.2.2.1 Imaging Methods}

The criterion standard measure of frontal-plane LE alignment is the hip-knee-ankle (HKA) angle, also known as the mechanical angle\textsuperscript{61, 62}. This is the angle subtended by a line from the centre of the femoral head to the knee (femoral mechanical axis) with a line from the knee to the centre of the tibial plafond or ankle talus (tibial mechanical axis). See Figure 2-1. For the purpose of this thesis, varus angles will be denoted negative and valgus angles positive\textsuperscript{61}.

“Normal” alignment in healthy adults is generally considered to be 1° to 1.5° of varus, or -1° to -1.5°\textsuperscript{64, 65, 92}.

\textbf{2.2.2.1.1 Full-length Lower Extremity Radiographs}

The HKA angle is measured from a full-length LE radiograph\textsuperscript{93}. In today’s diagnostic imaging departments, commonly three or four digital radiographs are “stitched” together to form a single full-length radiograph\textsuperscript{94-96}. These radiographs were used in recent large research studies.
Figure 2-1: Varus knee illustrating the mechanical and anatomic axes and angles. Modified from Cooke & Sled. 

FM – femoral mechanical axis
TM – tibial mechanical axis,
FA – femoral anatomic axis
TA – tibial anatomic axis
HKA – hip-knee-ankle angle (mechanical angle)
FS-TS – femoral shaft-tibial shaft angle (anatomic angle)
The FS-TS angle is approximately 4° to 5° valgus compared to the HKA angle.
for the assessment of frontal-plane alignment but are less often used for smaller studies and in clinical situations. The use of a full-length radiograph allows the effect of deformities of the femoral and tibial shafts to be appreciated as influencing the HKA angle. Arguments against full-length radiographs, compared to knee radiographs, include that they require specialized equipment and technician training, are more costly and expose the patients or participants to higher doses of radiation, particularly to the pelvis. However, the current use of digital technology has reduced the radiation exposure and technical difficulty to some extent.

The points used for determining the HKA angle are somewhat standardized. The centre of the femoral head is found by placing a circle template over the femoral head on the radiograph, then marking the centre of this circle. There are several locations which may be used for the points at the knee. Many use a single point, often the centre of the tibial spines. Moreland et al. used a single point at the knee that was the mid-point of several measured knee locations (the mid-point between the medial and lateral contours of the knee at the joint line, the mid-point of a line drawn across the tibial plateau, the centre of the femoral notch, the centre of the tibial spines and midpoint of a line drawn across the femoral condyles). Others prefer to use the centre of the femoral intercondylar notch as the distal point for the femoral mechanical axis, and the centre of the tibial interspinous groove as the knee point for the tibial mechanical axis. Using two points at the knee is preferred because it allows for the investigation of the femoral and tibial contributions to the HKA angle, and to observe the presence of knee subluxation. See Figure 2-1. The centre of the talus or tibial plafond at the ankle is determined using a ruler placed on the radiograph.

With the increasing availability of digital radiographs, researchers have reported that there were no differences in HKA angle measurements calculated manually from radiograph film
using a goniometer, and those calculated using software and digital radiographs [intraclass correlation coefficient (ICC) > 0.93; Pearson’s correlation r = 0.98]\textsuperscript{101-103}. Computer-assisted methods do take significantly less time (1.08 minutes per radiograph versus 4.9 minutes, p < 0.001) and are significantly easier to use as measured on a one-to-ten Likert scale (p = 0.03)\textsuperscript{101, 104, 105}. Intra-rater and inter-rater reliability were very high for both methods (computer-assisted methods ICC > 0.90, p < 0.05; manual measurements ICC > 0.86, p < 0.05)\textsuperscript{101, 104, 106, 107}. Generally, ICCs of less than 0.4 can be considered poor, ICCs between 0.4 and 0.75 can be considered fair to good and ICCs above 0.75 can be considered excellent\textsuperscript{108}. Computer-assisted methods are more precise. The minimal detectable change (MDC) stipulates the smallest change in the HKA angle that can be detected beyond random error\textsuperscript{109}. The MDC\textsubscript{95} for computer-assisted methods is 0.4°, compared to 1.6° for the manual method\textsuperscript{102}.

### 2.2.2.1.2 Knee Radiographs

Because full-length LE radiographs are not always available, knee radiographs are often used to estimate the HKA angle, as they are commonly taken for the clinical assessment of knee OA\textsuperscript{66, 83}. The angle calculated on a knee radiograph is called the femoral shaft-tibial shaft (FS-TS) angle, or the anatomic angle\textsuperscript{62}. This is the angle subtended by a line from the centre of the femoral shaft to the knee (femoral anatomic axis) and a line from the centre of the tibial shaft to the knee (tibial anatomic axis). The tibial anatomic axis is often very similar to the tibial mechanical axis. See Figure 2-1. Again, one or two points at the knee may be chosen to determine the anatomic axes\textsuperscript{110}. The femoral and tibial shaft points are generally measured 10 centimetres (cm) from the knee joint, to accommodate the portion of the long-bone shafts commonly seen on a knee radiograph\textsuperscript{62, 65}.
There are concerns that the FS-TS angle does not produce an accurate estimate of the HKA angle. The FS-TS angle is offset towards valgus compared to the HKA angle by approximately 4° to 6° for healthy individuals and 1.5° to 7° in individuals with knee OA, with a low to high correlation between the two measurements, \( r = 0.34 \) to 0.88, \( p < 0.005 \) in participants with knee OA. The offset tends to be greater in men than women, and one study surprisingly found no offset in a cohort of women without knee OA. While a correction of 4° to 6° to the FS-TS angle may be used in research to measure alignment, not all authors do this. Hinman et al. created a regression equation \[ \text{HKA} = 0.915 \times (\text{FS-TS}) + 13.895; r = 0.88 \] to estimate the HKA angle from the FS-TS angle; this equation has been used in at least one research study.

The FS-TS angle is more variable than the HKA angle. This variability is particularly important because the FS-TS angle is often used by surgeons in planning for surgical correction of deformity. The difference between the HKA and FS-TS angles is significantly greater in individuals with knee OA compared to healthy controls (t-test, \( p < 0.001 \)). In two studies, the FS-TS angle measured with a short femoral anatomic axis was 4.0° to 4.2° more valgus than the HKA angle, but with a long femoral anatomic axis the difference was 5.8° and when using the entire femoral shaft the difference was 4.9° to 5.9°. This illustrates how the shape of the femoral shaft has an impact on the relationship between the HKA angle and the FS-TS angle. In order of importance, lateral bowing of the femoral shaft, tibial bowing and the angle between the tibial plateau and the tibial shaft all influence the relationship between the HKA angle and the FS-TS angle. Therefore it has been recommended that the HKA angle, measured from a full-length LE radiograph should be used to ensure an accurate measurement of LE alignment.
2.2.2.2 Non-Imaging Methods

While full-length LE radiographs are the criterion standard for the calculation of the HKA angle, other non-radiographic options also have their place in research and patient care. Many non-imaging methods have advantages in that they can be performed in a physician’s or physiotherapist’s office. No ionizing radiation is administered, making these tests suitable for screening purposes and to monitor deformity over time. Tests with excellent correlation to the HKA angle could also be used for monitoring the effect of treatment for malalignment or OA. These tests can also be used on populations for which radiography would not be suitable, such as children, pregnant women and healthy control participants in clinical studies. Most are very quick and the results are available immediately.

2.2.2.2.1 Clinical Methods

Various clinical methods have been used to estimate the HKA angle. Goniometry in the standing weight-bearing position has been assessed in several research papers, with mixed results. Kraus et al. found a moderate relationship between goniometry measures and the HKA angle as measured from a full-length radiograph (Pearson’s $r = 0.70$, $p < 0.0001$). Test-retest reliability of the goniometric method was high (ICC 0.94). Hinman and colleagues and Riddle found much smaller relationships (Pearson’s $r = 0.32$, $p = 0.12$; Pearson’s $r = 0.50$ to $r = 0.54$ respectively, significance not mentioned) and suggested that goniometry should not substitute for a full-length LE radiograph. The technique places the arms of the goniometer along the centres of the thigh and calf, and therefore estimates the FS-TS angle rather than the HKA angle. However, correlations were not performed to compare goniometric estimates of the FS-TS angle with the FS-TS angle measured from radiographs.
Rather than measuring the FS-TS angle, another method used to assess alignment was to identify the horizontal centres of the ankle and knee joints, and to use the umbilicus as the proximal point. An extended-arm goniometer was used to measure the resulting angle. This reliable measure (test-retest reliability ICC 0.85) was highly correlated to the HKA angle \((r = 0.75, p < 0.001)\), with an offset of \(8.1^\circ\) varus, but again does not estimate the true HKA angle. Body mass index (BMI) did not influence this measurement.

Other techniques have been suggested. One method was to use callipers to measure the distance between the knees (for individuals with varus deformity) or ankles (for individuals with valgus deformity). The resulting horizontal measurement was highly correlated to the HKA angle (Pearson’s \(r = 0.76, p < 0.001\)). Another method used callipers to measure the distance from a plumb line held in a haemostat and dangled between the legs to the medial joint line of the knee (for individuals with varus alignment) or the medial malleolus (for individuals with valgus alignment) (Pearson’s \(r = 0.71, p < 0.001\) for the horizontal measurement from the plumb line to the knee or ankle relative to the HKA angle) and a third method used an inclinometer to determine the angle of the tibia with respect to the vertical (Pearson’s \(r = 0.80, p < 0.001; r = 0.84, p < 0.001\) for the tibial angle to the HKA angle). Regression equations were calculated to estimate the HKA angle.

2.2.2.2 Position- and Motion-Capture Methods

Infrared tracking systems are commonly used in gait studies, but have also been used for static measurements of LE alignment. Infrared tracking devices are taped onto a participant’s skin over pre-determined landmarks, and position-capture or motion-capture cameras are used to record the position of the markers. Knee and hip range of motion may be used to allow one
or more cameras to extrapolate the joint centres in two or three dimensions\textsuperscript{118}. In other cases, anthropomorphic measurements and palpation are used to estimate the joint centres\textsuperscript{67, 118, 119}. Controlling the exact location of the markers is difficult due to variations in soft tissue between individuals, contributing to error\textsuperscript{117}. In addition, movement of muscles and joints during gait or range of motion may cause the markers to move on the skin\textsuperscript{117}. Test-retest reliability measurements of $\pm 3^\circ$ for LE frontal-plane alignment measurements have been reported\textsuperscript{117}. In two studies of individuals with knee or hip OA, position-capture estimated the HKA angle to be $0.3^\circ$ to $3.5^\circ$ more varus compared to the true angle measured from the radiograph, with a high correlation between the HKA and the resulting angle (Pearson’s $r = 0.74$ to $r = 0.91$, $p = 0.001$)\textsuperscript{118, 119}. A three-dimensional gait analysis technique showed an even higher correlation with the HKA angle (Pearson’s $r = 0.93$, $p < 0.001$) with a varus offset of $3.9^\circ$ in patients with knee OA\textsuperscript{67}. One advantage of using position- or motion-capture is that alignment can be observed in individuals while also evaluating gait parameters\textsuperscript{119}. However, these methods tend to be time-consuming and require specialized equipment and expertise not commonly available in the clinical setting.

2.2.2.2.3 Frontal-Plane Photographs

Frontal-plane photographs are another approach for the assessment of LE frontal-plane alignment. Photographs are quick, inexpensive and can be performed in the clinical setting. Patients do not need to remain in one position for long\textsuperscript{120}. Photographs also have the advantage over other non-imaging methods of being a permanent record, useful for reanalysis or to demonstrate change over time. Goniometric measurements performed directly on a patient are awkward and subject to error\textsuperscript{66, 68}, but frontal-plane alignment measurements of the LEs...
determined with a protractor on a photograph are reliable (inter-rater reliability: ICC 0.700 to 0.839; test-retest reliability: ICC 0.627 to 0.904)\textsuperscript{70}.

Research supports the use of photographs to measure posture and frontal-plane alignment; these studies are often performed on children, youth and young adults\textsuperscript{121,122}. Analyses of frontal-plane LE alignment on photographs using computer software shows good to excellent intra-rater reliability (ICC\textsubscript{3,1} 0.67 to 0.96) and inter-rater reliability (ICC\textsubscript{2,1} 0.91 to 0.96)\textsuperscript{123}.

There have been two investigations where photographs were used specifically to estimate frontal-plane LE alignment\textsuperscript{70,71}. Schmitt et al.\textsuperscript{71} estimated the HKA angle while Moncrieff and Livingston estimated the FS-TS angle\textsuperscript{70}. Both used moderate numbers (16 to 20) of young, healthy participants with low to moderate BMIs\textsuperscript{70,71}. Intra-rater (ICC = 0.63 to 0.99, p < 0.05), inter-rater (ICC = 0.70 to 1.00, p < 0.05) and test-retest reliability (ICC = 0.65 to 0.90; p < 0.05) were moderate to excellent\textsuperscript{70,71}. The photographic estimation of the HKA angle was highly correlated to the measurement of the HKA angle from a full-length LE radiograph (Pearson’s r = 0.98, p < 0.001) with an offset of 0.9°\textsuperscript{71}. Moncrieff and Livingston\textsuperscript{70} did not compare their FS-TS angle measurements to radiographic FS-TS angle or HKA angle measurements.

Some features of these two studies should be addressed in further studies. Standardization of LE position is important for reliable and accurate measurements\textsuperscript{61,71,93}. Ideally, the stance position should be such that the knee will flex in the sagittal plane\textsuperscript{93}. Rotation of the LEs in standing varies considerably between individuals and a single stance position does not necessarily allow this to occur\textsuperscript{93}. Increased external rotation has been shown to increase the appearance of varus malalignment\textsuperscript{124}. Schmitt et al.\textsuperscript{71} had participants stand with their feet positioned straight forwards, while Moncrieff and Livingston\textsuperscript{70} had their participants stand either
with the medial borders of their feet touching or in a self-selected position. These positions might have produced undesirable limb rotation and have introduced variability and error. In fact, Schmitt et al. 71 also had some participants stand in 30° of external rotation, which produced a significant change in the HKA angle compared to participants standing with feet pointed forwards (ICC 0.66, p < 0.001 for the comparison of the HKA angle calculated with the subjects in these two LE rotation positions).

The precision of the instruments used is also important. Schmitt et al. 71 used custom computer software (mechanical desktop 4 power pack, AutoCAD 2000) to measure the HKA angle from the photographs to one tenth and one one-hundredth of a degree. Radiographic measurements of the HKA angle, analysed using computer-assisted techniques, are generally measured to one tenth of a degree 66, 92, 100. On the other hand, Moncrieff and Livingston 70 printed the photographs (20.3 cm by 15.2 cm) and used a goniometer with 18 cm arms and 1° gradations to measure the FS-TS angle. They estimated the FS-TS angle to one tenth of a degree. The small photographs and imprecise goniometer may have led to less-than-ideal precision of the measurements. Since the variation in alignment is only from approximately 15° varus to 15° valgus, and change occurs slowly, a more-precise measurement would be able to detect change earlier. Reliability coefficients (ICC) were higher for frontal-plane LE alignment determined with the computer-assisted techniques over the manual method 70, 71.

To estimate the HKA angle from a photograph, the centres of the hip, knee and ankle must be estimated 66, 69, 70. Sticky dots may be placed on the participant before the photograph is taken, or the centres may be estimated by identifying soft tissue or bony points on the LE photograph directly 66, 69, 70. Error may be introduced in either method because of differences in bone structure and soft tissue coverage, and differences in quadriceps contraction, which changes
the position of the patella \textsuperscript{70}. While reliability and validity studies of palpation of the anterior pelvic, knee and ankle bony prominences are lacking, palpation of posterior pelvic bony prominences is known to be associated with poor intra- and inter-rater reliability, and limited validity \textsuperscript{125,126}. If palpation and the placement of sticky dots are performed, the challenge is to palpate accurately. If measurements are taken from the radiograph without dots, palpation is not an issue, however soft tissue coverage and identification of landmarks may still be problematic.

There are several methods of identifying the centres of the femoral head, knee and ankle on a photograph \textsuperscript{67,116,118,119,127-129}. Many methods were initially introduced in position- and motion-capture studies, which use palpation to place markers on the skin \textsuperscript{67,116,118,119}. The knee centre is commonly estimated as the midpoint between the medial and lateral femoral condyles or the medial and lateral joint lines while the ankle centre is estimated as the midpoint between the medial and lateral malleoli, or the midpoint of the ankle joint \textsuperscript{67,116,118,119}.

The centre of the femoral head is the most difficult point to estimate \textsuperscript{118,127}. Neither Schmitt et al. \textsuperscript{71} or Moncrieff and Livingston \textsuperscript{70} attempted to identify this point on their photographs. Schmitt et al. \textsuperscript{71} instead chose the most proximal point for the anatomic axis as the centre of a line positioned horizontally across the uppermost thigh. The line was drawn on the photograph, not on the participant in the laboratory. Moncrieff and Livingston \textsuperscript{70} determined the most proximal point of the anatomic axis as a point marked on the participant at the intersection of a line drawn distally from the anterior superior iliac spine and a line drawn medially from the greater trochanter. While these points approximate the position of the shaft of the femur, the location of the centre of the femoral head should be estimated to determine the HKA angle.

Several methods of estimating the centre of the femoral head exist in the static- and motion-capture literature \textsuperscript{116,127-129}. While many of these methods are based on three dimensions,
they can be adapted to determine the femoral head position in the frontal plane $^{127,128}$. In one method the femoral head is estimated to be a certain percentage of the distance between the right and left anterior superior iliac spines (ASIS) medially and a certain percentage of this distance inferiorly $^{116,127,130}$. A variation of this method uses a percentage of the distance between the ASIS and the pubic tubercle for the inferior distance $^{128,131}$. Another method positions the centre of the femoral head 1.5 cm or 2.0 cm inferior to the midpoint between the ASIS and the top of the symphysis pubis $^{128,129}$. Calibration of the photograph with a ruler is required for this calculation. Finally, one study used separate regression equations for men and women to estimate the centre of the femoral head from the distance between the right and left ASIS $^{132}$. The validity (comparison with centre of the femoral head seen on a radiograph) of the regression equation method has been studied (Pearson’s correlation $r = 0.76$) $^{132}$ and the resulting HKA angle estimates from several of these methods have been compared with HKA angles measured from radiographs (Pearson’s correlations $r = 0.76$ to 0.93, with a bias of $0.3^\circ$ to $3.9^\circ$) $^{67,116,118,132}$.

In summary, because frontal-plane alignment is an important risk factor for the onset and especially the progression of knee OA, it is regularly assessed for research and clinical purposes. While the criterion standard measure of frontal-plane alignment is the HKA angle measured from a full-length LE radiograph, other methods such as knee radiographs and pelvis-to-ankle photographs can be used to estimate the HKA angle. These two methods will be studied for their ability to estimate the HKA angle, in Chapters 3 and 4.
2.3 Measurement of Tibiofemoral Osteoarthritis Severity

Assessment of the presence and severity of knee OA is performed for diagnosis, to monitor progression over time and to guide treatment decisions. Assessments are also used to guide participant inclusion or exclusion in research studies and to stratify participants according to OA severity. Individual characteristics such as biometrics (BMI, age etc.), involvement of other joints, family history and history of injury are commonly correlated to measures of knee OA severity to investigate risk factors. Studies of potentially disease-modifying OA drugs and other treatments also use knee OA severity assessments as outcome measures. In this section the literature on methods used to measure TF OA severity from radiographs and MRIs will be presented. Because the research presented in this thesis focuses on the medial and lateral TF compartments of the knee joint as observed on a frontal-plane radiograph, this review will as well. However, it is acknowledged that many of the newer scales also include the patellofemoral joint. As well, some scales are used for joints other than the knee, notably the hip and finger joints.

2.3.1 Severity Measurements from Radiographs

For both clinical and research purposes, the severity of knee OA is most commonly measured from radiographs. The acquisition of knee radiographs has evolved, with newer techniques providing more accurate and reliable measurements. This section will begin with a brief overview of the methods available to obtain the requisite knee radiographs. There are several methods of quantifying the severity of knee OA on a radiograph. Global scales tend to be
ordinal scales that have specific descriptions of the evidence required to determine which severity level is assigned to a joint. Composite scales score several features of OA individually, then add them to create a total score. Finally, grades or measurements representing levels of OA severity may be assigned to individual features of knee OA, most commonly osteophytes and joint space narrowing (JSN, an ordinal scale which estimates the width of the joint space on a scale of zero to three or four) or joint space width (JSW, a continuous scale which measures the width of the joint space, often in millimetres (mm)). Each of these methods is discussed in the following sections of the review.

2.3.1.1 Methods Used to Acquire Suitable Radiographs

For over 40 years radiographs have been taken with the patient weight-bearing equally on both legs with the knees in full extension and an anteroposterior x-ray beam directed horizontally at the level of the lower patella. In this position, however, the x-ray beam is not aligned with the tibial plateau. Alignment of the anterior and posterior margins of the tibial plateau with the x-ray beam is necessary in order to accurately assess the knee for OA changes, particularly JSN and JSW. Other disadvantages of radiographs taken with the knee in full extension are that variations in the degree of end-range extension can occur and the medial meniscus contributes to the observed joint space. Also in full extension an individual does not bear weight on the part of the femoral condyles which show the greatest wear; this instead occurs when the joint is in slight flexion. Therefore, four radiograph acquisition protocols with the knee in slight flexion have been developed. The semi-flexed protocol positions the patient standing with knees flexed approximately 7°. Fluoroscopy is used to visualize the anterior and posterior margins of the medial tibial plateau to ensure they are horizontal and
superimposed, and an anteroposterior radiograph is taken with the x-ray beam angled horizontally \[161\]. The other three protocols take the radiographs with a posteroanterior x-ray beam aligned parallel to the medial tibial plateau \[145, 159, 161-163\]. For the Lyon schuss protocol, the patient stands facing the radiograph cassette, with the thigh and knee placed against the cassette and the tip of the first toe in line with its lower edge, placing the knee in 25° to 30° of flexion \[144, 145\]. The x-ray beam is angled approximately 10° caudally. Fluoroscopy is used to visualize the tibial plateau as described for the semi-flexed protocol.

While protocols using fluoroscopy can achieve highly reproducible joint positions, the use of fluoroscopy is expensive, time-consuming and subjects the patient to additional ionizing radiation \[143\]. To overcome these problems, two non-fluoroscopic protocols were developed \[162, 163\]. In the metatarsophalangeal (MTP) protocol, the patient stands facing the radiograph cassette and his or her first MTP joint is placed below the edge of the radiograph cassette \[162\]. Then the knee is flexed to touch the cassette, producing approximately 7° of knee flexion \[162\]. The feet are externally rotated 7.5° to 10° degrees and the position recorded on a foot map and the x-ray beam is positioned horizontally \[144, 162\].

The fixed-flexion protocol is based on the Lyon schuss protocol but without fluoroscopy and with the x-ray beam directed caudally at a consistent 10° angle \[163\]. Surprisingly, even the Lyon schuss method, with fluoroscopy, only accurately positions the medial tibial plateau 77.9% of the time \[164\]. Therefore the fixed-flexion protocol was modified such that if the resulting radiograph showed poor positioning of the tibial plateau, up to four repeat radiographs could be taken, with the beam angled a further 1° to 2° cranially or caudally \[165-167\]. Currently the fluoroscopic protocols are being phased out in favour of digital radiography and two ongoing large multicentre studies use the fixed-flexion protocol \[2, 97, 144, 168\]. Use of a standardized
technique with well-trained technicians has been emphasized, especially with respect to limb rotation, knee flexion and equal weight bearing.61, 143, 169.

2.3.1.2 Global Scales

Global scales are ordinal scales that have specific descriptions for each grade.81, 146-148 Each level describes one or more features of OA that must be met for that particular level to be ascribed to a radiographic image. Global scales require an individual’s particular presentation of OA to “fit” the criteria for a given level of the scale. The earliest and by far the most commonly-used global scale is the Kellgren-Lawrence (KL) grading scale.81 Others include those developed by Ahlback,146 Sundaram et al.,148 and Brandt et al.147.

2.3.1.2.1 Kellgren-Lawrence Grading Scale

The KL scale, first described in 1957, gives an overall score of OA severity from zero to four.81, 170 In their initial publication the authors considered the following features evidence of OA: osteophytes on the joint margins or the tibial spines; periarticular ossicles; narrowing of joint space associated with sclerosis of subchondral bone; small pseudocystic areas, usually in the subchondral bone; and altered shape of the bone ends.81 Both tibiofemoral compartments of the knee were assessed using a standard set of radiographs for reference.81 Considering all features of OA, a grade of zero (no OA), one (doubtful OA), two (minimal OA), three (moderate OA), or four (severe OA) was given.81 Inter-rater reliability was reported (Pearson’s r = 0.83), but the authors acknowledged that one of the two readers consistently assessed the radiographs as showing more severe OA, illustrating the difficulty of using Pearson’s correlation coefficients to
adequately assess reliability. Intra-rater reliability was the same (Pearson’s correlation of $r = 0.83$)\textsuperscript{81}.

In 1963 an atlas (republished in 2005\textsuperscript{171}) was produced by Kellgren et al.\textsuperscript{170} which included written descriptions of each grade:

- **Grade 1:** doubtful narrowing of joint space and possible osteophytic lipping,
- **Grade 2:** definite osteophytes and possible narrowing of joint space,
- **Grade 3:** moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends, and
- **Grade 4:** large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends.

Later, in a 1977 publication, Lawrence\textsuperscript{172} described the grades as such:

- **Grade 1:** minute osteophyte of doubtful significance the only feature
- **Grade 2:** definite osteophyte, joint space unimpaired
- **Grade 3:** moderate diminution of joint space, and
- **Grade 4:** joint space greatly impaired, subchondral sclerosis.

OA incidence is defined by a KL grade of two\textsuperscript{81}. The KL scale was adopted by the World Health Organization in 1961 and has remained the most prominent scale for diagnosing OA and grading OA severity\textsuperscript{173}. Its use as a standard for radiographic knee OA was reconfirmed at the third International Symposium on Rheumatic Disease in New York in 1966\textsuperscript{174}.

Despite its widespread use, there are concerns about the KL scale\textsuperscript{173, 175, 176}. As evident in the above descriptions, osteophytes must be present for a KL grade other than zero to be given. Osteophytes often appear early in the course of the disease, before JSN occurs, because the bone responds to the stresses of OA earlier than the articular cartilage wears. However, the hallmark
feature of OA may be cartilage loss, which is typically estimated on a radiograph as JSN. Even so, there are a wide variety of presentations of knee OA. Therefore, if the presence of osteophytes is necessary to diagnose radiographic OA, some individuals might be misdiagnosed using the KL scale. As an example, one study showed that nine participants with significant JSN but no osteophytes (and therefore KL grade zero) all had definite OA changes identified on arthroscopy. For the Framingham Osteoarthritis Study, Felson et al. created a new KL grade two category for radiographs showing JSN without osteophytes. None of their participants actually fit this new category, highlighting the controversy over presentations of knee OA as seen on a radiograph.

A second important issue is that there are multiple descriptions of the KL grades which create variability in interpreting the grades. Kellgren and Lawrence themselves wrote three different descriptors, and other researchers have used additional variations. This can cause a different cohort of research participants to be identified as having, or not having, OA, and creates difficulty in comparing research studies on knee OA.

Several authors have assessed the intra- and inter-rater reliability of the KL scale. Intra-rater reliability (Cohen’s weighted kappa 0.50 to 0.88; Cohen’s kappa 0.84 to 0.99; Spearman’s correlation coefficient 0.89; ICC 0.85 to 0.93) and inter-rater reliability (Cohen’s weighted kappa 0.56 to 0.80; Cohen’s kappa 0.59 to 0.76; Spearman’s correlation coefficient 0.85; ICC 0.68 to 0.84) generally fall in the moderate to excellent range. A lack of sensitivity to change using the KL scale has been reported, and although it was not created to follow change in OA severity over time, it is frequently used for this purpose. There are only five grades, and the scale is not linear, hence the difference between grades.
one and two is not necessarily the same as the difference between grades three and four. Differentiating between grades zero and one, and one and two can be especially difficult. The border between “possible osteophytic lipping (grade one)” and “definite osteophytes (grade two)” is very subjective and the “narrowing of joint” in the grade three description can include joints with almost no JSN to joints with almost no joint space left. In order to increase its sensitivity to change, Felson et al. proposed two changes to the KL scale: grade two to include the requirement of both osteophytes and JSN, and a new grade, two/osteophyte, which describes a knee with osteophytes but no JSN. They do admit that further changes, while addressing some of the problems, might also further the confusion created because of different definitions of the scale.

KL grades have been compared to OA changes seen on MRIs of the knee as well as by arthroscopy. MRIs are assessed using ordinal or continuous scales to measure many individual features of OA, some of which may also be observed on radiographs. KL grades are moderately to poorly correlated with cartilage lesions (Spearman’s correlation $r = 0.55$, $p < 0.01$) and cartilage volume (Pearson’s correlation $r = -0.30$ to $-0.49$ depending on location, $p < 0.01$) as measured from MRI. Correlations of KL grade to cartilage damage seen at arthroscopy are similar to those measured from MRI (Pearson’s correlation $r = 0.49$, CI 0.38 to 0.59), with a higher association for the medial compartment. These results suggest that the KL scale, with its emphasis on osteophytes, has some limitations for the grading of knee OA severity.
2.3.1.2.2 Other Global Scales

Global scales other than the KL scale tend to focus on one feature of knee OA, with other features considered secondary. Ahlback published descriptions of six stages of knee OA based on the combination of JSN and bone attrition only. Stages zero to two describe JSN only, with progressive bone attrition described in stages three to five. Ahlback and Rydberg described the stages in a further publication with altered wording. Thirty five years after the initial description, two studies showed that intra-rater (Cohen’s weighted kappa 0.17 to 0.35; Cohen’s kappa 0.15 to 0.76) and inter-rater reliability (Cohen’s weighted kappa 0.18 to 0.45; Cohen’s kappa -0.01 to 0.21) of the Ahlback scale were variable but tended to be poor. Dieppe et al. subsequently improved the reliability by using a template showing typical bone contour, to be laid over a knee radiograph.

Sundaram et al. created a seven-point radiographic scale to assess the entire TF joint for knee OA after tibial dome osteotomy. Their grading system was very similar to KL in that osteophytes were considered the initial presentation of the disease, with JSN being identified at grade three. Psychometric testing was not performed on this scale and it does not seem to have been used since.

Finally, Brandt et al. created a JSN-weighted scale that they contrasted to the KL scale (with its osteophytic weighting). Secondary features included subchondral sclerosis, geodes and osteophytes. Brandt scale scores were compared to cartilage damage seen at arthroscopy; the Pearson’s correlation coefficient was $r = 0.56$ (CI 0.46 to 0.65). As this result was no better than that for the osteophyte-weighted KL scale ($r = 0.49$, CI 0.38 to 0.59), the authors questioned whether radiography was adequate to assess early OA. This scale has been used to classify
research participants for orthopaedic surgical outcomes research but has not appeared to be used recently.

2.3.1.3 Composite Scales

2.3.1.3.1 Currently-Used Scales

Composite scales score several features of OA individually, then add them to create a total score. Felson et al. studied several radiographic features of OA and found that a combination of one or two features [osteophyte grade two or more (on a zero-to-three scale), or JSN grade two or more (on a zero-to-three scale) and a bony feature such as a cyst, sclerosis or small osteophyte], each scored individually, correlated best with clinical symptoms of pain and crepitus, lending support to the usefulness of composite scales. Altman et al. also discovered that a sum of the individual scores for JSN, bone spurs, sclerosis, attrition and alignment was more sensitive to change over time than each individual score. Unlike global scales, composite scales are able to follow the course of individual OA features, and can respond to change in individuals with a variety of knee OA presentations.

Two scales were designed to follow the development of knee OA in individuals with anterior cruciate ligament tears. Satku et al.’s scale grades osteophytes, peaking of the tibial spine, JSN and subchondral sclerosis or cysts in several locations in the knee, each on a scale of zero to one or two, to give a total score of 14. Kannus et al. created a complicated scale that measured osteophytes, subchondral sclerosis, flattening of the femoral condyles, subchondral cysts, ligament calcification, JSN and angular deformity at a variety of locations within the knee. Individual scores were out of three to 12, for a total score of 100. Lower scores denoted more severe disease. This scale was very involved and cumbersome, which would
limit its use to research. It was reported to have good to excellent intra-rater reliability (Cohen’s kappa 0.70) and inter-rater reliability (Pearson’s correlation 0.94; Spearman’s correlation 0.90).

McAlindon et al. 140 created a scale to investigate the association between knee pain, disability, knee strength and radiographic score. They scored JSN, osteophytes and sclerosis in several compartments of both knees to sum to a possible score of 30. Intra-rater reliability was moderate (Cohen’s kappa of 0.57). Another scale was created by Merchant et al. 150 to follow individuals after ankle or lower leg injuries to investigate the onset of knee OA changes. A “normal” joint was given a score of ten and points were subtracted for osteophytes, JSN, degenerative cysts and subchondral sclerosis observed in both TF compartments. Psychometric testing was not reported.

2.3.1.3.2 The Unicompartmental Osteoarthritis Grade

The unicompartmental osteoarthritis grade (UCOAG) was created in 1999 by Cooke et al. 82, who wished to create a scale that had high reliability and sensitivity to change, and was correlated with changes in alignment and deformity caused by OA. The UCOAG scores femoral osteophytes (scored out of three), JSN (scored out of three), tibial erosion (scored out of four) and subluxation (scored out of three) for a total possible score of 13. Only the most-affected TF compartment is scored, because OA is most often a focal disease. Patellofemoral OA is not assessed because only frontal radiographs are used. While femoral osteophytes are included, tibial osteophytes are excluded in order to prevent over-weighting the scale with osteophytes and because tibial osteophytes frequently decrease in size as OA worsens and the knee subluxes. Tibial erosion is included because it is common and may contribute to joint instability as it
progresses. Similarly subluxation, a feature unique to the UCOAG, is incorporated because it also contributes to joint instability and disability. The UCOAG is modestly correlated to frontal-plane alignment (Pearson’s correlation $r = 0.51$, $p < 0.001$). Sclerosis is not included because bone density is highly variable between people and is affected by obesity and variations in image quality. Equal weight is given to osteophytes, JSN and subluxation, and slightly more weight to tibial erosion. This approach was intended to reduce the emphasis of one feature (i.e. osteophytes) over another and provide for a balanced opportunity for sensitivity to change in those with different presentations of OA.

While the UCOAG was introduced in 1999, its psychometric properties have not yet been fully investigated. Initial results found an inter-rater reliability (Cohen’s weighted kappa) of 0.92 using anteroposterior full-extension radiographs. The UCOAG has been used in at least one published research study and to date there has been no alteration in the feature descriptions.

2.3.1.4 Individual Osteoarthritis Feature Scales and Measurements

Apart from the KL scale, the most common method to assess knee OA severity is to assign grades (usually ordinal) to individual features of OA such as osteophytes, JSN and sclerosis. An atlas is used to guide interpretation of each feature. Joint space may also be assessed quantitatively as joint space width (JSW), for example, in mm. While each individual feature only describes one aspect of OA, the benefit of individual scales is their ability to monitor change over time. The most-often used individual OA feature scale was created by Altman et al. as described below.
2.3.1.4.1 Osteoarthritis Research Society International Atlas

The second most-used grading system (after KL), the Osteoarthritis Research Society International (OARSI) atlas, was created by Altman et al. 74 (the San Francisco Conference Group) in 1987. For the knee, five OA features were assessed [JSN, spur formation, loss of bone stock (attrition), subchondral bony sclerosis and frontal-plane alignment], each scored from zero to three. Medial and lateral TF compartments were assessed separately (except for alignment), giving nine individual scores. A total score was not calculated. Initial intra-rater reliability scores (measured with ICCs) for each feature varied from 0.40 to 1.0, although it is important to note that only three radiographs were used for this analysis 74. Inter-rater reliability scores (measured with ICCs) were slightly lower, varying between 0.32 and 0.86, with JSN having the best reliability 74. In all cases medial compartment scores were more reliable than lateral compartment scores. Altman et al. 74 did not speculate on the reasons for this. JSN and bone spurs were the individual OA features found to be most sensitive to change over time 74.

In order to standardize the interpretation of radiographs, OARSI published another radiographic atlas in 1995 showing the spectrum of severity of three osteoarthritic features (JSN, marginal osteophytes and subchondral sclerosis), each scored from zero to three 80. These images were no longer available to be republished so an updated atlas, available electronically, was published in 2007, emphasising OA changes of medial and lateral femoral and tibial plateau osteophytes, medial and lateral JSN, medial tibial attrition, medial tibial sclerosis and lateral femoral sclerosis 200. A modified version of the OARSI JSN scale was also created, whereby if JSN had increased over time, but not enough to warrant the next grade on the zero to three scale, a one-half grade was assigned 84. This modification enhanced sensitivity to change 84.
Grades assessed using the OARSI atlas have moderate to good reliability, with JSN more reliable than osteophytes. Intra-rater reliability (Cohen’s kappa 0.57 to 0.91 for osteophytes, 0.77 to 0.83 for sclerosis and 0.68 to 0.80 or ICC 0.79 to 0.95 for JSN) is somewhat higher than inter-rater reliability (Cohen’s kappa 0.33 to 0.88 for osteophytes, 0.77 for sclerosis, and 0.48 to 0.70 or ICC 0.66 to 0.87 for JSN).

Comparison of the OARSI atlas to findings from arthroscopy has been performed. Osteophytes show moderate sensitivity (49% to 67%) compared to arthroscopy however the other OA features show fair to poor sensitivity (3% to 46%). Specificity of all features is good to excellent (73% to 100%) relative to arthroscopic findings.

2.3.1.4.2 Other Individual Osteoarthritis Feature Scales

Thomas et al. and Cooper et al. created ordinal scales for individual features of knee OA, similar to the OARSI scale. Thomas et al. scored osteophytes, JSN, sclerosis and cysts, each on a scale of zero to three. Cooper et al. scored these same four features, plus abnormality of the bony contour, each on a scale of zero to two. However, neither scale has been used extensively. More extensive use was made of an atlas produced by Spector et al. which scored TF osteophytes, sclerosis, JSN and cortical collapse, each on a scale of zero to one or three. It was updated two years later to include the skyline view to assess the patellofemoral joint. Intra-rater reliability (Cohen’s kappa 0.41 to 0.96) and inter-rater reliability (Cohen’s kappa 0.30 to 0.90) for osteophytes and JSN scored according to the Spector et al. scale ranged from fair to excellent.

Scott et al. published an atlas similar to the OARSI atlas which scores eight individual features of knee OA (medial and lateral osteophytes, medial and lateral JSN, medial and lateral
subchondral sclerosis, osteophytes of the tibial spines and chondrocalcinosis) each on a scale from zero to one or three. Both medial and lateral TF compartments were included. This atlas was created for the Baltimore Longitudinal Study of Aging and is now referred to as the Scott Feature Based Scoring System\textsuperscript{208}. It has been used in epidemiological studies and as an outcome measure\textsuperscript{209-211}. Intra-rater reliability (ICC 0.80 to 0.89) and inter-rater reliability (ICC 0.40 to 0.87) have been tested for osteophytes, JSN and sclerosis scored with this system and ranged from fair to excellent\textsuperscript{157,184}.

The Nottingham Logically Derived Line Drawing Atlas (LDLDA) consisted of line drawings (simple black drawings which illustrate the OA features of the knee), rather than photographs of radiographs\textsuperscript{154}. JSN and osteophytes were scored on a scale of zero to three. The authors felt that line drawings could overcome some issues with the OARSI atlas\textsuperscript{74}, such as differences in magnification between radiographs and more than one OA feature shown on a particular radiograph. Grades for lateral TF compartment JSN and lateral tibial osteophytes and medial femoral osteophytes assigned using the LDLDA were significantly different (higher or lower) than grades assigned using the OARSI atlas (p < 0.05)\textsuperscript{154}. Grades assigned using the LDLDA have been used to describe the participant sample in epidemiological studies\textsuperscript{212}, and as outcome measures\textsuperscript{213}. Also tested were variations of the scoring system described in the LDLDA, using grading scores from minus one to three, four and five\textsuperscript{214}, and from minus three to three, minus four to four, and minus five to five\textsuperscript{215}. The authors expected that sensitivity to change might be enhanced with some of these variations, but did not actually test this hypothesis\textsuperscript{214,215}. Finally one of the modified scales was tested using an acetate overlay placed directly on the radiograph, to aid in determining the grades\textsuperscript{216}. Reliability for each of these modified scales was as good as or better than the original scale\textsuperscript{214,216}.
Two scales use computer software to quantitatively assess knee radiographs for OA changes. The Knee Images Digital Analysis (KIDA) was an interactive software tool created for the Cohort Hip and Cohort Knee (CHECK) study. Joint space width, osteophyte area, subchondral bone density, joint angle and tibial eminence height were measured using continuous scales. While intra- and inter-rater reliability were excellent, only good-quality radiographs could be fully analysed by the software, and careful participant positioning was particularly important.

Knee OA Computer-Aided Diagnosis (KOACAD) was a fully automated diagnostic system that measured joint space area, minimum JSW, osteophyte area and TF angle on continuous scales. It was created for the Research On OA Against Disability (ROAD) study. The intra-rater reliability (ICC) for all parameters was 1.0. Sensitivity to change has not been investigated, but the authors claimed that quantitative radiograph analysis can be as sensitive as quantitative MRI. Normative and threshold parameters were acquired for males and females of different age groups.

2.3.1.4.3 Continuous Measurements of Joint Space Width

While JSN is graded on an ordinal scale as an individual feature of OA, ratio-scale measurements of JSW are frequently used to monitor the progression of OA over time, and to monitor the effects of treatment. JSW can be measured in four ways. The minimum JSW is measured as the narrowest distance between the femur and tibia. JSW can be measured between a specified location on the distal femur and a specified location on the proximal tibia. Mean JSW can be calculated within an area of interest. Finally, the area of joint space within pre-determined medial and lateral boundaries can be
calculated\textsuperscript{144, 155}. There is debate over which method is most responsive to change, although minimum JSW is most often used\textsuperscript{155, 168}. Of these methods, only minimum JSW can be performed manually using a ruler or calipers or a magnifying glass with an internal measurement scale\textsuperscript{144}. For the other measurements, semi-automated and automated computer methods must be used. These methods use computer algorithms to detect the joint margins and perform the joint space calculations\textsuperscript{153, 164, 223, 224}. Because of the limited human input they show promise for high levels of reliability\textsuperscript{144, 153, 164, 223, 225-229}. For example, Pearson’s correlations of $r = 0.98$ to $r = 0.99$ were computed for intra- and inter-rater reliability for a completely automated measurement tool created by Dacre and colleagues\textsuperscript{228, 229}. Many of the computer programs are semi-automated and allow for the reader to make adjustments to the computer-detected landmarks as needed\textsuperscript{153, 164, 223, 224}.

The joint space seen on a radiograph is an accepted proxy for the thickness of articular cartilage and narrowing of the joint space suggests degradation of the cartilage\textsuperscript{75}. It has been suggested that 36 months is the shortest period over which change in the JSW would be expected to be seen on a radiograph\textsuperscript{7}; however properly performed standardized radiographic technique and computerized quantitative measurements may shorten this period to as little as 18 months\textsuperscript{221}. Similarly, in a study of individuals without knee OA, the minimal relevant change in JSW was 0.59 mm if standardized guidelines and fluoroscopy were used, but 1.29 mm if neither was used\textsuperscript{169}. Variability exists, but the rate of joint space narrowing in individuals with OA is estimated to be between 0.13 and 0.25 mm a year\textsuperscript{230-232}.

Since JSW is used as a proxy for cartilage loss, it has been compared to changes in cartilage observed at arthroscopy and with MRI. JSW is specific (95% to 100%) but less sensitive (7% to 46%) to detect articular cartilage damage as seen at arthroscopy\textsuperscript{203}. Moderate to
good correlations (Spearman’s correlation $r = 0.58$ and Pearson’s correlation $r = 0.86$) have been reported between JSW and articular cartilage damage seen on MRI. Poor sensitivity and correlation results might be because at least some of the change in JSW is related to meniscal changes such as thinning and subluxation, depending on the radiographic protocol used.

In summary, the assessment of knee OA severity as seen on a radiograph is important for diagnosis and monitoring of disease progression. Global, ordinal and individual feature scales, and the quantitative measurement of JSW are used to assess knee radiographs. The KL, UCOAG and OARSI JSN grading scales will be assessed further in Chapters 5 and 6.

### 2.3.2 Severity Measurements from Magnetic Resonance Images

While radiography has a long history of use for the assessment of knee OA, MRI is a newer assessment modality, with benefits and drawbacks. It has been used in recent large-scale epidemiological and clinical trial studies, however it has not yet been embraced for clinical practice. This may be related to lack of availability, higher cost, greater technician training, long examination time and longer reading times compared to radiographs. Some argue that MRI in the clinical setting is unnecessary if it will not change treatment decisions.

The biggest advantage of MRI is its ability to visualize, in three dimensions, all of the tissues present in the knee, including bone, cartilage, ligaments, synovium and meniscus. In particular, cartilage is observed directly, instead of indirectly using a measure of JSN or JSW on a radiograph. Also, unlike radiographs, there are no issues with magnification, distortion and superimposition. MRI can detect pre-morphologic changes, possibly allowing for earlier diagnosis.
Intra-rater reliability (pooled ICC 0.77 to 0.94) and inter-rater reliability (pooled ICC 0.80 to 0.93) calculations for knee tissues including the cartilage, synovium, meniscus, ligament and bone, evaluated with MRI, were good to excellent²³⁹. Sensitivity to change was measured with the standardized response mean (SRM), which is the ratio of the mean change score divided by the standard deviation of the change scores²⁴⁰. It is used as an estimate of change in a particular measure, standardized to the variability between participants²⁴⁰. SRM values of 0.20 or less represent a trivial response to change, SRM values from 0.20 to 0.50 are small, SRM values from 0.50 to 0.80 are moderate and SRM values greater than 0.80 represent a large response²⁴⁰. The sensitivity to change of MRI measures of various tissues in knees with OA showed an SRM of -0.05 to -3.27, with an SRM of -0.86 for medial TF quantitative cartilage morphology, which is the most commonly assessed tissue for knee OA on an MRI²³⁹. Compared to arthroscopy, considered the criterion standard, the sensitivity of MRI to detect abnormalities in articular cartilage varied between 26% and 96%²⁴¹. Specificity varied between 50% and 100%, and accuracy varied between 49% and 94%²⁴¹. There are many MRI protocols and reading methods and some are more sensitive, specific and accurate than others.

2.3.2.1 Ordinal Scales

2.3.2.1.1 Whole-Organ Magnetic Resonance Imaging Score

The first ordinal scale score produced to assess knee OA on MRIs was the “whole-organ magnetic resonance imaging score” (WORMS) score¹⁹¹. The WORMS score includes five joint articular features¹⁹¹. Each feature is scored in several sub-regions of the knee, including the anterior, central and posterior regions of the medial and lateral tibia and femur, and the medial and lateral aspects of the patella. The articular features are: cartilage morphology (each sub-
region scored out of six), osteophytes (two additional sub-regions, superior and inferior tips of the patella; each sub-region scored out of seven), bone attrition (each sub-region scored out of three), bone marrow lesions (additional sub-region of tibial inter-spinous region; each sub-region scored out of three), and subchondral cysts (includes tibial inter-spinous region; each sub-region scored out of three). Several non-articular features are also scored: meniscal tears (six sub-regions, each scored out of four), joint effusion (scored out of three), meniscal extrusion (medial and lateral, each scored out of two), synovitis (intercondylar and infrapatellar, each scored out of three), collateral ligaments (medial and lateral, each scored out of two), cruciate ligaments (anterior and posterior, each scored zero or one), meniscal cysts (medial and lateral, each scored zero or one), popliteal cyst (scored out of three), anserine bursitis (scored zero or one), patellar bursitis (scored zero or one), TF cyst (scored zero or one) and loose bodies (scored zero or one). The medial and lateral anterior femoral compartments are considered to be part of the patellofemoral joint, rather than the TF joint. While the components of the WORMS score sum to a maximum of 380, most individuals with knee OA have relatively low WORMS scores. In one sample of 19 individuals with KL grades of two and three, the average WORMS was 60 (standard deviation 33) \(^{191}\). In another study, 70 individuals with knee OA had an average WORMS score of 64.5 (standard deviation 16.5) \(^{242}\). WORMS scoring takes approximately 80 minutes per MRI \(^{243}\). A combination of WORMS cartilage, osteophytes and synovitis scores is moderately correlated with KL score \((r = 0.51, p < 0.01)\) \(^{242}\).

The intra-rater reliability of WORMS was very high, for example, ICCs of 0.95 for meniscal pathology, 0.96 for cartilage morphology and 0.98 for bone marrow edema \(^{244}\). Inter-rater reliability results (ICC) varied from 0.61 for bone attrition to 0.99 for cartilage morphology, suggesting that these components of the WORMS are a reliable criterion standard \(^{191, 239, 245, 246}\).
WORMS is also particularly sensitive. Two studies found that up to 90% of middle-aged and elderly individuals without knee pain have abnormalities on knee MRI assessed with WORMS.

Semi-quantitative scores like WORMS were initially designed for OA assessment at one point in time, but are now being used for longitudinal studies. SRMs over 24 weeks for cartilage morphology was -0.18 to -0.50 (depending on knee compartment) and for marginal osteophytes was -0.27 to -0.40. These SRMs are small, possibly because of the short duration between baseline and follow-up.

2.3.2.1.2 Other Semi-Quantitative Scales

Since WORMS was created, another semi-quantitative scale, the Boston Leeds Osteoarthritis Knee Score (BLOKS) was created. It has been compared to WORMS, and while neither scale is consistently better than the other, they each have specific strengths and weaknesses. In an attempt to harvest the strengths from both WORMS and BLOKS, Hunter et al. created the MRI Osteoarthritis Knee Score (MOAKS). It has very good to excellent intra- and inter-rater reliability (Cohen’s kappa intra-rater reliability results from 0.61 for bone marrow lesions to 1.00 for meniscus morphology and inter-rater reliability results from 0.36 for cartilage depth to 0.97 for meniscus morphology).

2.3.2.2 Continuous Measurements

Cartilage volume, surface area and thickness, as well as effusion volume, bone marrow lesion volume and synovium can be directly and quantitatively measured on an MRI. Quantitative analyses are generally automatic or semi-automatic and depend less on reader
experience and expertise than semi-quantitative analyses. Quantitative measurements are suited to monitoring the progression of OA, especially in individuals with existing disease. SRMs for change in measurements of various tissues over one year are trivial to moderate. Some expect that quantitative MRI analyses will replace JSW on radiographs as the preferred outcome measure for trials of disease-modifying OA drugs.

2.4 Concluding Remarks

This review of the literature focused on two evaluations related to TF OA, the assessment of TF frontal-plane alignment and the measurement of TF OA severity on a radiograph. It has presented information on the current state of knowledge in these two areas. Research presented in the next two chapters of this thesis will examine the relationship between the FS-TS angle and the HKA angle as measured on a radiograph and the evaluation of a measure of the HKA angle from frontal-plane photographs. HKA angle measurements acquired without the use of ionizing radiation will be of use to physiotherapists and other clinicians for screening and monitoring patients.

The subsequent two chapters will investigate the psychometric properties of the UCOAG grading scale. The validity and sensitivity to change of the KL and OARSI JSN scales will also be investigated. While these two scales are most-studied, they are not ideal. The UCOAG grading scale shows promise to be a reliable and valid scale which is sensitive to change, however it has not yet been thoroughly assessed. If this composite scale has good reliability, validity and sensitivity to change, it will be a good choice to quantify the severity of OA on a radiograph for research and clinical applications.
Chapter 3


Published as:

3.1 Abstract

Objectives: Researchers commonly use the femoral shaft-tibial shaft (FS-TS) angle from knee radiographs to estimate the hip-knee-ankle (HKA) angle in studies examining risk factors for knee osteoarthritis (OA) incidence and progression. The objective of this study was to determine the relationship between the HKA and FS-TS angles, depending on the method of calculating the FS-TS angle and the direction and degree of knee deformity.

Methods: One hundred and twenty full-length digital radiographs were assigned, with 30 in each of four alignment groups (0.0° to 4.9°, and ≥ 5.0° of varus and valgus), from a large cohort of persons with and at risk of knee osteoarthritis. The HKA angle and 5 measures of the FS-TS...
angle (using progressively shorter shaft lengths) were obtained using Surveyor™ Analysis Software, OAISYS Inc. The offsets between the HKA angle and the different versions of the FS-TS angle were calculated, with 95% confidence intervals (CI). Pearson correlations were calculated.

**Results:** In varus limbs use of a shorter shaft length increased the offset between the HKA and FS-TS angles from 5.1° to 7.0°. The opposite occurred with valgus limbs (from 5.0° to 3.7°). Correlations between the HKA and FS-TS angles for the whole sample of 120 individuals were excellent (r = 1.00 to 0.88). However, correlations for individual alignment groups were low to moderate, especially for the shortest-shaft FS-TS angle (r = 0.41 to 0.66).

**Conclusions:** The offsets obtained using the shorter FS-TS angle measurements vary depending on direction and degree of knee deformity, and therefore may not provide reliable predictions for the HKA angle. We recommend that full-length radiographs be used whenever an accurate estimation of the HKA angle is required, although broad categories of alignment can be estimated with the FS-TS angle.

### 3.2 Introduction

Symptomatic knee osteoarthritis (OA) with radiographic changes was estimated to affect between 6.7% and 16.7% of individuals over 45 years old in a 2005 review of studies performed in the United States. This rate is increasing, primarily due to demographic factors such as aging of the population, increasing rates of obesity and an increasing prevalence of traumatic osteoarthritis. Varus or valgus alignment of the lower limb has been shown to increase the risk of progression of knee OA. More specifically, the odds ratio (OR) of OA
progression in the medial tibiofemoral compartment for those with varus deformity has been calculated to be between 2.90 and 10.96 \cite{84, 258, 259, 261}. For progression of lateral compartment OA in individuals with valgus deformity the OR ranges from 1.39 to 10.44 \cite{84, 258, 259, 261}.

The hip-knee-ankle (HKA) angle is a measure of lower limb alignment, defined as the angle between the mechanical axes of the femur and the tibia (Figure 3-1). The HKA angle is measured from a full-length lower-limb radiograph. In healthy adults with a neutral alignment, the HKA angle is between 1.0° and 1.5° of varus \cite{263, 264}. The femoral shaft-tibial shaft (FS-TS) angle (also known as the anatomic angle) is the angle between the anatomic axes of the femur and the tibia (Figure 3-1).

Some researchers advocate the use of the FS-TS angle taken from radiographs of the knee to estimate the HKA angle, with or without an offset, which is the difference between the HKA and FS-TS angles \cite{265-267}. They argue that there is a high correlation ($r = 0.65$ to 0.88) between the HKA and FS-TS angles, and that there are several advantages of a knee radiograph over a full-length one. However, others argue that in order to obtain the best estimate of mechanical alignment, the HKA angles must be directly measured from full-limb radiographs, because using a knee radiograph limits the accuracy of the measurement \cite{268, 269}. Deformities of shafts of the long bones might alter the relationship between the HKA and FS-TS angles, as may subluxation at the knee \cite{63, 269, 270}.

One factor which might influence the ability of the FS-TS angle to accurately estimate the HKA angle is the method used to calculate the FS-TS angle. Statistically significant differences in FS-TS angle measurements have been found depending on how the anatomic axes were measured \cite{264, 271}. The FS-TS angle is commonly measured on knee radiographs using lines drawn from the knee to a point 10 centimetres (cm) along the shafts of the long bones \cite{265, 267, 268}.
Figure 3-1: Diagram of a full-length lower limb radiograph with a varus alignment.

Mechanical and anatomic axes as well as the various angles are represented. The points marked on the radiograph in order to calculate the hip-knee-ankle (HKA) angle and the various femoral shaft-tibial shaft (FS-TS) angles are numbered 1 to 13.
1 – centre of head of femur
2 – femoral intertrochanteric point
3 – ⅓ femoral shaft point
4 – ½ femoral shaft point
5 – ⅓ femoral shaft point
6 – 10 cm femoral shaft point
7 – femoral intercondylar point
8 – tibial interspinous point
9 – 10 cm tibial shaft point
10 – ⅓ tibial shaft point
11 – ½ tibial shaft point
12 – ⅔ tibial shaft point
13 – tibial mid-plafond point.

FS – femoral shaft (femoral anatomic axis)
FM – femoral mechanical axis
TS – tibial shaft (tibial anatomic axis)
TM – tibial mechanical axis
HKA – hip-knee-ankle angle
FS-TS – femoral shaft-tibial shaft angle

Modified from Cooke et al. 61, with permission.
However, the use of other locations for the shaft points might change the relationship of the FS-TS angle to the HKA angle. Therefore, we wished to compare several different versions of the FS-TS angle, using different points of origin, to estimate the HKA angle. An important consideration is that for the results to be useful the shaft points must be visible on commonly acquired radiographs.

It is also possible that the relationship between the HKA and FS-TS angles might vary with respect to the nature (varus or valgus) and severity of deformity. We were unable to find any prior studies that evaluated this question. Therefore, we wished to study this relationship in cohorts of individuals with mild and severe varus and valgus deformities.

Thus, the aim of the current study was to determine the relationship between the HKA and FS-TS angles in participants with or at high risk of knee osteoarthritis. We asked three research questions: Does the relationship between the FS-TS and HKA angles differ depending on direction and magnitude of knee deformity?, Does the shaft length used to determine the FS-TS angle affect the ability to accurately estimate the HKA angle?, and What proportions of the femoral and tibial shafts are seen on a typical knee radiograph? The results of this study will inform researchers who perform clinical and epidemiological studies about which method of measuring lower limb alignment best suits their needs.
3.3 Participants and Methods

3.3.1 Radiograph Selection

The database of full-length lower limb radiographs from the Multicenter Osteoarthritis (MOST) Study was used to select images for this study (Ancillary Study AS06-03; Analysis Plan AP09-03, see Appendix A). The MOST study was approved by institutional review boards at the University of Iowa, University of Alabama, Birmingham, University of California, San Francisco and Boston University Medical Campus and participants provided written informed consent. All of the participants in the MOST study either had knee OA or were at high risk for developing knee OA. This included individuals who were overweight or obese, those with current knee pain or a history of knee injury or surgery. Individuals were excluded if they had rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, Reiter’s syndrome, significant kidney disease, cancer, bilateral knee replacement, were unable to walk without assistance or were planning to move out of the study area in the next three years. Full-length films were obtained from 1598 participants, according to the method of Sharma et al., with both right and left limbs viewed. Participants stood with knees in full extension, with the tibial tubercles facing forwards. Various joint angles (including the HKA angle) and limb lengths had previously been determined as described by Cooke et al. The reliability of this technique has been confirmed [inter-reader reliability for the HKA angle: Intraclass correlation coefficient (ICC) = 0.995 (95% confidence interval [CI], 0.994 - 1); intra-reader reliability for the HKA angle: ICC = 0.998 (95% CI, 0.998 - 1); inter-reader reliabilities for other angles between the femur and tibia: ICCs between 0.839 and 0.993; intra-reader reliabilities for other angles between the femur and tibia: ICCs between 0.908
To avoid selecting both limbs from the same participant only right limbs were selected. Limbs that showed fractures, pins or plates and hip or knee replacements were excluded, as were those where full analysis was not possible because of poor image quality or because a portion of the limb was not visible on the image. Past traumatic injury or implant placement might have altered the relationship between the HKA and the FS-TS angles and led to additional variability between the angles. Finally, chosen images must have had a ruler to allow for scaling. Thus 1240 limb images were available for analysis. From these, 30 right limbs were randomly selected for each of four categories, based on the HKA angle; group 1: HKA angle of 5.0° varus or greater, group 2: HKA angle from 0.0° up to and including 4.9° varus, group 3: HKA angle from 0.1° up to and including 4.9° valgus, and group 4: HKA angle of 5.0° valgus or greater. Group one was chosen from 181 individuals (14.6% of the available limb images), group two from 598 individuals (48.2%), group three from 406 individuals (32.7%) and group four from 55 individuals (4.4%). We attempted to select each group so that it would contain balanced representation of the sexes. While equal numbers of radiograph images were selected for each sex for three of the four HKA angle-based alignment groups, only two male participants had valgus deformities of greater than 5°. Both were included in group four. The groups were compared with respect to demographic variables [age, weight, height, body mass index (BMI) and Kellgren-Lawrence grade (KL)] using t-tests for continuous variables and Chi-square ($\chi^2$) for ordinal variables.
3.3.2 Measurements

A custom version of Surveyor™ 2.0 software from Orthopedic Alignment & Imaging Systems Inc. (OAISYS) was used to determine the HKA angle and several variations of the FS-TS angle on the full-length radiographs (Figure 3-1). Points were placed on the images with digital “tools” (center-line, circle, ruler), using strict criteria to minimize bias. For example, the centre of the femoral head was defined as the exact centre of a circle placed as closely as possible around the edge of the femoral head and the mid-shaft points were positioned exactly half-way across the shaft, using the ruler tool. For measurements of the HKA angle, points placed at the centre of the femoral head, the femoral intercondylar notch, the tibial interspinous groove and at the centre of the tibial plafond were used. The first two points defined the femoral mechanical axis and the second two points defined the tibial mechanical axis. The angle at the intersection of the two lines was the HKA angle, with negative numbers indicating varus alignment and positive numbers indicating valgus alignment. For the full-length FS-TS angle, the points were located at the intertrochanteric point between the greater and lesser femoral trochanters in line with the femoral neck axis and at the femoral intercondylar notch (femoral anatomic axis), as well as at the tibial interspinous groove and at the centre of the tibial plafond (tibial anatomic axis). The angle between these axes defined the full-length FS-TS angle. Three additional points were located on the mid-shaft of the femur, two thirds, one half and one third of the length of the femoral shaft from the knee to the intertrochanteric point. Corresponding points were located on the mid-shaft of the tibia. Finally, points were located on the femoral and tibial shafts 10 cm from the knee points. The shaft points were used to calculate four different FS-TS angles, described as the \( \frac{2}{3} \) FS-TS, \( \frac{1}{2} \) FS-TS, \( \frac{1}{3} \) FS-TS and 10 cm FS-TS angles. To minimize bias, the
points were marked in proximal to distal order, and the resulting angles were not reviewed until after all points were marked. The images were analysed in order of acquisition rather than by group.

3.3.3 Data Analysis

Mean offset was defined as the mean HKA angle minus the mean FS-TS angle. Mean offsets and 95% CI between the HKA angle and the different methods of calculating the FS-TS angle were determined for the complete sample of 120 limbs and for each alignment group. Paired t-tests were used to determine if the (HKA – FS-TS) offset was significantly different from the (HKA – 10 cm FA-TS) offset. Pearson correlation coefficients (r) were used to compare the HKA angle and the different methods of calculating the FS-TS angle, for the complete sample and separately for each alignment group. To determine if the relationship between the FS-TS and HKA angles differed depending on direction and magnitude of knee deformity, the size of the mean offsets was examined between alignment groups and compared to that of the complete sample. CIs and correlation coefficients were used to study the ability of the various FS-TS angle measurements to accurately estimate the HKA angle. Regression equations were determined to describe the relationship between the HKA and 10 cm FS-TS angles for the complete dataset and for each alignment group.

To determine any effect of sex on the results we carried out a 2-way analysis of variance (ANOVA) to compare sex differences with group, sex and group*sex as factors, for all of the alignment angles and mean offsets. Post-hoc Tukey analyses were performed as needed. Group four was not included due to insufficient numbers of male participants.
All statistical analyses were performed using Minitab\textsuperscript{TM} statistical software (Release 15.1.30, Minitab Inc., State College, Pennsylvania). Statistical significance was set at $\alpha = 0.05$ (two-sided) for all tests.

Finally, the proportion of the femoral and tibial shafts visible on a typical knee radiograph was determined. Typical radiograph cassettes and digital radiograph systems have an exposure area that is 16.5 or 17.0 inches [419 or 432 millimetres (mm)] long. All 120 right limbs were reviewed to determine what proportion of the shaft lengths would be visible on a 419 mm long image.

### 3.4 Results

Demographic data for each of the four groups is presented in Table 3-1. KL grade was significantly associated with group ($\chi^2 = 55.8$, $p<0.0001$). Participants with greater deformity (varus and valgus) more often had osteoarthritis, based on KL grade.

Lower limb angles are presented in Table 3-2. To investigate the relationship between the HKA - FS-TS angle offset, the FS-TS angle shaft length and alignment group, mean offsets and 95% CIs were calculated and plotted (Table 3-2 and Figure 3-2). The average offset between the mechanical and anatomic axes (full-length FS-TS angle) for the entire dataset was $-5.0^\circ$ (95% CI, -5.1, -4.9). However, when the sample was broken down into alignment groups, substantial variability was evident. For limbs with a varus deformity the magnitude of the offset increased as the shaft length for the FS-TS angle calculation decreased. But for limbs with a valgus deformity,
Table 3-1: Demographic data, with mean and standard deviation, for each alignment group.

<table>
<thead>
<tr>
<th>Group</th>
<th>HKA</th>
<th>Sex % of women</th>
<th>Age mean (yrs.) SD (yrs.)</th>
<th>Weight mean (kg) SD (kg)</th>
<th>Height mean (cm) SD (cm)</th>
<th>BMI mean (kg/m²) SD (kg/m²)</th>
<th>K/L Grade % grade 2 or greater</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete Dataset</td>
<td>61%</td>
<td>63.0 8.4</td>
<td>89.3 16.0</td>
<td>168.2 9.1</td>
<td>31.6 5.8</td>
<td>51%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>≥ 5.0° Varus</td>
<td>50%</td>
<td>62.4 9.5</td>
<td>93.7 15.6</td>
<td>167.4 10.9</td>
<td>33.8 6.8</td>
<td>70%</td>
</tr>
<tr>
<td>2</td>
<td>between 0.0° – 4.9° Varus</td>
<td>50%</td>
<td>61.5 8.0</td>
<td>88.8 18.3</td>
<td>168.3 8.5</td>
<td>31.4 6.0</td>
<td>33%</td>
</tr>
<tr>
<td>3</td>
<td>between 0.0° – 4.9° Valgus</td>
<td>50%</td>
<td>63.9 9.1</td>
<td>86.8 13.1</td>
<td>172.2 9.4</td>
<td>29.2 3.3</td>
<td>20%</td>
</tr>
<tr>
<td>4</td>
<td>≥ 5.0° Valgus</td>
<td>93%</td>
<td>64.2 7.0</td>
<td>87.8 16.4</td>
<td>165.2 6.1</td>
<td>32.2 5.9</td>
<td>80%</td>
</tr>
<tr>
<td>Significant Differences (p &lt; 0.05)</td>
<td>none</td>
<td>none</td>
<td>3 &amp; 4 p = 0.025</td>
<td>1 &amp; 3 p = 0.022</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3-2: Means and 95% confidence intervals (CI) for lower limb angles and (HKA - FS-TS) offsets, divided by sex.
Group 4 was not included in the analysis of variance to compare group and sex due to insufficient male participants. * sex main effect, p < 0.05
HKA – hip-knee-ankle (angle) FS-TS – femoral shaft – tibial shaft (angle)

a) Lower limb angles

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex</th>
<th>HKA mean (°) CI (°)</th>
<th>FS-TS mean (°) CI (°)</th>
<th>⅔ FS-TS mean (°) CI (°)</th>
<th>½ FS-TS mean (°) CI (°)</th>
<th>⅓ FS-TS mean (°) CI (°)</th>
<th>10cm FS-TS mean (°) CI (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete Dataset</td>
<td>73 ♀</td>
<td>1.4 (0.0, 2.8)</td>
<td>6.4 (5.1, 7.8)</td>
<td>7.3 (6.0, 8.5)</td>
<td>6.8 (5.6, 7.9)</td>
<td>6.2 (5.1, 7.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47 ♂</td>
<td>-2.0 (-3.4, -0.5)</td>
<td>3.0 (1.6, 4.5)</td>
<td>4.2 (2.8, 5.5)</td>
<td>4.1 (2.8, 5.4)</td>
<td>4.0 (2.8, 5.3)</td>
</tr>
<tr>
<td>1</td>
<td>≥ 5.0° Varus</td>
<td>15 ♀</td>
<td>-7.0 (-7.6, -6.3)</td>
<td>-1.5 (-2.3, -0.7)</td>
<td>0.1 (-0.7, 1.0)</td>
<td>0.3 (-0.6, 1.2)</td>
<td>0.2 (-0.9, 1.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 ♂</td>
<td>-7.7 (-8.7, -6.7)</td>
<td>-3.6 (-0.5, -0.3)</td>
<td>-1.2 (-2.0, -0.3)</td>
<td>-0.9 (-2.1, 0.2)</td>
<td>-0.7 (-2.0, 0.6)</td>
</tr>
<tr>
<td>2</td>
<td>between 0.0° – 4.9° Varus</td>
<td>15 ♀</td>
<td>-2.6 (-3.4, -1.8)</td>
<td>2.3 (1.5, 3.1)</td>
<td>3.8 (2.8, 4.7)</td>
<td>3.9 (2.9, 4.8)</td>
<td>3.6 (2.3, 4.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 ♂</td>
<td>-2.0 (-2.6, -1.4)</td>
<td>2.9 (2.3, 3.5)</td>
<td>4.3 (3.6, 5.0)</td>
<td>4.5 (3.7, 5.3)</td>
<td>4.6 (3.8, 5.3)</td>
</tr>
<tr>
<td>3</td>
<td>between 0.1° – 4.9° Valgus</td>
<td>15 ♀</td>
<td>2.0 (1.2, 2.7)</td>
<td>7.1 (6.4, 7.8)</td>
<td>7.4 (6.6, 8.2)</td>
<td>6.6 (5.6, 7.5)</td>
<td>5.7 (4.7, 6.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 ♂</td>
<td>2.0 (1.4, 2.6)</td>
<td>6.9 (6.3, 7.5)</td>
<td>7.6 (6.8, 8.3)</td>
<td>7.2 (6.4, 8.0)</td>
<td>6.9 (6.1, 7.7)</td>
</tr>
<tr>
<td>4</td>
<td>≥ 5.0° Valgus</td>
<td>28 ♀</td>
<td>7.7 (7.0, 8.5)</td>
<td>12.6 (11.8, 13.4)</td>
<td>12.9 (12.0, 13.8)</td>
<td>11.9 (11.0, 12.9)</td>
<td>11.0 (10.0, 12.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ♂</td>
<td>12.0 (9.9, 14.0)</td>
<td>17.3 (15.0, 19.7)</td>
<td>17.3 (16.2, 18.3)</td>
<td>15.6 (15.0, 16.2)</td>
<td>14.4 (14.0, 14.8)</td>
</tr>
</tbody>
</table>
### b) Lower limb (HKA – FS-TS) offsets

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex</th>
<th>HKA – FS-TS mean (°) CI (°)</th>
<th>HKA – ⅓ FS-TS mean (°) CI (°)</th>
<th>HKA – ½ FS-TS mean (°) CI (°)</th>
<th>HKA – ⅓ FS-TS mean (°) CI (°)</th>
<th>HKA – 10 cm FS-TS mean (°) CI (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73 ♀</td>
<td>-5.0 (-5.2, -4.9)</td>
<td>-5.9 (-6.2, -5.6)</td>
<td>-5.4 (-5.8, -5.0)</td>
<td>-4.7 (-5.3, -4.2)</td>
<td>-4.6* (-5.2, -4.0)</td>
</tr>
<tr>
<td></td>
<td>47 ♂</td>
<td>-5.0 (-5.2, -4.8)</td>
<td>-6.1 (-6.4, -5.8)</td>
<td>-6.1 (-6.5, -5.6)</td>
<td>-6.0 (-6.6, -5.4)</td>
<td>-6.5* (-7.3, -5.8)</td>
</tr>
<tr>
<td>1</td>
<td>≥ 5.0° Varus</td>
<td>15 ♀</td>
<td>-5.4 (-5.7, -5.1)</td>
<td>-7.1 (-7.5, -6.7)</td>
<td>-7.3 (-7.8, -6.7)</td>
<td>-7.2 (-8.0, -6.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 ♂</td>
<td>-5.1 (-5.4, -4.8)</td>
<td>-6.5 (-7.1, -6.0)</td>
<td>-6.8 (-7.7, -5.9)</td>
<td>-7.0 (-8.2, -5.7)</td>
</tr>
<tr>
<td>2</td>
<td>between 0.0° – 4.9° Varus</td>
<td>15 ♀</td>
<td>-4.9 (-5.3, -4.6)</td>
<td>-6.4 (-7.0, -5.8)</td>
<td>-6.5 (-7.2, -5.7)</td>
<td>-6.2 (-7.3, -5.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 ♂</td>
<td>-4.9 (-5.2, -4.6)</td>
<td>-6.3 (-6.8, -5.8)</td>
<td>-6.5 (-7.1, -5.8)</td>
<td>-6.5 (-7.3, -5.8)</td>
</tr>
<tr>
<td>3</td>
<td>between 0.1° – 4.9° Valgus</td>
<td>15 ♀</td>
<td>-5.1 (-5.4, -4.9)</td>
<td>-5.4 (-5.9, -5.0)</td>
<td>-4.6 (-5.2, -4.0)</td>
<td>-3.8 (-4.6, -3.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 ♂</td>
<td>-4.9 (-5.3, -4.5)</td>
<td>-5.6 (-6.3, -5.0)</td>
<td>-5.3 (-6.0, -4.5)</td>
<td>-4.9 (-5.7, -4.1)</td>
</tr>
<tr>
<td>4</td>
<td>≥ 5.0° Valgus</td>
<td>28 ♀</td>
<td>-4.9 (-5.0, -4.7)</td>
<td>-5.2 (-5.6, -4.8)</td>
<td>-4.2 (-4.8, -3.7)</td>
<td>-3.2 (-4.0, -2.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ♂</td>
<td>-5.4 (-5.6, -5.1)</td>
<td>-5.3 (-6.3, -4.3)</td>
<td>-3.7 (-5.1, -2.2)</td>
<td>-2.5 (-4.9, 0.0)</td>
</tr>
</tbody>
</table>
**Figure 3-2:** Mean offsets (with 95% confidence intervals) between the hip-knee-ankle (HKA) angle and the different methods of determining the femoral shaft-tibial shaft (FS-TS) angles, for each alignment group.

* p < 0.05 for comparison to HKA – FS-TS offset
the magnitude of this offset decreased. Similarly, the data for the individual alignment groups revealed much weaker correlations. There was also a significant difference between the HKA – FS-TS angle and HKA – 10 cm FS-TS angle offsets for both varus alignment groups; however there was only a significant difference between the HKA – FS-TS angle and the HKA – 10 cm FS-TS angle offsets for the more severe valgus group.

The linear regression equation to describe the relationship between the HKA and 10 cm FS-TS angles for the complete dataset was: HKA angle = -5.94 + 1.111 * 10 cm FS-TS (p = 0.000). Linear regression equations for each alignment group were as follows: group 1 (severe varus) HKA angle = -7.34 + 0.266 * 10 cm FS-TS angle (p = 0.026); group 2 (mild varus) HKA angle = -3.47 + 0.267 * 10 cm FSTS angle (p = 0.013); group 3 (mild valgus) HKA angle = 0.039 + 0.298 * 10 cm FSTS angle (p = 0.004); and group 4 (severe valgus) HKA angle = 2.67 + 0.491 * 10 cm FSTS angle (p = 0.000).

For groups one to three there were no group*sex interactions or sex main effects for the angles and offsets, with two exceptions. Sex had a main effect for the 10 cm FS-TS angle and the HKA – 10 cm FS-TS angle offset (Table 3-2). However, we did not find significant offset sex differences in the offsets among alignment groups.

The FS-TS angle shaft length appears to influence the ability to estimate the HKA angle. When we examined the correlation between the HKA and FS-TS angles, we found that correlations for the entire sample were high (r > 0.88) (Table 3-3). However the correlations were much weaker for shorter-shaft FS-TS angle measurements. Despite the sex main effect for the HKA – 10 cm FS-TS angle offsets, the correlations for the entire sample divided into males (r = 0.87) and females (r = 0.89) were very similar.
Table 3-3: Pearson correlations (r) between the hip-knee-ankle (HKA) angle and the different methods of measuring the femoral shaft-tibial shaft (FS-TS) angle.

<table>
<thead>
<tr>
<th>Group</th>
<th>HKA vs. FS-TS r</th>
<th>HKA vs. ½ FS-TS r</th>
<th>HKA vs. ⅓ FS-TS r</th>
<th>HKA vs. 10 cm FS-TS r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete Dataset</td>
<td>1.0</td>
<td>0.98</td>
<td>0.92</td>
</tr>
<tr>
<td>1</td>
<td>≥ 5.0° Varus</td>
<td>0.95</td>
<td>0.85</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>between 0.0° – 4.9° Varus</td>
<td>0.90</td>
<td>0.74</td>
<td>0.62</td>
</tr>
<tr>
<td>3</td>
<td>between 0.1° – 4.9° Valgus</td>
<td>0.87</td>
<td>0.73</td>
<td>0.60</td>
</tr>
<tr>
<td>4</td>
<td>≥ 5.0° Valgus</td>
<td>0.98</td>
<td>0.91</td>
<td>0.82</td>
</tr>
</tbody>
</table>

p < 0.05 in each case.
Finally, we investigated how much of the femoral and tibial shafts were visible on a typical knee radiograph. Presuming that the knee was centered perfectly on the image, a 419 mm long radiograph image showed approximately 208 mm above and 208 mm below the joint line. One-third of the tibial and femoral shafts were seen on all images, as were the 10 cm points. One-half of the femoral shaft was seen on the shortest limbs (23% of the 120 limbs in the sample) and one-half of the tibial shaft was seen on most of the limbs (92% of the limbs). The two-thirds femoral and two-thirds tibial shaft points were not seen on limbs of any length.

3.5 Discussion

Several studies have investigated the relationship between the HKA and FS-TS angles. The current study, to our knowledge, is the first to suggest that the relationship between the HKA and FS-TS angles differs depending on the direction and degree of deformity of the lower limb. Also, we found that using shorter FS-TS shaft lengths to estimate the HKA angle weakened the relation of the anatomic angle with the mechanical angle in the overall sample. The relation of these two measures was especially attenuated when both shorter shaft lengths were used and subcategories of alignment were studied.

The demographic data of the four alignment groups did not differ significantly, with a few exceptions. Only two males in the entire MOST database had HKA angles of greater than 5° valgus. The rarity of valgus deformity in males has been noted previously and partly explains the difference in height between groups three and four. Those with greater deformities (varus or valgus) had higher KL grades and tended to have higher values for BMI.

With only two exceptions no significant differences were found between the sexes with respect to the various angles and offsets, similar to prior results from adults with and without
However, in contrast to the current study, a difference has previously been found in the HKA angle and the HKA – FS-TS angle offsets between the sexes. The offset for females has been reported to be between 3.0° and 3.5° while that for males was between 4.7° to 6.4°. Chang et al. reported the opposite trend, with females having a larger offset than males (7.3° versus 6.0°; the FS-TS angle measured with 15 cm shaft lengths), at least for individuals with knee OA. Further comparisons of males and females need to be performed to confirm if real differences exist and the direction of these differences.

Research question one asked whether the relationship between the FS-TS and HKA angles differs depending on the direction and magnitude of knee deformity. The average offset between the HKA and full-length FS-TS angles of 5.0° in the current study is similar to findings from other studies where 4° - 6° is typically considered as the difference. However, our data show that the HKA – FS-TS angle offset varied as a function of the degree of deformity, especially for the FS-TS measures made using shorter shaft lengths. Specifically, for varus limbs the offset increased and for valgus limbs, it decreased. Therefore when dealing with individuals with significant varus or valgus deformity, it would be inaccurate to use 5.0° as the difference between the HKA angle and the shorter-shaft versions of the FS-TS angle, as the FS-TS angle measurements vary widely from the HKA angle.

Research question two asked whether the shaft length used to determine the FS-TS angle affects its ability to accurately estimate the HKA angle. As the FS-TS shaft length decreased, the confidence limits around our offsets increased and the correlations for the individual alignment groups decreased from greater than r = 0.87 (for the HKA – full-length FS-TS angle offset) to less than r = 0.66 (for the HKA – 10 cm FS-TS angle offset) (Table 3-3), contributing to a poor estimate for the HKA angle. Prior studies, using 10 cm shaft lengths and the same knee points as
in the current study, have found poor to excellent correlations between the HKA and FS-TS angles \((r = 0.27\) to \(r = 0.88\) for the FS-TS angle obtained with standing radiographs\(^{63, 265, 267, 268}\) and \(r = 0.66\) to \(r = 0.75\) for the FS-TS angle obtain with fixed-flexion radiographs\(^{100, 265}\)). Some of these studies had a wide variety of participants with varus, valgus and neutral lower limb alignments \((r = 0.27\) to \(0.80)\)\(^{63, 100, 265}\) while others only used participants with medial compartment OA, which is associated with varus alignment \((r = 0.34\) to \(0.88)\)\(^{267, 268}\). When our participant sample was broken down into alignment strata the correlations become much weaker, especially for the \(\frac{1}{3}\) and 10 cm FS-TS angle calculations. One limitation of performing correlations on subgroups of a dataset is that because each group is limited to individuals within a small range of the HKA angle values, the correlations will be attenuated. However, we also found that the correlations became weaker as the FS-TS shaft lengths decreased, even for the individual alignment groups, which showed high correlations for the HKA versus FS-TS angles, but much smaller correlations for the HKA versus 10 cm FS-TS angles. Confidence limits around the HKA angles imputed were wide enough to suggest caution when using \(\frac{1}{3}\) FS-TS and 10 cm FS-TS measurements to estimate the HKA angles.

Several authors have reported results similar to those in the current study when comparing different methods of calculating the FS-TS angles, including greater variation between the mechanical axis and distal femoral anatomic axis than the full-length anatomic axis, and a higher correlation \((r = 0.65, p < 0.0001)\) between the HKA and FS-TS angles calculated using the mid-diaphyseal lines of the femur and tibia than the 10 cm FS-TS measurement \((r = 0.34, p = 0.005)\)\(^{268, 271}\). As well, the FS-TS measurements taken using a 15 cm shaft length \((r = 0.81\) for males, \(r = 0.88\) for females) had greater correlations to the HKA angles than those taken using a 10 cm shaft length \((r = 0.69\) for males, \(r = 0.80\) for females)\(^{63}\). These studies lend support to our
contention that short-shaft FS-TS angle measurements increase uncertainty if used to estimate the HKA angle.

HKA angle measurements allow the opportunity to study the contribution of various parts of the limb to alignment\textsuperscript{263, 275}. Geometric changes in the shafts of the bones may cause some of the discordance between the HKA and FS-TS angles\textsuperscript{63, 263}. These changes might predispose individuals to knee OA or may be brought on by bone remodeling that occurs with OA development\textsuperscript{63, 263}.

Research question three asked what proportions of the femoral and tibial shafts are seen on a typical knee radiograph. Much of the prior research comparing the HKA and FS-TS angles uses FS-TS angle measurements calculated using a 10 cm shaft length\textsuperscript{265, 267, 268}. The results show that one-third of the femoral and tibial shafts are visible on the average cassette, even for the tallest participants. Unfortunately, the correlations are similarly poor for the 10 cm and $\frac{1}{3}$ FS-TS angle comparisons to the HKA angle.

One limitation to this study is that the various FS-TS angle measurements were determined from full-length radiographs rather than anteroposterior knee radiographs which are commonly used in research investigating the incidence and progression of knee OA. The FS-TS angles calculated from full-length radiographs and anteroposterior knee radiographs have never been compared, however Kraus et al.\textsuperscript{265} found a good correlation ($r = 0.73$, $p < 0.0001$) between the FS-TS angle measured from semiflexed knee radiographs and the FS-TS angle measured from full-length radiographs.

This study has practical implications with respect to the measurement of lower-limb alignment for research purposes. There are significant limitations to using the FS-TS angle to predict lower-limb alignment, especially when an accurate measurement of mechanical alignment
is required and we recommend that the HKA angle be used to determine lower-limb alignment. However, for samples with a variety of varus and valgus limbs and where broad categories of alignment are required for large numbers of persons in a study, the FS-TS angle could be used with the correction factors we provide to categorize participants as varus or valgus, with the recognition that limbs close to neutral will be hard to accurately classify. For subgroup studies, such as those of medial knee OA, categorizing limbs will produce more accurate estimates (i.e. all will probably be varus), but since there is uncertainty around each of the correction factors (see confidence intervals in tables), estimation of the HKA angle from the FS-TS angle is imperfect and using the FS-TS angle to guess the exact HKA angle in individuals is problematic. The linear regression equations may also be used to estimate the HKA angle using the 10 cm FS-TS angle. Hinman et al.\textsuperscript{66} also created a similar linear regression: $\text{HKA} = 13.90 + 0.915 \times 10 \text{ cm FS-TS}$, $r = 0.88$. The offset between the HKA and FS-TS angles in their sample of 40 individuals with symptomatic knee OA was only 1.1°.

This caution also pertains to the use of lower-limb alignment to estimate joint space narrowing in the progression of knee OA. If the FS-TS angle is used to estimate the HKA angle, which in turn is used to estimate joint space narrowing, any error will be compounded. In individuals with severe valgus deformity, valgus malalignment severity would be underestimated using the FS-TS angle, and thus any joint space change would be underestimated. Conversely, for individuals with severe varus deformity, the degree of varus malalignment would be overestimated using the FS-TS angle and any joint space change would be overestimated.

In conclusion, we recommend that full-length radiographs be used whenever an accurate estimation of the HKA angle is required. This is because the offset between the HKA and short-shaft FS-TS angle measurements is variable, and is influenced by the direction and degree of
malalignment of the lower limb. Imprecision around the correction factor would make it challenging to accurately predict an individual’s mechanical angle. However, broad categories of alignment in groups of persons can be estimated using short limb films, especially if the sample includes a variety of limbs that are varus, neutral and valgus.
Chapter 4

Standardized standing pelvis-to-floor photographs for the assessment of lower extremity alignment

4.1 Abstract

**Objectives:** The first purpose of this study was to determine how to estimate the centres of the knee and ankle and a proximal femoral point on a photograph. The second purpose was to assess the intra-rater, inter-rater and test-retest reliability and validity of a pelvis-to-floor photograph for the estimation of the hip-knee-ankle (HKA) angle.

**Methods:** Sticky dots were placed on participants’ anterior superior iliac spines, the superior pubic symphysis and at the knee and ankle. One radiograph and one photograph were taken with the participant standing in a standardized position and the dots were removed. After thirty minutes the dots were positioned again and a second photograph was taken. The HKA was measured from each radiograph. The proximal thigh point and estimated centres of the knee and ankle which estimated the HKA angle most accurately were determined using Pearson’s correlation coefficients. HKA angles were measured from the photographs (HKA-P). Reliability was tested using intraclass correlation coefficients (ICC$_{(2,1)}$), Bland-Altman analyses and the minimal detectable change (MDC$_{95}$). Validity was tested using a Pearson’s correlation coefficient and Bland-Altman analysis.
Results: Fifty adults participated. Using the selected points the HKA-P angle was 5.0° more varus than the HKA angle. Intra-rater (ICC_{(2,1)} > 0.985), inter-rater (ICC_{(2,1)} = 0.988) and test-retest reliability (ICC_{(2,1)} = 0.903) were excellent with no discernible bias. The MDC_{95} was 2.69°. The HKA-P was highly correlated to the HKA (r = 0.92).

Conclusions: The HKA-P angle may be used in place of the HKA angle, for clinical, research and screening purposes.

4.2 Introduction

Knee osteoarthritis (OA) is a common cause of pain and physical disability in older adults, with a prevalence of 5.4% to 38% \(^6\)\(^{-13}\). Malalignment of the lower extremities is one risk factor for knee OA onset and progression \(^{35, 52, 53, 83-86}\). Varus alignment causes the weight-bearing axis (a line drawn from the centre of the femoral head to the centre of the ankle) to be positioned medial to the knee, resulting in increased loading in the medial tibiofemoral compartment \(^{52}\). Valgus alignment has the opposite effect, with decreasing loading in the medial tibiofemoral compartment; severe valgus deformity will also cause increased loading in the lateral tibiofemoral compartment. Increased loading is believed to contribute to the onset and progression of OA in the respective compartment \(^{52}\). Progression of existing knee OA is highly associated with varus [odds ratio (OR) 2.90 to 10.96, \(p < 0.05\)] and valgus (OR 3.42 to 10.44, \(p < 0.05\)) deformities \(^{35, 52, 53, 83-86}\). The association of knee OA onset and malalignment is smaller (varus OR 2.1, \(p < 0.05\); valgus OR 2.5, \(p < 0.05\)) \(^{35, 83}\). Conservative and surgical treatment options exist to attenuate the loads or modify varus or valgus deformity. Conservative options include orthotics and bracing while a high tibial osteotomy is a common surgical procedure
intended to reduce deformity and therefore decrease the compressive force in the affected tibiofemoral compartment\cite{35,54}.

Frontal-plane lower extremity (LE) alignment is therefore an important assessment, for its potential impact on the development and progression of knee OA and in the evaluation of the effectiveness of treatment options. The “gold standard” measure of frontal-plane LE alignment is the hip-knee-ankle angle (HKA), measured from a full-length LE radiograph\cite{61}. The HKA angle occurs at the intersection of a line drawn from the centre of the head of the femur through the centre of the knee (femoral axis) with a line drawn from the centre of the knee through the centre of the ankle (tibial axis). Varus angles are denoted in negative degrees and valgus angles are positive\cite{61}.

Full-length LE radiographs are not always ideal to determine the HKA angle; stated reasons include expense, lack of specialized equipment and concern over ionizing radiation\cite{62}. There is a need for an accurate way to estimate the HKA angle which can be performed in a clinic setting and not require the use of radiographs. The method should be fast, simple and reliable, use readily available equipment and create a permanent record for follow-up comparisons. Such a technique could be used to screen for deformity, for treatment planning, as a clinical outcome measure and for clinical research\cite{62}. Goniometry has been used in the standing position to estimate the HKA angle, with Pearson’s correlations that range from 0.32 (p = 0.12) to 0.75 (p < 0.001) and biases that range from 3.3° to approximately 8.1°\cite{62,66,68,69}. Other methods tested include measuring the distance between the knees (for individuals with varus deformity, standing with ankles touching) or between the ankles (for individuals with valgus deformity, standing with knees touching) (correlation with the HKA angle: Pearson’s r = 0.76, p < 0.001), measuring the distance from the knee or ankle to a plumb line secured in a haemostat and held between the legs.
(correlation with the HKA angle: Pearson’s $r = 0.71$, $p < 0.001$) and using an inclinometer to determine the angle of the tibia with respect to the vertical (correlation with the HKA angle: Pearson’s $r = 0.80$, $p < 0.001$) $^{66}$. While these methods may be reliable [intraclass correlation coefficient (ICC) 0.84 to 0.97] $^{62,66,69}$, the application of these techniques with the patient in standing is awkward for the clinician and no permanent visual record is maintained.

Standardized pelvis-to-floor photographs may be a viable alternative to full-length LE radiographs for the estimation of the HKA angle. They fulfill the requirements for a suitable clinical assessment tool and are less awkward to perform than the other methods listed above. Prior research using samples of 16 to 20 healthy young adults with relatively low body mass index (BMI, kg/m$^2$) suggests that the HKA estimated on a photograph (HKA-P) has moderate to very high intra-rater (ICC = 0.63 to 0.99), inter-rater (ICC = 0.83 to 0.99) and test-retest (ICC = 0.70 to 0.90) reliability and is highly correlated to the HKA angle ($r = 0.98$) with a bias of $0.9^\circ$ $^{70,71}$. The first purpose of this study was to determine the location of points that estimate the centres of the knee and ankle and a proximal femoral point on a photograph, in order to provide the best estimate of the HKA angle. The second purpose of this study was to determine the intra-rater, inter-rater and test-retest reliability and concurrent validity (correlation to the HKA angle) of the HKA-P angle in a larger number of adults compared to previous studies, with a range of ages and BMI scores which are more representative of the general population.
4.3 Participants and Methods

4.3.1 Participants

Fifty adult participants (age 18 and over) who could stand without assistance for 20 minutes with knees extended and weight born equally on both LEs were recruited from the community. Recruitment posters were placed on the University campus and throughout the wider community. Information was also placed on web sites that target seniors and retired individuals. This sample size was assumed to be large enough to represent the population distribution (i.e. the mean of the sample was representative of the mean of the population), and to enable an adequate distribution of BMI and alignment. Potential participants were not accepted if they had a recent traumatic injury to the knee or ankle, (since a large joint effusion could make the identification of landmarks difficult) or contraindications to radiography (pregnancy, cancer or other serious illness). The study was approved by the University Health Sciences Research Ethics Board. Participants gave informed consent (Appendix B).

4.3.2 Measurements

4.3.2.1 Standing Full-length Lower Extremity Radiograph

One weight-bearing, full-length LE anteroposterior digital radiograph was taken in a standardized position (see Figure 4-1). The participant, dressed in shorts and in bare feet, stood on a step-stool which had a calibrated template attached to the top (Figure 4-2). He or she stood
Figure 4-1: Participant set-up for the radiograph and first photograph.

1. Anterior superior iliac spine
2. Superior border of the pubic symphysis
3. Inferior pole of the patella
4. Mid-point between the medial and lateral joint lines
5. Point visually inspected to be at the centre of the ankle between the extensor hallucis longus and the extensor digitorum longus tendons
6. Mid-point between the medial and lateral malleoli
**Figure 4-2:** Calibrated template, used to position the participants’ feet accurately and to measure lower extremity rotation.

Modified from Orthopedic Alignment and Imaging Systems, Inc.
with heel centres 9 centimetres (cm) apart and the lower limbs rotated such that the axis of knee flexion was in the frontal plane. This was checked by having the participant repeatedly flex and extend the knee while adjusting foot rotation until an imaginary axis located medial to lateral across the knee was visually estimated to be in the frontal plane. Foot rotation was observed from the protractor on the calibrated template and recorded so that the position could be reproduced accurately. A ruler with 2.0 millimetre (mm) radiopaque tantalum beads glued at 10 cm intervals was placed beside the participant in the same frontal plane as the knees. The participant held onto horizontal bars attached to the radiograph frame at hip level, for stability and to keep the arms from obstructing the images. Nineteen millimetre sticky paper dots (Avery, Pickering, ON) with 2.0 mm tantalum beads taped to the centre were placed over the following landmarks bilaterally: anterior superior iliac spine (ASIS), superior border of the symphysis pubis (SPS), inferior pole of the patella, the mid-point between the medial and lateral joint lines of the knee, the mid-point between the medial and lateral malleoli and a point visually inspected to be at the centre of the ankle between the extensor hallucis longus and the extensor digitorum longus tendons. See Figure 4-1. The landmarks were determined by palpation and the use of calipers.

An OPTIMA XR640 x-ray machine with a 16-inch-by-16-inch [40.3 cm by 40.3 cm] digital detector (General Electric, model number 2393824) was used. Three or four individual radiographs were taken (depending on the leg length of the participant), from the pelvis to the ankles, and “stitched” together automatically using software included with the x-ray machine. The radiograph was reviewed for quality before the participant moved. No identifying information except for participant number was included on the radiographs.
4.3.2.2 Standing Pelvis-to-floor Lower Extremity Photograph

Two photographs were taken. The first one was taken immediately before or after the radiograph; the participant did not move (see Figure 4-1). For the second photograph, the participant stood in the same standardized position, but on the floor in front of a white background. The stickers, this time with black dots in the centre rather than tantalum beads, were positioned as stated above. The ruler was positioned beside the participant. Arms were held across the chest. A Canon PowerShot SD800IS (Cannon Canada Inc., Mississauga, ON) digital camera (7.1 megaPixels) attached to a tripod was used with the lens of the camera positioned at the level of the participant’s knee joint line. A 3.1x optical zoom setting was used. Parallax was minimized by locating the camera the maximum distance possible (3.0 metres) away from the participant allowed due the dimensions of the room. The photograph was checked for quality before the participant moved and a second photograph was taken if needed.

4.3.3 Procedure

To allow for the investigation of test-retest reliability, there were two testing sessions, separated by a 30-minute break which allowed any skin irritation from the stickers to resolve, so that the examiner did not know where the stickers were placed during the first session. At the first testing session the participant provided informed consent and markers were placed as described above. Next, one photograph and one radiograph were taken, in random order, while the participant remained stationary. The stickers were removed and the participant changed back into his or her street clothes.
Thirty minutes later the participant returned but to a different room because the x-ray room was no longer available. Age, sex, height and weight were recorded. The participant changed back into his or her shorts, the stickers were re-applied and the standardized position was recreated with the help of the recorded foot angle from the template. A second photograph was taken, following the same protocol as the first photograph.

A customized imaging analysis software program (Surveyor™ image analysis program 3.1, Orthopedic Alignment and Imaging Services, Inc.) was used to determine the joint centres and the HKA and HKA-P angles. Three readers were trained on the use of the software. Training took one hour one-on-one (either in-person or by Skype) and was followed by practice. A test batch of nine images was read twice and analysed for intra-rater and inter-rater reliability. A two-page instruction guide was given to all readers and reviewed regularly to ensure consistency.

The radiographs were de-identified and randomized and the right and left knees were assessed by one reader to obtain the HKA angle. The photographs taken first in the protocol (in the x-ray room) were de-identified and the right and left knees were assessed to obtain the HKA-P angle by three readers twice each, at least two weeks apart. The photographs were randomized separately for each instance of reading. Finally, the photographs taken in the second session (in the second room) were randomized and assessed once, by reader one, to obtain the HKA-P angle.

4.3.4 Determination of Frontal-Plane Alignment - Radiograph

The HKA angle for each radiograph was determined by only one experienced reader with the Surveyor™ image analysis program (see Figure 4-3). The HKA angle was calculated as the
Figure 4-3: Determination of the hip-knee-ankle (HKA) angle with a full-length lower extremity radiograph.

1. Centre of the femoral head
2. Tibial interspinous groove
3. Centre of the tibial plafond
angle between a line drawn from the centre of the femoral head to the tibial interspinous groove and a line drawn from the tibial interspinous groove to the centre of the tibial plafond. HKA angle calculations using this software have excellent intra- and inter-rater reliability (ICC > 0.995).

4.3.5 Identification of Knee and Ankle Joint Centres and a Proximal Femoral Point on a Photograph

Radiographs and photographs from the first 18 participants were used to identify the knee joint centre that would be visible on a photograph. For the ankle joint centre and the proximal femoral point only 17 radiographs were available because one radiograph or photograph did not adequately show all of the landmarks. It was deemed that assessing this number several months before the validity analyses would not bias the reader and this sample size would give a reasonable estimate of the suitability of each point. Images were assessed by one reader using the Surveyor™ image analysis program.

4.3.5.1 Knee

The potential for three different points to estimate the centre of the knee on radiographs was determined. All three points could be seen on both the radiographs and photographs. The first was the inferior pole of the patella (designated on the radiograph by a tantalum bead). This point had been observed on radiographs in the past. The second point was the mid-point between the medial and lateral joint lines (designated by a tantalum bead seen on the radiograph). The third point was the mid-point of a horizontal line drawn across the knee on the radiograph.
where the medial contour of the soft tissue of the knee changed from convex to concave\(^{71}\), or, if the change in contour was not obvious, the mid-point between the most convex and concave curves. The actual HKA angle was determined on the radiographs then the three test points were substituted for the tibial interspinous groove point and estimated HKA angles were calculated using each of these points.

4.3.5.2 Ankle

Four potential points to estimate the centre of the ankle were substituted for the centre of the tibial plafond on radiographs and the resulting estimated HKA angles were compared to the actual HKA angle. The first point was the mid-point between the medial and lateral malleoli, marked by a tantalum bead placed on the participant\(^{128}\). The second point was the centre of the ankle determined by visual inspection which was also marked by a tantalum bead\(^{116}\). The third point was located as the mid-point of a line drawn on the radiograph between the medial and lateral malleoli\(^{71}\). The fourth point was the mid-point of a line drawn on the radiograph between the two most concave contours of the distal tibial shaft, a novel point that appeared easy to locate and was positioned along the tibial shaft. A fifth option was also assessed, this time using a photograph to estimate the HKA angle. The fifth point was the mid-point of a line drawn horizontally at the crease where the ankle meets the foot. This point was suggested because this crease is more often seen on a photograph than are the malleoli.

4.3.5.3 Proximal Femoral Point

The location of the centre of the femoral head is difficult to estimate on a photograph. Therefore, instead of locating the exact centre of the femoral head, we strove to determine a
proximal femoral point which would allow the HKA angle to be estimated accurately and reliably. Six potential points were considered and the resulting estimated HKA angles were calculated from photographs. Photographs rather than radiographs were used because two of the options used the contours of the thigh, which were not visible on the radiographs. Estimated HKA angles were compared to the actual HKA angle. For the first point, the distance between the right and left ASIS markers was measured. The point was located 32% of this distance medial to, and 34% of this distance distal to the ASIS \(^{130,277}\). The second point was similar, and located 14% of the inter-ASIS distance medial to, and 79% of this distance distal to the ASIS \(^{131}\). The third point was positioned 2 cm distal to the mid-point of a line drawn between the ASIS and the SPS \(^{129}\). The fourth point was located a distance medial to the ASIS determined by the following regression equations: \([-6.81 + (0.379 \times \text{ASIS} - \text{ASIS})]\) for men and \([-6.25 + (0.362 \times \text{ASIS} - \text{ASIS})]\) for women \(^{132}\). The inter-ASIS distance was measured in mm. The fifth point was the mid-point of the width of the uppermost thigh \(^{71}\) and the sixth point was the mid-point of the width of the thigh located one-half of the way between the knee and the uppermost inner thigh, a novel point that showed promise in preliminary evaluations.

4.3.6 Determination of Frontal-Plane Alignment - Photograph

The HKA-P angle for each photograph was determined using a custom version of the Surveyor™ image analysis program (see Figure 4-4). The preferred estimated knee and ankle joint centre points and the proximal femoral axis point, as identified from the above evaluation, were marked on each photograph. The HKA-P angle was calculated automatically by the
Figure 4-4: Determination of the hip-knee-ankle angle with a full-length lower extremity photograph (HKA-P).

1. Proximal femoral point [32% of the inter-anterior superior iliac spine (ASIS) distance medial to and 34% of this distance distal to the ASIS]
2. Estimated centre of the knee (mid-point of a horizontal line drawn across the knee where the medial contour of the soft tissue of the knee changes from convex to concave)
3. Estimated centre of the ankle (mid-point of a line drawn straight across at the crease where the ankle meets the foot)
software as the angle between a line from the proximal femoral point to the estimated centre of the knee and a line from the estimated centre of the knee to the estimated centre of the ankle.

4.3.7 Data Analysis

Analyses were performed using Minitab (version 15.1.30.0, Minitab Inc., State College, PA) and MedCalc (version 12.2.1.0, MedCalc Software, Mariakerke, Belgium). Statistical significance was accepted for p-values of less than 0.05.

4.3.7.1 Identification of Knee and Ankle Joint Centres and a Proximal Femoral Point on a Photograph

Estimated HKA angles calculated using the various joint centre estimates were correlated to the actual HKA angle using Pearson’s correlation coefficients. The Bland-Altman bias, along with its 95% confidence interval (CI), between each estimated HKA angle and the actual HKA angle were also calculated. The knee, ankle and proximal femoral points which produced the estimated HKA angle with the highest correlation to the actual HKA angle and the smallest 95% CI were chosen to create the HKA-P angle. The Bland-Altman bias must be added to the HKA-P angle in order to approximate the HKA angle.

4.3.7.2 Reliability of the HKA-P Angle

Photographs taken at the first testing session were used to assess intra-rater and inter-rater reliability. ICC_{(2,1)} and Bland-Altman analyses were used to assess the intra-rater reliability between reading times one and two for readers one, two and three, separately.
ICC\(_{(2,1)}\) was also used to assess the inter-rater reliability between readers one, two and three for reading time one\(^{278}\). Additionally, Bland-Altman analyses were used to assess the inter-rater reliability between each pair of readers for the first reading time.

HKA-P angle readings from the first reading time of the first testing session and from the second testing session were used to assess test-retest reliability. Readings from reader one were used for these analyses. ICC\(_{(2,1)}\) and Bland-Altman analyses were used to determine the level of agreement between HKA-P angles from the two testing sessions.

The rating system for ICCs documented by Fleiss\(^{108}\) was used in this study. ICCs of less than 0.4 were designated poor, ICCs between 0.4 and 0.75 as fair to good and ICCs above 0.75 as excellent\(^{108}\). ICC\(_{(2,1)}\) results for reliability should be greater than 0.75 (i.e. highly reliable) if they are to be useful clinically. In order for the HKA-P angle to be clinically acceptable, it was decided by the authors that the estimated Bland-Altman bias between readers or reading times or photograph sessions should not be greater than 2° and the limits of agreement for each type of reliability should not be beyond ± 3°.

Finally, the minimal detectable change at the 95% level (MDC\(_{95}\) ) was calculated using the test-retest reliability data\(^{279}\). The MDC\(_{95}\) is the smallest change that can be detected beyond random error\(^{109}\). The MDC\(_{95}\) was calculated as [standard error of the measurement (SEM) * 1.96 * \(\sqrt{2}\)], where the 1.96 is the z-score associated with a 95% confidence interval and the \(\sqrt{2}\) accounts for the fact that there are two measurements, each associated with their own error component\(^{279-281}\). The SEM, or the difference between the observed and true values, was calculated as \([s \sqrt{(1 - r)}]\), where \(s\) is the standard deviation of the samples and \(r\) is the correlation coefficient, in this case the ICC\(_{(2,1)}\)\(^{280,282}\). To use this version of the SEM, the sample variances
must be the equal and the participants must be truly stable. Both of these assumptions held true for our participants.

4.3.7.3 Concurrent Validity Between the HKA and HKA-P Angles

The relationship between the HKA and HKA-P angles measured from the first photograph session (as assessed by reader one the first time) was examined using a Pearson’s correlation coefficient and Bland-Altman analysis. Photographs from the first testing session were used because they were taken in the exact same position as the radiographs and thus would give an accurate estimate of the correlation between photograph and radiograph.

Pearson’s correlations were interpreted using the following descriptors: a value of 0.80 or higher was considered to show a very high correlation, a value of 0.60 to 0.80 indicated high validity, 0.30 to 0.60 indicated moderate validity and a value of less than 0.30 indicated low validity.

4.4 Results

4.4.1 Participants

Fifty eligible adults participated in the study, 14 males and 36 females. Demographic information is presented in Table 4-1. Our participant sample showed a large range of variability with respect to all recorded demographic characteristics. The HKA-P and HKA angle values were determined for the right and left knees. Reliability and validity calculations for the right and left knees were not significantly different; therefore data for the right knee only are presented.
Table 4-1: Demographic characteristics of the participant sample.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>41.8</td>
<td>21.5</td>
<td>20 – 86</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>169.6</td>
<td>8.5</td>
<td>157 - 196</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>71.5</td>
<td>14.7</td>
<td>48.0 – 124.3</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>24.7</td>
<td>3.9</td>
<td>17.1 – 37.2</td>
</tr>
</tbody>
</table>
4.4.2 Identification of Knee and Ankle Joint Centres and a Proximal Femoral Point on a Photograph

The Pearson’s correlations between the estimated HKA angles produced with each of the tested points for the knee and ankle centres and the proximal femoral point and the actual HKA angle are presented in Error! Not a valid bookmark self-reference. The Bland-Altman biases for each estimated HKA angle produced with the tested points are also found in Error! Not a valid bookmark self-reference.

The third knee point, the mid-point of a horizontal line drawn across the knee where the medial contour of the soft tissue of the knee changed from convex to concave (Pearson’s r = 0.95, Bland-Altman bias of -0.5°) was chosen as the preferred point to estimate the centre of the knee. The fifth option for the ankle point, which was the mid-point of a line drawn horizontally at the crease where the ankle meets the foot (Pearson’s r = 0.99, Bland-Altman bias of -0.6°), was chosen to estimate the centre of the ankle. While the third option for the ankle point, the mid-point of a line drawn on the photograph between the medial and lateral malleoli, produced the same correlation and bias results, the fifth option was preferred because the malleoli are not always visible on photographs. The first option for the proximal femoral point (Pearson’s r = 0.94, Bland-Altman bias of 5.0°, CI 4.7, 5.3), was chosen because it had the highest correlation to the HKA angles and the narrowest 95% CI. The bias between the estimated HKA and the HKA angles was considered reasonable and simple to use. An overall bias of 5.0° was defined between the HKA and HKA-P angles, because the biases for the knee and ankle centres were small compared to the bias for the proximal femoral point.
Table 4-2: Pearson’s correlations (r) and Bland-Altman biases between estimated hip-knee-ankle (HKA) angles calculated with different estimates of the knee and ankle joint centres and the proximal femoral point, and the actual HKA.

CI – confidence interval

<table>
<thead>
<tr>
<th>Joint</th>
<th>Point</th>
<th>Point Description</th>
<th>Pearson’s r (95% CI)</th>
<th>Aland-Altman bias (°) (95% CI for bias)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knee</td>
<td>1</td>
<td>inferior pole of the patella, marked with a dot</td>
<td>0.55 (0.27, 0.74)</td>
<td>0.1 (-0.6, 0.9)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>mid-point between the medial and lateral joint lines, marked with a dot</td>
<td>0.80 (0.64, 0.89)</td>
<td>0.6 (0.2, 1.1)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>mid-point of a horizontal line drawn across the knee where the medial contour of the soft tissue of the knee changes from convex to concave</td>
<td>0.95 (0.91, 0.98)</td>
<td>-0.5 (-0.7, -0.3)</td>
</tr>
<tr>
<td>Ankle</td>
<td>1</td>
<td>mid-point between the medial and lateral malleoli, marked with a dot</td>
<td>0.93 (0.86, 0.96)</td>
<td>-2.4 (-2.7, -2.2)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>visually inspected centre of the ankle, marked with a dot</td>
<td>0.94 (0.89, 0.97)</td>
<td>-2.2 (-2.4, -2.0)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>mid-point of a line drawn on the photograph between the medial and lateral malleoli</td>
<td>0.99 (0.98, 1.00)</td>
<td>-0.6 (-0.7, -0.5)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>mid-point of a line drawn on the photograph between the two most concave contours of the distal tibial shaft</td>
<td>0.98 (0.97, 0.99)</td>
<td>-0.8 (-1.0, -0.7)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>mid-point of a line drawn horizontally at the crease where the ankle meets the foot</td>
<td>0.99 (0.97, 0.99)</td>
<td>-0.6 (-0.7, -0.4)</td>
</tr>
<tr>
<td>Proximal Femoral Point</td>
<td>1</td>
<td>32% of the inter-ASIS distance medial to, and 34% of this distance distal to the ASIS</td>
<td>0.94 (0.88, 0.97)</td>
<td>5.0 (4.7, 5.3)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>14% of the inter-ASIS distance medial to, and 79% of this distance distal to the ASIS</td>
<td>0.92 (0.84, 0.96)</td>
<td>-0.6 (-1.1, -0.0)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2 cm proximal to the mid-point of a line drawn between the ASIS and the SPS</td>
<td>0.93 (0.86, 0.96)</td>
<td>2.8 (2.3, 3.2)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>point medial to the ASIS determined by separate regression equations for men and women</td>
<td>0.87 (0.76, 0.93)</td>
<td>-1.6 (-2.0, -1.2)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>mid-point of the width of the uppermost thigh</td>
<td>0.86 (0.73, 0.93)</td>
<td>-2.0 (-2.5, -1.4)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>mid-point of the width of the thigh located one-half of the way between the knee and the uppermost inner thigh</td>
<td>0.85 (0.71, 0.92)</td>
<td>-0.5 (-1.0, 0.1)</td>
</tr>
</tbody>
</table>
4.4.3 Reliability of the HKA-P Angle

A summary of the HKA-P and HKA angle values is shown in Table 4-3. ICC\(_{(2,1)}\) and Bland-Altman biases and limits of agreement results are presented in Table 4-4. Intra-rater, inter-rater and test-retest reliability were excellent, with high correlations, no discernible bias and small limits of agreement. A sample Bland-Altman plot, for intra-rater reliability, is presented in Figure 4-5. The SEM for test-retest reliability was 0.97° and the resulting MDC\(_{95}\) was 2.69°.

4.4.4 Concurrent Validity Between the HKA and HKA-P Angles

The Pearson’s correlation between the HKA and HKA-P angles was 0.92 (p < 0.0001). The HKA-P angle was an average of 4.5° more varus than the HKA angle, with limits of agreement between -6.9° and -2.1°. See Figure 4-6.
Table 4-3: Hip-knee-ankle angle assessed from a photograph (HKA-P) and hip-knee-ankle (HKA) angle results assessed from a radiograph for the right knee.

<table>
<thead>
<tr>
<th></th>
<th>mean (°)</th>
<th>standard deviation (°)</th>
<th>range (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HKA-P first session</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reader 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first reading</td>
<td>-4.9</td>
<td>3.1</td>
<td>-12.3 – 0.9</td>
</tr>
<tr>
<td>second reading</td>
<td>-4.8</td>
<td>3.1</td>
<td>-12.4 – 1.1</td>
</tr>
<tr>
<td>reader 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first reading</td>
<td>-5.0</td>
<td>3.1</td>
<td>-12.7 – 1.0</td>
</tr>
<tr>
<td>second reading</td>
<td>-5.0</td>
<td>3.2</td>
<td>-12.8 – 3.4</td>
</tr>
<tr>
<td>reader 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first reading</td>
<td>-4.9</td>
<td>3.1</td>
<td>-12.5 – 2.0</td>
</tr>
<tr>
<td>second reading</td>
<td>-4.9</td>
<td>3.1</td>
<td>-12.1 – 2.2</td>
</tr>
<tr>
<td><strong>HKA-P second session</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reader 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first reading</td>
<td>-5.1</td>
<td>3.1</td>
<td>-12.7 – 1.7</td>
</tr>
<tr>
<td><strong>HKA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reader 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first reading</td>
<td>-0.4</td>
<td>3.2</td>
<td>-8.9 – 7.4</td>
</tr>
</tbody>
</table>
Table 4-4: Intraclass correlation coefficient (ICC\(_{2,1}\)) and Bland-Altman analysis results for intra-rater, inter-rater and test-retest reliability for the hip-knee-ankle angle assessed from a photograph (right knee).

<table>
<thead>
<tr>
<th></th>
<th>ICC(_{2,1}) (95% confidence interval)</th>
<th>Bland-Altman Bias (°)</th>
<th>Bland-Altman Limits of Agreement (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intra-rater Reliability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 1</td>
<td>0.995 (0.991, 0.997)</td>
<td>-0.10</td>
<td>-0.68, 0.46</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.985 (0.974, 0.991)</td>
<td>0.01</td>
<td>-1.08, 1.09</td>
</tr>
<tr>
<td>Reader 3</td>
<td>0.996 (0.993, 0.998)</td>
<td>-0.01</td>
<td>-0.56, 0.54</td>
</tr>
<tr>
<td><strong>Inter-rater Reliability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Readers 1 &amp; 2</td>
<td>0.988 (0.981, 0.993)</td>
<td>0.06</td>
<td>-1.02, 1.15</td>
</tr>
<tr>
<td>Readers 1 &amp; 3</td>
<td></td>
<td>0.01</td>
<td>-0.86, 0.88</td>
</tr>
<tr>
<td>Readers 2 &amp; 3</td>
<td></td>
<td>-0.05</td>
<td>-0.92, 0.81</td>
</tr>
<tr>
<td><strong>Test-retest Reliability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photographs 1 &amp; 2</td>
<td>0.903 (0.835, 0.944)</td>
<td>0.20</td>
<td>-2.4, 2.9</td>
</tr>
</tbody>
</table>
Figure 4-5: Bland-Altman plot of intra-rater reliability for Reader 1.

LOA – limits of agreement [mean of (Time 1 – Time 2) ± 1.96 * standard deviation of (Time 1 – Time 2)].
Figure 4-6: Bland-Altman plot for the hip-knee-ankle (HKA) angle and the hip-knee-ankle angle measured from a photograph (HKA-P) (concurrent validity).

LOA – limits of agreement [mean of (HKA-P – HKA) ± 1.96 * standard deviation of (HKA-P – HKA)].
4.5 Discussion

Our results show that the HKA-P angle as estimated using the selected proximal femoral, knee and ankle points seen on a frontal-plane photograph had excellent intra-rater, inter-rater and test-retest reliability. Furthermore, the HKA-P and HKA angles were highly correlated, with the HKA-P angle 4.5° more varus than the HKA angle.

The knee, ankle and proximal femoral points chosen to determine the HKA-P angle were upheld by the strong reliability and validity results. Several possibilities for each point were determined from the literature and tested on a limited selection of participants. The chosen knee point was the mid-point of a horizontal line positioned across the knee at the point where the medial contour of the knee changed from convex to concave. While not all knees showed a distinct point at which this transition began, the readers were able to estimate it reliably. Similarly at the ankle, the estimation of where the ankle meets the foot may not always be perfect because of differences in light and shadows on the photographs, but the readers were able to mark this point reliability as well. Because these points are determined directly on the photograph, it is not necessary to use palpation and calipers to position sticky dots directly on a patient’s knee or ankle. This makes the procedure much more convenient and faster for the clinician or researcher.

The centre of the femoral head is difficult to estimate on a photograph, and therefore we instead aimed to find a proximal femoral point that would produce a HKA-P angle with the highest correlation to the HKA angle, as well as the smallest confidence interval. This was accomplished with a point located 32% of the inter-ASIS distance medial to, and 34% of this distance inferior to the ASIS. The chosen point is very simple to determine, and the ASIS is easy to palpate, even on individuals with a higher BMI. Overlying soft tissue must be accommodated
for, though, to ensure that the sticky dot is located directly over the ASIS. This point is not altered by rotating the LE, so it may be placed before the participant is positioned on the foot template.

Reliability results for the HKA-P angle were similar to or higher than those reported from similar studies performed on small numbers of young, healthy participants (ICCs of 0.627 to 0.997 for intra-rater reliability, 0.827 to 0.989 for inter-rater reliability and 0.700 to 0.904 for test-retest reliability)\textsuperscript{70,71}. Although inter-rater reliability is generally lower than intra-rater reliability\textsuperscript{152}, our results were the same for both types of reliability which highlights the consistency achievable when readers are trained in this method to determine the HKA-P angle. The reliability of the HKA-P angle was similar to that of the HKA angle [intra-reader reliability ICC 0.998 (95% CI 0.998, 1.000); inter-reader reliability ICC 0.995 (95% CI 0.994, 1.000)]\textsuperscript{106}.

There were several potential sources of variability between readers and reading times. Some sources which may have affected the readers’ decision of where to place the proximal femoral, knee centre and ankle centre points included glare from the camera flash on the background surface, shadows, body hair and body habitus (which caused differences in soft tissue contours). Despite these potential causes of variability, ICC results were consistently 0.99 for intra-rater and inter-rater reliability. No bias for reliability was found, which confirmed that our readers did not alter their technique between readings, and that they were very similar in their application of the written guidelines. The Bland-Altman limits of agreement show that 95% of the data points were within 1.1° of the mean HKA-P angle for intra-rater and inter-rater reliability, which confirms the small amount of variability between reading sessions and between readers.
Test-retest reliability may be affected by additional potential sources of variability, such as participant LE positioning, palpation of the ASIS and placement of the sticky dots, position of clothing and upper extremities, and set-up of the camera, including distance from lens to the participant’s knee, horizontal position of the lens, position of the lens at the level of the patella, optical zoom setting and ambient lighting. As with intra-rater and inter-rater reliability, there was no bias in the test-retest reliability data, which indicates that the protocol was consistently followed in the two testing situations. The Bland-Altman limits of agreement show that 95% of the data points are within 2.6° of the mean HKA-P angle.

The HKA-P angle was highly correlated to the HKA angle, which suggests that the HKA-P angle may be used confidently to estimate the HKA angle, provided that 4.5° is added to the HKA-P angle to adjust for the bias between the two alignment measures. Schmitt et al. 71, using a participant sample of 10 individuals with moderate BMI (19 to 28 kg/m²) also found a very high correlation (Pearson r = 0.98, p < 0.001) between the HKA-P and HKA angles, with the HKA-P angle 0.9° more varus than the HKA angle. We show similar results in a larger and more-varied sample. The MDC₉₅ was 2.69°, so a difference of 3° between individuals or between baseline and follow-up would be considered a true difference. This MDC is small enough to be useful clinically for monitoring the progression of deformity and the potential for onset or progression of knee OA. Moncrieff and Livingston 70 reported SEMs of 1.00° to 1.82°, resulting in MDC₉₅ calculations of 2.77° to 5.04°. A MDC₉₅ of 5.04° would be too large to be clinically useful as the typical range of the HKA angle is approximately -15° to 15°.

Three studies used goniometry performed directly on the participant to estimate the HKA angle 62, 66, 69. While test-retest reliability measures were excellent (ICC 0.84 to 0.94), the correlation between goniometric alignment and the HKA angle varied between insignificant
Goniometric alignment varied from 4.4° more valgus than the HKA angle (with the femoral axis located along the femoral shaft) to approximately 8.1° more varus than the HKA angle (with the femoral axis located along a line from the knee to the umbilicus). It is likely that the physical difficulty of manipulating a goniometer while sitting or kneeling on the floor in front of a participant contributed to the reduced reliability of these methods compared to the HKA-P angle. The goniometric methods used to approximate femoral anatomical alignment (the angle between the femoral and tibial shafts) do not estimate the HKA angle as consistently as those using photographs.

Standardization of the testing position was very important. Changes in limb rotation and foot position can alter the HKA angle, therefore, these parameters must be controlled for in the determination of the HKA-P angle as well. Prior studies using photographs had participants stand in a self-selected position, in the Romberg stance position (with medial borders of feet touching), with feet pointing straight forwards or in 30° external rotation. None of these positions account for the variability between individuals with respect to rotation of the femur and tibia, flexibility of the feet (for example, pes planus leads to internal rotation of the tibia), and the relative length of the hip musculature (for example, a tight piriformis can lead to excessive external rotation of the hip when in a self-selected stance position). Others use anatomical landmarks based on such features as the patella and the tibial tubercle; however, these too can vary between individuals. To accommodate between-participant variability, we used a template to position participants with their heels a consistent 9 cm apart, and altered the foot position so that the knees flexed within the sagittal plane, in other words, the axis of rotation for
knee flexion was in the frontal plane. This procedure reduced the likelihood of rotational error. Foot position was recorded from the foot template, for consistency at the second testing session.

Training for readers to assess the photographs for the HKA-P angle was not difficult, and took one to two hours depending on the user’s level of experience with the computer software. A two-page instruction sheet was provided for reference and was reviewed before commencing the second reading session. Readers’ backgrounds varied, and included one physiotherapist, one physiotherapist/researcher and one non-healthcare professional. Therefore, the determination of a reliable and valid HKA-P angle does not require intense training or a specialized background.

The HKA-P angle is ideal to use in a clinical setting, by physicians, physiotherapists, kinesiologists and nurse practitioners, without the cost, ionizing radiation or inconvenience associated with full-length LE radiography. It can be used in clinical and research applications to screen and assess individuals at risk of knee OA, and possibly to monitor the progression of varus or valgus deformity. It can also contribute to treatment decision-making and may be used as an outcome measure to determine the effectiveness of interventions for malalignment including orthotics, high tibial osteotomy or total knee arthroplasty. Several features of the HKA-P angle make it easy to use with patients in a clinical setting. Preparing the patient with sticky dots on the ASISs and positioning him or her on a template is simple, convenient and safe for the patient. The patient can wear any clothing that allows access to the skin over the ASISs and does not obscure the knee. A foot template is needed but a ruler is not required for the chosen proximal femoral, knee and ankle points. A room allowing at least 3 m between the patient and the camera is suggested to reduce parallax on the photograph. Digital photographs can be easily and safely archived for future reference and comparison to follow-up photographs.
Software which provides tools to measure distances and angles is recommended to analyse photographs for the HKA-P angle. This makes it easier to calculate the distance between the left and right ASISs and the mid-point of the lines at the knee and ankle. The proximal femoral point can then be determined using the required percentage distance between the two ASISs. A calibrated ruler makes this easier but is not necessary as other measurements (for example, pixels) can be used to calculate the percentages. While others have magnified the photographs and used a goniometer to measure the HKA-P angle\textsuperscript{69, 70}, we do not recommend this method as it might affect the precision of the measurements and lead to lower reliability\textsuperscript{70}.

There are a few limitations of this study and a few areas to investigate further. While we had a good range of individuals with respect to age and BMI, our participant sample does not focus solely on those most at risk of knee OA (i.e. individuals over 50 years old, with obesity, knee pain or prior knee injury). Nonetheless the reliability findings are strong, suggesting that the method could be used to measure the HKA-P angle in a population more at risk for knee OA as well. The photographs were somewhat less clear when enlarged and a camera with a resolution higher than 7.1 megaPixels would improve the photograph quality. Study of the sensitivity to change of the HKA-P angle has yet to be performed. A longitudinal study of individuals at high risk for progression of varus or valgus deformity would be required to test this psychometric property.

In conclusion, the best points to estimate the centres of the knee and ankle and the proximal femoral point have been determined. The HKA-P angle determined from a pelvis-to-ankle photograph using these points can confidently be used in place of the HKA angle determined from a full-length LE radiograph. The correlation between the HKA-P and HKA
angles is high, however the HKA-P angle is 4.5° more varus and this bias must be applied to accurately estimate the HKA angle.
Chapter 5

Reliability of the Unicompartmental Osteoarthritis Grade (UCOAG) for the radiographic assessment of knee osteoarthritis

5.1 Abstract

Objectives: The unicompartmental osteoarthritis grading (UCOAG) scale is a composite scale for grading knee radiographs for the severity of knee osteoarthritis (OA). The purpose of this study was to determine the intra-rater, inter-rater and test-retest reliability of the UCOAG when applied to posteroanterior fixed-flexion knee radiographs.

Methods: One sample of 100 knee radiographs was selected from the Multicentre Osteoarthritis Study to study intra- and inter-rater reliability. A magnetic resonance imaging (MRI)-based score was used to ensure that selected radiographs represented a wide range of knee OA severity. Three readers applied the UCOAG to each radiograph twice, two weeks apart. A second sample of 100 radiograph pairs, of individuals which did not change over 15 or 30 months, was selected to study test-retest reliability. One reader applied the UCOAG to the radiographs. Intraclass correlation coefficients (ICC_{(2,1)}) and Cohen’s weighted kappas were used to determine the level of each type of reliability. The minimal detectable change (MDC_{95}) was calculated using data from sample two.
**Results:** ICC\(_{(2,1)}\) results were 0.82 to 0.91 for intra-rater reliability, 0.77 for inter-rater reliability and 0.84 for test-retest reliability. Cohen’s weighed kappa results were 0.65 to 0.75 for intra-rater reliability, 0.47 to 0.61 for inter-rater reliability and 0.64 for test-retest reliability. The (MDC\(_{95}\)) was 2.61.

**Conclusions:** The UCOAG has moderate to excellent reliability, which is comparable to that of other radiographic knee OA grading scales. A change of three UCOAG grades specifies a change in knee OA severity. The UCOAG is recommended for clinical and research purposes.

### 5.2 Introduction

Osteoarthritis (OA) of the knee is a significant cause of disability in our aging population. A United States study using data procured between 1991 and 1997 determined that 16.4% of individuals 45 and over, and 20.8% of those 65 to 74 had symptomatic knee OA\(^{10}\). Knee radiographs are commonly used to diagnose knee OA and monitor progression. Reliable methods of scoring radiographic features of knee OA are important for grading OA severity, monitoring progression over time and determining the effect of interventions.

The most commonly-used grading scale for knee OA as seen on a radiograph is the Kellgren-Lawrence (KL) scale, which grades features of OA on a scale from zero to four\(^{81,170}\). The KL scale has been used extensively for grading the presence and severity of OA in epidemiological studies but was not designed for monitoring disease progression\(^{2,10,81,97}\). Grade one is described as “doubtful narrowing of joint space and possible osteophytic lipping” while grade two is described as “definite osteophytes and possible narrowing of joint space”\(^{170}\). Grade three is described as “moderate multiple osteophytes, definite narrowing of joint space and some
sclerosis and possible deformity of bone ends” and grade four is described as “large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends” \(^{170, 178}\). On the KL scale, radiographic OA is defined by a grade of two or greater \(^{81}\). Intra-rater reliability has been assessed as substantial to very good (Cohen’s weighted kappa 0.66 to 0.88, Spearman’s correlation coefficient 0.89) and inter-rater reliability as substantial (Cohen’s weighted kappa 0.56 to 0.80, Spearman’s correlation coefficient 0.85) following the suggestions for the interpretation of kappa made by Landis and Koch \(^{73, 181, 182, 286, 287}\).

Despite its widespread use, the KL scale has limitations \(^{173}\). The KL scale emphasizes osteophytes as the hallmark feature of knee OA \(^{73, 81, 175, 288}\). Osteophytes must be present for a KL grade other than zero to be given and joint space narrowing (JSN), which is considered to show radiological evidence of cartilage loss, is not required until grade three \(^{175, 176, 289}\). Brandt et al. \(^{147}\) showed that nine participants with significant JSN but no osteophytes (and therefore KL grade zero) all had definite OA changes identified on arthroscopy. Differentiating between grades zero and one, and one and two can be especially difficult \(^{175, 180, 186}\). A lack of sensitivity to change has also been reported \(^{75}\), and although the KL scale was not developed to monitor change in OA severity over time, it is frequently used for this purpose \(^{175, 290, 291}\). Finally, multiple versions of the descriptors assigned to each grade have been developed, creating variability in interpreting the grades \(^{73, 175, 178}\).

JSN on its own can be used as a radiographic measure of knee OA severity as well as change over time. JSN can be determined using an ordinal scale, from zero to three \(^{200}\). Medial and lateral compartments may be assessed separately or scoring from the two compartments can be combined \(^{153, 181}\). Joint space width (JSW) also can be calculated on a ratio scale by using a ruler or with a computerized algorithm \(^{73, 74}\). Mean or minimum JSW measurements are used,
depending on the method \textsuperscript{153, 222}. European and United States regulatory guidelines specify that JSW, as measured on a radiograph, is an objective outcome measure for assessing the effectiveness of potentially disease-modifying drugs for knee OA \textsuperscript{138}.

Measurement of JSN and JSW is influenced by many factors \textsuperscript{138, 222, 235}. Small variations in knee flexion or rotation on a radiograph image result in different parts of the cartilage being visible on the image \textsuperscript{138, 222}. In addition, meniscal subluxation associated with knee OA can influence the observed JSN or JSW \textsuperscript{138, 235}. Therefore, it is not surprising that JSN and JSW are only moderately correlated with cartilage loss as seen on arthroscopy, and cartilage volume measured from magnetic resonance imaging (MRI) \textsuperscript{147, 233, 292}. This lack of association is also influenced by the fact that while radiography assesses the thickness of the cartilage, arthroscopy only assesses the surface condition of the articular cartilage \textsuperscript{293}. Intra-rater reliability varies considerably depending upon which method is used (Cohen’s kappa values of 0.41 to 0.83 for ordinal estimates of JSN, 0.08 to 0.86 for ruler measurements of JSW and 1.00 for computer analyses of JSW) \textsuperscript{73}. Inter-rater reliability is also dependent on method and Cohen’s kappa results are generally slightly lower than for intra-rater reliability \textsuperscript{73}. The measurement of JSN and JSW does not consider other features of knee OA, such as osteophytes, bony erosion, sclerosis, subluxation or changes in alignment. Composite scores that include several key characteristics of knee OA may be better able to provide a more sensitive and reliable measure of OA severity \textsuperscript{140, 149-151}. Unlike the global KL scale, which requires the full description of a grade to be met in order to assign that grade, and JSN or JSN, which assesses only one feature of knee OA, composite measures sum the scores for several individual features of OA to obtain a total severity score \textsuperscript{140, 149-151}. These scores have the potential to detect several features of OA in a wide variety
of presentations of OA. Four composite scales, published by Kannus et al., McAlindon et al., Merchant et al. and Satku et al., score knee OA features such as JSN, osteophytes, sclerosis and cysts in several locations of the knee to create total scores of 100, 30, 10 or 14, respectively. For the Merchant et al. and Kannus et al. scales lower scores denoted more severe disease. The Kannus et al. scale has good to excellent reliability however it is very cumbersome and best-suited to research applications. None of these scales has been used extensively.

Another composite scale, the unicompartmental osteoarthritis grade (UCOAG) was introduced by Cooke et al. in 1999 to grade the severity of knee OA and to monitor its progression. JSN, femoral osteophytes, tibial erosion and subluxation are scored on a scale of zero to three or four, resulting in a total score out of thirteen. A significant difference between the UCOAG and other composite scales is that only the tibiofemoral (TF) compartment most-affected by OA is scored. While femoral osteophytes are included, tibial osteophytes are excluded in order to prevent over-weighting the scale with osteophytes and because tibial osteophytes frequently decrease in size as OA worsens and the knee subluxes. Tibial erosion is included because it is common and may contribute to joint instability as it progresses. Subluxation is also incorporated because it is a feature of joint instability. Sclerosis was not included because bone density is highly variable between people and is affected by obesity and variations in image quality. Initial inter-rater reliability testing was excellent (Cohen’s weighted kappa of 0.92, p < 0.001) using anteroposterior radiographs taken in full extension. Unpublished data demonstrated that intra-rater reliability using the same radiographs was very high (Cohen’s weighted kappa of 0.96, p < 0.001).
Recent protocols for knee radiograph acquisition place the knee in slight (7° to 10°) to moderate (25° to 30°) flexion while weight bearing. The greatest amount of wear on the femoral condyles occurs with the knee in some flexion and the medial meniscus does not contribute to the observed joint space in flexion, thus producing a more accurate measure of JSW. Therefore, there is higher sensitivity to detect knee OA changes with radiographic protocols which place the knee in some flexion. Knee OA grading scales also show greater reliability using radiographs taken with these protocols over those taken with the knee in full extension. Emphasis has been placed on the importance of using a standardized protocol to direct radiographic procedure. This allows for the accurate measurement of change over time, and enables the comparison of individuals within a research study and the comparison of findings across studies.

To ensure that grades obtained using the UCOAG are replicable, reliability of the UCOAG must be assessed. Therefore the purpose of this study was to determine the intra-rater, inter-rater and test-retest reliability of the UCOAG, obtained using posterior-anterior (PA) fixed-flexion TF radiographs.

Intra-rater reliability is the consistency achieved when one reader assesses a radiograph more than once. Inter-rater reliability is the consistency achieved when more than one reader assesses the same radiograph. Test-retest reliability is the consistency achieved when one reader assesses two radiographs of the same person, taken at two different times but which are deemed not to show change in OA severity. This absence of change is ideally determined using a method that gives more certainty than the one being assessed (i.e. the UCOAG). MRI may be considered a suitable criterion standard for this purpose because it is able to assess all tissues of the knee, without distortion or superimposition.
5.3 Participants and Methods

5.3.1 Radiograph Selection

5.3.1.1 Intra-rater and Inter-rater Reliability

One hundred PA fixed-flexion knee radiographs (left or right), taken at baseline, were selected from the Multicenter Osteoarthritis Study (MOST) database (Ancillary Study AS11-01; Analysis Plan AP11-10, see Appendix C). The MOST database consists of information on 3026 persons between the ages of 50 and 79 years with, or at risk of developing knee OA, including individuals who are overweight or obese, those with knee pain at the time of entry into the study and those with a history of knee injury or surgery \(^2,84\). Individuals were excluded from entry into the study if they had a diagnosis of rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, Reiter’s syndrome, significant kidney disease or cancer, and/or if they had bilateral knee replacement, were unable to walk without assistance or were planning to move out of the study area within three years of initiating participation \(^2\). Intake assessment included demographic data, patient questionnaires (Western Ontario and McMaster Universities Index \(^{298,299}\), Knee Injury and Osteoarthritis Outcome Score \(^{300,301}\), modified Late-Life Function and Disability Instrument \(^{302}\), Modified Physical Activity Scale for the Elderly \(^{303}\) and the Short Form-12 Health Survey \(^{304}\), a subjective interview and a physical examination \(^2\). Diagnostic imaging, performed at intake, 15 months and 30 months, included standing PA fixed-flexion knee radiographs and MRI images of the knee \(^2\). Further detail on MOST is available at http://most.ucsf.edu/default.asp (accessed 2011 to 2013).
The sample size of 100 participants was determined using the precision (confidence intervals) of the estimated limits of agreement according to Bland and Altman (1986)\(^{283}\); the 95% CI of the estimated limits of agreement are 1.96 +/- \sqrt{(3/n)}s, where n is the number of subjects and s is the standard deviation of the differences between the two repeated measurements. A sample size of 100 would thus give a 95% CI of 0.34s, so if we expect s to be approximately 2 on our scale of 0-13, the 95% CI would be 0.68, which is reasonable\(^{305}\).

An MRI-based score was used to ensure that a wide range of knee OA severity was represented in the 100 radiographs selected for this study. Potential participants were stratified for severity of knee OA using a custom summed Whole-organ magnetic resonance imaging score (WORMS), obtained from baseline knee MRI images\(^{191}\). The WORMS score assesses five articular features, each in 14 to 16 sub-regions of the TF and patellofemoral compartments of the knee\(^{191}\). Twelve non-articular features, such as meniscal tears and joint effusion, are also scored\(^{191}\). The custom summed WORMS score used for selection of knees for this study was made up of the scores for the medial and lateral tibial (anterior, central, posterior) and femoral (central, posterior) sub-regions for the following features of knee OA: cartilage morphology, osteophytes, bone attrition and meniscal extrusion for a maximum total score of 164. These four features were chosen to represent the components of OA that we were most interested in; i.e. those in the TF compartment which correspond to the features assessed by the UCOAG. Left and right knees were divided into four strata according to the WORMS summed scores and 25 knee radiographs were selected from each stratum.

While baseline MRIs were performed on 5036 knees in the MOST database, only 2243 knees had WORMS scores, as WORMS scoring was performed only for participants who were involved in specific longitudinal studies. For the 2243 knees available for selection, the highest
custom WORMS summed score was 122 out of 164. It was not possible to simply divide the 2243 knees into four groups with custom WORMS summed scores equally spaced from 0 to 41, 42 to 82, 83 to 123 and 124 to 164 because there would be no individuals in the most-severe group. Therefore, custom WORMS summed score ranges of 0 to 19, 20 to 39, 40 to 59 and 60 to 164 were used to divide the cohort into four groups. There were 1396 knees in the first group, 535 in the second group, 205 in the third group and 107 in the fourth group. These were the ranges which allowed for 25 knees to be selected in the most severe group, with the same proportion of knees with medial and lateral TF compartment involvement in each group.

To ensure that a representative number of knees with, or at risk of medial or lateral TF OA were included within each group, Osteoarthritis Research Society International (OARSI) JSN scores, from zero to three, were used to assess which TF compartment had the greater involvement of OA. MOST considers the TF compartment with the greater JSN grade as the one with OA, or potential for OA. If JSN grades were equal for both medial and lateral TF compartments, lower-limb alignment, as measured using the hip-knee-ankle (HKA) angle was used. Participants were considered to have greater potential for medial TF compartment involvement if the HKA angle was less than or equal to -1° (i.e. varus) and lateral TF compartment involvement if the HKA angle was greater than -1° (i.e. neutral or valgus).

Eighteen (72%) of the 25 knees selected in each of the four WORMS strata were of knees with, or at greater risk for medial TF OA and seven (28%) of the knees had, or were at risk for lateral TF OA. These proportions represent the proportions of medial and lateral JSN found on knee radiographs of the participants in the MOST database. Fifty-one percent of the resulting sample was of right knees. No individual had both knees selected.
5.3.1.2 Test-retest Reliability

To assess test-retest reliability, 100 radiograph pairs were selected from knees which had no change in cartilage morphology over a 15- or 30-month period. Because knee OA is manifested largely as change in articular cartilage, cartilage morphology was used as the WORMS feature to assess change. Since knee OA is a focal disease, we followed the TF sub-region with the worst cartilage morphology score over 15 or 30 months. If this score increased the knee was not eligible for selection. The number of knees that met the criteria for no change was 1352.

Potential knees were stratified for knee OA severity as described for intra- and inter-rater reliability. The four ranges determined using the custom WORMS score were the same. For test-retest reliability, there were 812 knees in the first group (WORMS scores from 0 to 19), 326 in the second group (WORMS scores from 20 to 39), 126 in the third group (WORMS scores from 40 to 59) and 88 in the fourth group (WORMS scores from 60 to 164). Fifty percent of the resulting sample was of right knees. No individual had both knees selected.

Images were de-identified by MOST and the contralateral knee was removed from the radiograph to prevent confusion. Images were numbered by MOST and a master list was retained by them until data collection was completed. PROC SURVEYSELECT procedure in Statistical Analysis Software (SAS®, version 9.2, SAS Institute Inc., Cary, NC) was used for participant selection.
5.3.2 Reader Training for the Unicompartmental Osteoarthritis Grade

The UCOAG grading scale includes four features of knee OA; JSN, scored out of three, femoral osteophytes, scored out of three, tibial erosion, scored out of four and subluxation, scored out of three. Individual feature scores are combined to create a composite score out of 13. Readers were recruited from a variety of backgrounds, including health care professionals and non-professionals. Training consisted of a four-hour group session conducted by the UCOAG developer. Printed instructions were provided to the readers. Topics covered included the assessment of alignment, which TF compartment to assess, detailed assessment of each feature of the UCOAG and the effect of rotation on UCOAG grades. Readers then practiced on a batch of 30 knee images during which they were able to ask for clarification, until they felt confident. A test batch of 24 images was evaluated by each trainee. The trainees’ UCOAG grades were reviewed for discrepancies and a one-hour review session was held in person or on-line. The reliability study began within one week after this review.

5.3.3 Procedure

Each PA fixed-flexion TF image was visualized on a computer screen using a custom imaging analysis program, Surveyor™ 3.1 (Orthopedic Alignment & Imaging Systems Inc., Kingston, ON). Image quality was optimized by adjusting for brightness, contrast and gamma, as needed. The reader noted which TF compartment (medial or lateral) of each knee was most affected according to the UCOAG criteria and analysed the four UCOAG features of this
compartment. The software recorded which knee compartment was read, the scores for each sub-category and the total score for each image.

5.3.3.1 Intra-rater and Inter-rater Reliability

Three readers scored each of the 100 images twice, at least two weeks apart. Image batches for each reader and time were randomized differently.

5.3.3.2 Test-retest Reliability

One reader scored the second sample of 200 images. Images were randomized with respect to knee and time.

5.3.4 Data Analysis

The UCOAG was analysed as a ratio scale and as an interval scale. The total UCOAG score was analysed, rather than the individual feature scores. Analyses were performed using Minitab (version 15.1.30.0, Minitab Inc., State College, PA) and MedCalc (version 12.2.1.0, MedCalc Software, Mariakerke, Belgium).

5.3.4.1 Intraclass Correlation Coefficients\( (2,1) \)

Intraclass correlation coefficients\( (2,1) \) (ICC\( (2,1) \)) were calculated to determine the level of agreement between the two scores obtained by each of the three readers (intra-rater reliability) and the first scores between the three readers for the same radiographs (inter-rater reliability) \(^{278} \).
ICC\textsubscript{(2,1)} was also used to determine the level of agreement between UCOAG scores on the paired radiographs (test-retest reliability)\textsuperscript{278}. The rating system used by Fleiss\textsuperscript{108} was used in this study. ICCs of less than 0.4 were considered poor, ICCs between 0.4 and 0.75 were considered fair to good and ICCs above 0.75 were considered excellent\textsuperscript{108}.

5.3.4.2 Cohen’s Weighted Kappas

Cohen’s weighted kappa is a more conservative method of comparing scores and is commonly used for nominal scales\textsuperscript{284}. Linear weightings were used, presupposing linearity among UCOAG grades. Cohen’s weighted kappas were used to determine the level of agreement between the two scores obtained by each of the three readers (intra-rater reliability), between the first scores obtained by all possible pairs of readers (inter-rater reliability) and between scores on the paired radiographs (test-retest reliability).

Cohen’s weighted kappa scores were interpreted such that scores of 0.0 to 0.20 showed slight agreement, 0.21 to 0.40 showed fair agreement, 0.41 to 0.60 showed moderate agreement, 0.61 to 0.80 showed substantial agreement and 0.81 to 1.00 showed very good agreement\textsuperscript{286, 287}.

5.3.4.3 Minimal Detectable Change\textsubscript{95}

The minimal detectable change at the 95\% level (MDC\textsubscript{95}) was calculated using test-retest reliability data. The MDC\textsubscript{95} is the smallest change that can be detected beyond random error\textsuperscript{109}. It provides useful information for clinicians or researchers because there needs to be a difference of at least the MDC\textsubscript{95} in order for a change on the UCOAG to be considered a true change. First, the standard error of the measurement (SEM) was calculated. This is defined as SEM = \( s/\sqrt{(1 - \text{reliability})} \), where “s” is the average standard deviation of the baseline and 30-month UCOAG
grades and “reliability” is the ICC\textsubscript{(2,1)} result \cite{280,282}. The MDC\textsubscript{95} was then calculated using the formula: $1.96 \times \sqrt{2} \times \text{SEM}$ \cite{109,281,307}. The 1.96 is the t-score for infinite degrees of freedom for a two-tailed test with $\alpha = 0.95$ and the $\sqrt{2}$ accounts for the presence of error in measurements taken from radiographs from two time points, baseline and follow-up \cite{281}. Turner et al. \cite{109} suggested that this was the best way to calculate the MDC for clinician-based clinimetric indices.

5.4 Results

5.4.1 Participants

The data for intra-rater reliability were obtained from three readers who evaluated 100 radiographs on two occasions. There were instances where the reader selected a different TF compartment to grade on the second occasion. This happened three times for reader one, and once for readers two and three. These cases were eliminated from the analysis leaving 97, 99 and 99 comparisons for readers one, two and three respectively. See Tables 5-1 and 5-2 for a description of the participant samples and the resulting UCOAG grades.

Inter-rater reliability was determined by comparing the UCOAG grades of the three readers on the first reading of 100 radiographs. Cases were only included in analyses where the three readers graded the same TF compartment. For the ICC\textsubscript{(2,1)} this resulted in a sample size of 97. For the reader pairs assessed with Cohen’s weighted kappa, this resulted in a sample size of 97 between readers one and two, 97 between readers one and three and 99 between readers two and three. See Tables 5-1 and 5-2.
Table 5-1: Description of participant samples.

<table>
<thead>
<tr>
<th></th>
<th>Intra-rater and Inter-rater Reliability mean (standard deviation)</th>
<th>Test-retest Reliability mean (standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>Males : Females</td>
<td>34:65</td>
<td>41:56</td>
</tr>
<tr>
<td>Age (years)</td>
<td>63.9 (8.0)</td>
<td>63.6 (8.7)</td>
</tr>
<tr>
<td>Body Mass Index (kg/m²)</td>
<td>30.2 (5.5)</td>
<td>31.3 (4.7)</td>
</tr>
<tr>
<td>Most-affected compartment as assessed by readers Medial : Lateral</td>
<td>81 : 18</td>
<td>81 : 16</td>
</tr>
<tr>
<td>WOMAC Physical Ability Score</td>
<td>15.5 (12.5)</td>
<td>18.4 (11.8)</td>
</tr>
<tr>
<td>WOMAC Knee Pain Score</td>
<td>3.4 (3.7)</td>
<td>4.2 (3.5)</td>
</tr>
<tr>
<td>20 m walk (average time, seconds)</td>
<td>17.1 (3.1)</td>
<td>17.0 (2.5)</td>
</tr>
<tr>
<td>5 chair stands (average time, seconds)</td>
<td>11.7 (3.5)</td>
<td>11.7 (3.4)</td>
</tr>
</tbody>
</table>

WOMAC – Western Ontario and McMaster Universities Arthritis Index; Physical Ability Score out of 68; Knee Pain Score out of 20.
Table 5-2: Unicompartmental osteoarthritis grades (UCOAG) for intra-rater, inter-rater and test-retest reliability.

<table>
<thead>
<tr>
<th></th>
<th>Intra-rater and Inter-rater Reliability</th>
<th>Test-retest Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reader 1 first reading</td>
<td>Reader 1 second reading</td>
</tr>
<tr>
<td>Number of participants</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>Mean (standard deviation)</td>
<td>4.2 (2.6)</td>
<td>3.6 (2.5)</td>
</tr>
<tr>
<td>Range</td>
<td>0-12</td>
<td>0-12</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
To determine test-retest reliability reader one evaluated 100 radiographic pairs. For three pairs, a different compartment was graded for the second radiograph of the pair; these cases were removed from the analysis leaving a sample of 97 paired radiographs. See Tables 5-1 and 5-2.

5.4.2 Intra-rater Reliability

Intra-rater reliability ICCs \(_{(2,1)}\) ranged from 0.82 to 0.91 which were excellent. Cohen’s weighted kappas were between 0.65 and 0.75, indicating substantial agreement. See Table 5-3. Investigation of intra-rater reliability results for individual features of the UCOAG was also conducted and none were found to have consistently poor reliability.

5.4.3 Inter-rater Reliability

The ICC \(_{(2,1)}\) of 0.77 showed excellent inter-rater reliability. See Table 5-3. The more conservative Cohen’s weighted kappas ranged from 0.47 to 0.61 and demonstrated moderate to substantial agreement between readers. Investigation of inter-rater reliability results for individual features of the UCOAG was also conducted and none were found to have consistently poor reliability.
Table 5-3: Measurements of intra-rater, inter-rater and test-retest reliability for the unicompartmental osteoarthritis grade (UCOAG).

<table>
<thead>
<tr>
<th></th>
<th>Reader</th>
<th>N</th>
<th>ICC$_{(2,1)}$ (95% CI)</th>
<th>Cohen’s Weighted Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intra-rater Reliability</strong></td>
<td>Reader 1</td>
<td>97</td>
<td>0.82 (0.70; 0.89)</td>
<td>0.65 (0.56; 0.74)</td>
</tr>
<tr>
<td></td>
<td>Reader 2</td>
<td>99</td>
<td>0.91 (0.87; 0.94)</td>
<td>0.75 (0.68; 0.82)</td>
</tr>
<tr>
<td></td>
<td>Reader 3</td>
<td>99</td>
<td>0.89 (0.83; 0.93)</td>
<td>0.70 (0.63; 0.77)</td>
</tr>
<tr>
<td><strong>Inter-rater Reliability</strong></td>
<td>Readers 1 vs. 2</td>
<td>97</td>
<td>0.77 (0.70, 0.84)</td>
<td>0.47 (0.37; 0.57)</td>
</tr>
<tr>
<td></td>
<td>Readers 1 vs. 3</td>
<td>97</td>
<td>0.77 (0.70, 0.84)</td>
<td>0.57 (0.48; 0.65)</td>
</tr>
<tr>
<td></td>
<td>Readers 2 vs. 3</td>
<td>99</td>
<td>0.61 (0.54; 0.69)</td>
<td></td>
</tr>
<tr>
<td><strong>Test-retest Reliability</strong></td>
<td>Time 1 vs. 2, reader 1</td>
<td>97</td>
<td>0.84 (0.71; 0.90)</td>
<td>0.64 (0.57; 0.72)</td>
</tr>
</tbody>
</table>

ICC – intraclass correlation coefficient  
CI – confidence interval
5.4.4 Test-retest Reliability

The ICC\(_{2,1}\) of 0.84 demonstrated excellent test-retest reliability and the Cohen’s weighted kappa of 0.64 also showed substantial agreement. See Table 5-3. The SEM was 0.94 and the resulting MDC\(_{95}\) was 2.61.

5.5 Discussion

The purpose of this study was to determine the intra-, inter- and test-retest reliability of the UCOAG grading scale when used to assess the severity of knee OA in the most-affected TF compartment as measured from PA fixed-flexion radiographs. The UCOAG has moderate to excellent intra-, inter- and test-retest reliability over a large range of knee OA severity.

The three readers were not informed of which TF compartment to assess with the UCOAG; they assessed the compartment which they deemed most-involved with respect to OA changes. For the intra- and inter-rater reliability assessments, all three readers agreed on the most-involved compartment, and at both time-points 97% of the time. Similarly, the same TF compartment was chosen 97% of the time for both of the images in the image pairs for the test-retest reliability study. These results are impressive, since some of the radiographs showed minimal or no signs of OA. This highlights the consistency of the UCOAG in the determination of which compartment is most affected with OA.

Intra-rater reliability of the UCOAG was substantial to excellent based on ICC\(_{2,1}\) (0.82 to 0.91) and Cohen’s weighted kappa (0.65 to 0.75) findings. These levels of agreement compare favourably with those from other knee OA grading scales. The KL scale has been
studied widely, and has reports of substantial to excellent intra-rater reliability (ICCs of 0.85 to 0.91, Cohen’s weighted kappas of 0.70 to 0.79 and Cohen’s kappas of 0.66 to 0.88)\textsuperscript{73, 157, 182, 184, 308}. Intra-rater agreement for ordinal measures of JSN is fair to excellent (ICCs of 0.41 to 0.95, Cohen’s weighted kappas of 0.75 to 0.87 and Cohen’s kappas of 0.41 to 0.83)\textsuperscript{73, 157, 182, 184, 308}. Some composite scales used for grading OA severity have not been tested for reliability\textsuperscript{150, 151}. However, Kannus et al.\textsuperscript{149} reported substantial intra-rater reliability (Cohen’s kappa of 0.70) and McAlindon et al.\textsuperscript{140} reported moderate agreement (Cohen’s kappa of 0.57).

Inter-rater reliability results are typically lower than intra-rater reliability\textsuperscript{152}; this trend was seen in our results (ICC\textsubscript{(2,1)} 0.77, Cohen’s weighted kappas of 0.47 to 0.61). The UCOAG thus had moderate to substantial inter-rater reliability. Again, this compares favorably with assessments of other scales, where there is a large amount of variability in published results\textsuperscript{73, 74, 149, 157, 182, 184, 309}. The KL scale has reported moderate to excellent inter-rater reliability (Cohen’s weighted kappa results of 0.56 to 0.80, Cohen’s kappa of 0.52 and ICCs of 0.68 to 0.93)\textsuperscript{73, 157, 182, 184, 309}. JSN ordinal measurements have fair to excellent agreement between readers (Cohen’s kappas for medial and lateral JSN of 0.30 to 0.71, ICCs of 0.36 to 0.85)\textsuperscript{73, 74, 157, 182, 184, 309}. The only composite scale to be tested for inter-rater reliability was that published by Kannus et al.\textsuperscript{149}, with a Pearson’s correlation of 0.94 and a Spearman’s correlation of 0.90. These assessment methods are not ideal for measuring reliability as they do not take into account systematic error and thus tend to give inflated scores\textsuperscript{279, 284}.

Analyses for test-retest reliability (ICC\textsubscript{(2,1)} of 0.84 and Cohen’s weighted kappa of 0.64) showed substantial to excellent agreement for UCOAG grades between paired radiographs. Because test-retest reliability includes the ability to replicate the radiograph set-up and positioning of the participant after 15 or 30 months, this result confirms the consistent application
of the radiograph protocol by the MOST team. Despite its importance, test-retest reliability is not routinely studied, and comparative results from other grading scales are not available. An MDC$_{95}$ of 2.61 suggests that a change of three UCOAG grades is necessary to show an actual change in knee OA severity. On a scale graded out of 13, this is a clinically meaningful result and can be used to monitor change in radiographic knee OA over time.

The data were reviewed to determine if there were any patterns or trends which would assist in increasing the reliability of the UCOAG. The four UCOAG features were investigated individually however none of them was more or less reliable than the others. Images of knees with less-severe OA were no less reliable than those of knees with more-severe OA. This is unlike the KL scale, where some have reported difficulty differentiating between grades zero and one, and one and two $^{175, 180, 186}$. Experience and training influence reliability $^{309-312}$. Professionals and experienced readers possess a wide range of experience with other scales and situations, which influences their interpretation of a grading scale $^{310}$. Variability between readers with different backgrounds and levels of experience is high $^{310, 311}$. Our readers consisted of one health care professional and two non-health care professionals. It is uncertain how our choice of readers affected our results. The non-health care professionals had excellent intra-rater reliability and it is possible that their lack of background in the area was actually of benefit, because they had no preconceptions or competing interpretations. On the other hand, a prior study in which the UCOAG was found to have very good inter-rater reliability (Cohen’s weighted kappa of 0.92) was performed with a highly-experienced orthopaedic surgeon and a physiotherapist $^{82}$. Based on the findings of the current study, both health-care professionals and non-professionals may be suitably trained to use the UCOAG to grade knee OA severity. Vignon et al. $^{312}$ and Guermazi et al. $^{309}$ emphasized that
for any one project one reader should be used and re-training should occur frequently, in order to maximize consistency. Cooper et al.\textsuperscript{152} noted large variations in intra-rater reliability between readers, which supports actively choosing a single reader with a high level of internal consistency. Increased training for the UCOAG beyond what was described for this study, including more feedback and more experience with unusual presentations of knee OA, might further increase reliability.

Many existing OA grading scales use atlases which contain radiographs of knees showing the characteristic appearance of the OA features at various levels of severity\textsuperscript{81, 141, 200}. Readers compare their images with those in the atlas to assign a grade. Atlases were not used in this study as one of the aims of the UCOAG was to have a simple scale that did not require an atlas. However, the use of an atlas might improve intra- and inter-rater reliability, as suggested by Vignon et al.\textsuperscript{312}.

The UCOAG shows sufficient reliability for research purposes, and can potentially be used to grade the presence and severity of radiological knee OA for small and large epidemiological studies. The UCOAG can also potentially be used clinically by physicians, to describe the radiological severity of knee OA in individual patients, to monitor the progression of OA over time and for treatment planning. The UCOAG has been included in the Knee Surgery Triage tool, which uses a combination of disability evaluation and radiographic grading to guide the decision to refer a patient for surgery\textsuperscript{313}.

Further psychometric testing of the UCOAG must also be performed to fully qualify it for these purposes. To test the validity of the UCOAG, grades should be correlated to a criterion standard of physical change caused by knee OA. The strongest criterion standard is joint change seen at knee surgery, however only individuals with higher levels of OA are eligible for surgical
intervention, and the resulting participant sample would be skewed. Another possible criterion standard is MRI assessment, and WORMS scores are available for many of the knees in the MOST database. For the UCOAG to be deemed suitable to assess change in OA severity over time, its sensitivity to change must also be tested. Radiograph pairs, taken at baseline and several months later, should be graded with the UCOAG, and the change in UCOAG grade compared to change graded with a criterion standard measurement. Again, a composite WORMS score may be used for this purpose.

UCOAG grades may also be compared to clinical measures of pain and dysfunction resulting from knee OA (for example, Western Ontario and McMaster Universities Arthritis Index, timed 20 m walk) in order to determine how UCOAG results relate to the patient’s experience of knee OA. Other severity grading scales for knee OA have not been closely aligned with the experience of OA \(^{10,314,315}\).

If the UCOAG is found to be a useful scale for the grading of TF OA, then it could be tested for its ability to grade OA in other joints (patellofemoral, hip, hand, shoulder). Modification might be required as each joint presents with slightly different features of OA.

In conclusion, the UCOAG has moderate to excellent intra-rater, inter-rater and test-retest reliability, which is comparable to that of currently-used scales. A change of three or more UCOAG grades indicates a true change in TF OA severity. The UCOAG is therefore recommended for clinical and research purposes, pending results from validity and sensitivity-to-change testing.
Chapter 6

Validity and sensitivity to change: A comparison of three scales for the radiographic assessment of knee osteoarthritis

6.1 Abstract

**Objectives:** The purpose of this study was to assess the concurrent validity and sensitivity to change of two existing knee osteoarthritis (OA) grading scales, the Kellgren-Lawrence (KL) and the Osteoarthritis Research Society International (OARSI) joint space narrowing (JSN) grading scales and a newer scale, the unicompartmental osteoarthritis grade (UCOAG), which grades JSN, femoral osteophytes, tibial erosion and subluxation to create a total score.

**Methods:** One sample of 72 posteroanterior (PA) fixed-flexion radiographs displaying mild to moderate knee OA was selected from the Multicenter Osteoarthritis Study to study validity. A second sample of 75 radiograph pairs, which showed an increase in OA severity from baseline to 30 months later, was selected to study sensitivity to change.

The three radiographic grading scales were applied to each radiograph in both samples.

Spearman’s rank correlation coefficients were used to correlate the radiographic grades and the change in grades over 30 months with a magnetic resonance imaging (MRI)-based score.

Standardized response means (SRM) were calculated using data from the second sample.

**Results:** Correlations between the KL, OARSI JSN and UCOAG grading scales and the MRI-based score were 0.830, 0.835 and 0.771 (p < 0.0001) respectively while correlations between change in the radiographic grading scales and change in the MRI-based score were 0.479, 0.515
and 0.473 (p < 0.0001) respectively. SRMs for the three scales were 0.96, 0.86 and 1.00 respectively.

**Conclusions:** All three radiographic grading scales can be used in place of MRI-based scores for the assessment of knee OA severity and change over time.

### 6.2 Introduction

Osteoarthritis (OA) of the knee affects 5.4% to 38% of older adults in Europe, the United States and Asia. This prevalence is expected to increase in the coming years, as the population in these regions ages and obesity and knee injury become more common. Knee OA causes knee pain, stiffness and disability; therefore it is important to be able to diagnose the condition early and to monitor progression over time so that treatment interventions may be administered early in the course of the disease.

The diagnosis of knee OA is determined with the presence of symptoms (for example, pain and stiffness) accompanied by radiographic changes. To facilitate objective and consistent assessments, radiographs are generally scored using grading scales. The most commonly-used grading scale is the Kellgren-Lawrence scale, which scores several features of OA in both the medial and lateral tibiofemoral (TF) compartments on an ordinal scale from zero to four. This global scale emphasizes osteophytes, and requires that all aspects of a grade description be met in order for a particular grade to be assigned. While the KL scale was intended to assess radiographs for the presence and severity of knee OA at one point in time, it has also been used to monitor change over time.
Another commonly-used scale is the Osteoarthritis Research Society International (OARSI) joint space narrowing (JSN) scale. This individual grading scale uses an atlas to compare radiographs to representative images and assign a grade for the severity of JSN from zero to three in either the medial or lateral TF compartment. The OARSI JSN scale is commonly used to monitor change in OA severity over time for epidemiological research. Both of these grading scales emphasize a single feature of knee OA (osteophytes for the KL scale and JSN for the OARSI JSN scale). A scale that includes several features of OA might be better for monitoring progression in people with a variety of presentations of OA. To address this issue a composite knee OA grading scale, the unicompartmental osteoarthritis grading scale (UCOAG), was designed to assess several features of knee OA individually but sum the scores for each feature to create a total score out of 13.

For grading scales to be recommended to assess knee OA on a radiograph, they must be valid (measure what they purport to measure) and sensitive to change. To assess concurrent validity, grades obtained from each scale must be compared to grades obtained from a criterion standard. One relevant criterion standard is observation of the knee joint made during surgery, however this is not a good choice because the study sample would consist solely of knees that required surgical intervention. Another criterion standard is knee OA severity measured from a magnetic resonance image (MRI). MRIs allow the observation of cartilage damage and eliminate issues due to magnification, distortion and superimposition. One MRI grading scale is the whole-organ magnetic resonance imaging score (WORMS), an ordinal, multi-feature grading scale for knee OA. Five articular features are assessed with the WORMS, each scored in several sub-regions of the articular surfaces of the tibia, femur and patella. Twelve non-articular features such as meniscal tears and joint effusion are also scored. While the features
included in the WORMS score sum to a maximum of 380, this is not generally done. Instead, the scores of individual features are more commonly used 191, 243, 247, 321. KL grades show moderate associations with cartilage lesions and volume as seen on MRI 78, 187. Comparisons of OARSI JSN and UCOAG grades to MRI findings have not been performed.

Sensitivity to change for radiographic grading scales is assessed using pairs of images taken from the same individual, at two time-points. The change in severity of knee OA observed using the radiographic grading scales is compared to the change in severity observed using a criterion standard. Again, WORMS scores are an appropriate criterion standard for this comparison. While sensitivity to change of the KL scale for knee OA has not been performed, results for hand OA showed a trivial to small responsiveness to change as measured with the standardized response mean (SRM), for change over one year (SRM 0.17 to 0.24, depending on the reader) 322. Correlations comparing sensitivity to change between MRI measures and the radiographic grading scales have not yet been assessed. The balanced and varied features of the UCOAG suggest that it might be sensitive to change in a variety of presentations of knee OA.

Therefore the first goal of this study was to determine if the KL, OARSI JSN and UCOAG grading scales were valid for measuring the severity of TF OA on a radiograph and to establish if one of these scales was superior to the others for this purpose. The second goal was to determine if the KL, OARSI JSN and UCOAG grading scales were sensitive to change in the severity of TF OA over a 30-month period and to ascertain if one of these scales was more sensitive than the others for detecting change over time.
6.3 Participants and Methods

6.3.1 Radiograph Selection

Knee radiographs were obtained from the Multicenter Osteoarthritis Study (MOST) database, a United States of America National Institutes of Health / National Institute on Aging-sponsored study on the prevention and treatment of knee OA (Ancillary Study AS11-01; Analysis Plan AP11-21, see Appendix D). The MOST study was approved by institutional review boards at the University of Iowa, University of Alabama, Birmingham, University of California, San Francisco and Boston University Medical Campus and participants provided written informed consent. This database consists of information on 3026 persons with, or at risk of developing knee OA, including individuals who are overweight or obese, those with knee pain at the time of entry into the study and those with a history of knee injury or surgery. The database includes information on individuals between the ages of 50 and 79 with most having mild or moderate knee OA. A detailed assessment at baseline included demographic data, patient questionnaires, a subjective interview and a physical examination of the lower extremity. Diagnostic imaging, performed at baseline and 30 months later, included bilateral weight-bearing fixed-flexion posteroanterior (PA) radiographs, taken using a Synaflexer™ positioning frame and the protocol by Peterfy et al. MRIs (1.0 tesla) of the knees were also acquired at baseline and 30 months. KL and OARSI JSN grades for knee radiographs and WORMS scores for MRIs performed at baseline and 30 months later were available for 1694 knees. Lower extremity alignment measurements, including the hip-knee-ankle (HKA) angle, measured on anteroposterior full-length radiographs were also available for each participant, from baseline
only. Further detail on MOST is available at http://most.ucsf.edu/default.asp (accessed 2011 to 2013).

6.3.1.1 Concurrent Validity

One sample of 72 PA fixed-flexion knee radiographs (left or right), taken at baseline, was selected from the MOST database to study concurrent validity. Sample size was calculated based on a Pearson’s correlation with two independent variables, a medium effect size, $\alpha = 0.05$ and statistical power ($1 - \beta$) = 0.80; it was estimated to be 67. This number was increased to 72 to facilitate the use of four strata of equal size ($n=18$) of OA severity in the experimental design.

To ensure that a wide range of knee OA severity was represented, potential participants were stratified according to a custom summed WORMS score. This score, computed specifically for this study, was made up of the individual scores for the medial and lateral tibial (anterior, central, posterior) and femoral (central, posterior) sub-regions for the following features of knee OA: cartilage morphology (ten sub-regions, each scored out of six), osteophytes (ten sub-regions, each scored out of seven), bone attrition (ten sub-regions, each scored out of three) and meniscal extrusion (two menisci, each scored out of two), for a maximum total of 164. These four features were chosen to represent the components of OA that we were most interested in; i.e. those in the TF compartment which correspond to the features assessed by the KL, OARSI JSN and UCOAG grading scales. Inter-rater reliability of the WORMS scores for these four features in the medial and lateral TF compartments has been assessed, with intraclass correlation coefficients of 0.65 to 0.92.

There were 1694 left and right knees which had WORMS scores from MRI. The cohort was divided into four groups using the following divisions of the custom summed WORMS
scores: 0 to 19, 20 to 39, 40 to 59 and 60 to 164. The range of scores in the highest score group was very large because most of the participants in the MOST study had mild to moderate knee OA and if the 1694 knees had been divided into four groups by dividing the custom summed WORMS score into four equal portions, the most-severe group would have no individuals in it\(^2\). There were 976 knees in the first group, 442 in the second group, 159 in the third group and 117 in the fourth group. To ensure that the same number of knees with, or at risk of medial and lateral TF compartment OA were included within each stratum, OARSI JSN grades and the HKA angle were used to assess which compartment had the greater involvement of OA\(^2,200\). MOST defined the most-affected TF compartment as the one with the greater OARSI JSN grade. If OARSI JSN grades were equal for both medial and lateral TF compartments, lower-limb alignment, as measured using the HKA angle was used\(^92\). Participants were considered to have medial TF compartment involvement if the HKA angle was greater than 1° of varus and lateral TF compartment involvement if the HKA was neutral (1° of varus to 1° of valgus) or greater than 1° of valgus\(^35,326\). Of the 1694 potential knees, 964 (57%) had, or were at risk for medial TF OA while 730 (43%) had, or were at risk for, lateral TF OA. For each of the sample groups of 18 participants, individuals were randomly selected in this proportion of medial and lateral involvement.

6.3.1.2 Sensitivity to Change

A second sample, of 150 PA fixed-flexion radiographs, was selected to study sensitivity to change. This sample consisted of paired radiographs for 75 knees, taken at baseline and 30 months later. The sample size estimation was the same as for participant sample one, but increased to 75 for simplicity (images are assessed in batches of 25).
To address the question of sensitivity to change, a minimal increase in OA severity was chosen to ensure the selection of participants which would allow the radiograph scales to be adequately tested. A small increase in severity would not be expected to be detected on a radiograph. Therefore an increase of at least 15% or 6 points (whichever was greatest) on the custom summed WORMS score (out of 164) from baseline to 30 months later was required. The 15% level was chosen because the UCOAG grading scale was estimated to have a minimal detectable change of 2 out of 13, which is approximately a 15% change. Therefore, we would not expect the UCOAG (or the other scales) to be sensitive enough to pick up a change of less than 15%.

An absolute minimum level of change from baseline to 30 months later was also required for the custom summed WORMS score. Otherwise, in a knee with a small amount of evidence of OA at baseline (for example, 14 points on the custom summed WORMS score), a 15% increase would be a small absolute number, which would not be detectable on a radiograph. To do this, we observed from the MOST database that for MRIs with a custom summed WORMS score of less than 40, there was a 75% chance of a KL grade of zero or one, which indicates no OA. However, for WORMS summed scores of 40 or greater, there was a 94% chance of a KL grade of two or greater, which indicates the presence of OA. We therefore calculated 15% of this score (40), which is six, as the minimal change that would be expected to be seen on a radiograph.

One hundred and seventy three knees met these criteria. Of these, 75 were randomly selected. Sixty-one percent of the resulting sample had the medial TF compartment most-severely affected with OA. PROC SURVEYSELECT procedure in Statistical Analysis Software (SAS®, version 9.2, SAS Institute Inc., Cary, NC) was used for participant selection.
6.3.2 Measurements

6.3.2.1 Kellgren-Lawrence Grades

Standing PA fixed-flexion radiographs were assessed to determine KL grades by two expert readers from the Boston University Clinical Epidemiology Research and Training Unit, a musculoskeletal radiologist and a rheumatologist, working independently. Radiographs were presented in random order and readers were blinded to clinical status. Baseline and follow-up films were scored while viewed simultaneously, with the chronological order of the images known to the readers. If there was a discrepancy between the results from the two readers for a particular participant, adjudication was used if the disagreement resulted in the participant being differently classified as having TF OA or not (KL grades of two or greater denote the presence of OA), or having a change in TF OA or not, over time. Adjudication consisted of the two readers and a third person meeting to achieve consensus. In other cases, the senior reader’s results were used. Inter-rater reliability has not been assessed for the version of the KL grading scale used by MOST.

KL grades (zero to four) were assigned to each knee; there was no distinction made between most- and least-affected TF compartments. Grade one describes “doubtful narrowing of joint space and possible osteophytic lipping”, grade two describes “definite osteophytes and possible joint space narrowing”, grade three describes “moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends” and grade four describes “large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends”. In the MOST protocol, for knees with a KL grade of four, a lateral radiograph was also viewed. A grade of 3.5, added by MOST to the
original KL scale, signified bone-on-bone cartilage erosion on the PA radiograph but some residual joint space seen on a lateral radiograph. For the present study, KL grades of 3.5 were changed to grades of four, in order to test the original KL grading scale. A KL grade of two or greater was considered indicative of incident TF OA and change in OA was present if any increase in KL grade occurred.

6.3.2.2 Osteoarthritis Research Society International Joint Space Narrowing Grades

OARSI JSN grades were assessed by expert readers from MOST, for PA fixed-flexion radiograph pairs presented in known chronological order, according to the same ordering, blinding and adjudication procedures used for KL grades, detailed above. Inter-rater reliability (Cohen’s kappa) for the two readers has been reported as up to 0.66 (p < 0.001).

OARSI JSN grades of zero to three were given for the most severely-affected TF compartment following the radiograph examples in the OARSI Radiographic Atlas. The grades corresponded to the following descriptors: zero - normal, one - mild change, two - moderate change, three - severe change. Additionally, in the MOST protocol, as the paired radiographs were assessed, if there was clear evidence of JSN worsening but not enough to assign the next grade, a half-grade was given for the second radiograph. In order to assess the original OARSI JSN grading scale, we changed all one-half grades to the lower integer (for example, a grade of 2.5 was changed to a grade of two). An increase in OA was designated if there was any increase in the OARSI JSN grade.
6.3.2.3 Unicompartmental Osteoarthritis Grades

One experienced reader, who was originally trained by the creator of the UCOAG, performed the UCOAG gradings on radiographs presented in random order and blinded to clinical status but with chronological order known for the radiograph pairs.

Each PA fixed-flexion knee image was visualized on a computer screen using a custom imaging analysis program, Surveyor™ 3.1 (Orthopedic Alignment & Imaging Systems Inc., Kingston, ON) as described in the literature. The four UCOAG features of the most severely-affected compartment were analysed: JSN (scored from zero to three), presence and size of femoral osteophytes (scored from zero to three), presence and degree of tibial erosion (scored from zero to four) and evidence of subluxation (scored from zero to three), resulting in a total score from zero to 13.

6.3.2.4 Whole-organ Magnetic Resonance Imaging Scores

WORMS scoring was performed on pairs of MRIs, presented in known chronological order, but blinded to clinical status and participant identifier. Readers were trained by MOST’s senior WORMS reader. Results for the most severely-affected TF compartment were used in this study. Because OA affects several articular tissues, the five articular features of the WORMS [(cartilage morphology (scored out of six), tibial and femoral osteophytes (scored out of seven), bone attrition (scored out of three), bone marrow lesions (scored out of three) and subchondral cysts (scored out of three)] were used to create a WORMS composite score for correlation to the KL, OARSI JSN and UCOAG grading scales, which was different from the one used for participant selection. The WORMS composite score included the score for the worst of the tibial (anterior, central, posterior) and femoral (central, posterior) sub-regions for each articular feature.
for a total maximum score of 22. Although the OARSI JSN scale assesses only one feature of knee OA, correlation to a WORMS composite score was appropriate because the OARSI JSN scale is commonly used on its own to measure knee OA progression\textsuperscript{317-320}. The creation of study-specific WORMS scores has been done previously to assess knee OA progression and for comparison to biochemical markers of OA\textsuperscript{327,328}.

To further investigate the components of the radiographic grading scales, individual features of the WORMS were also correlated to corresponding features of the UCOAG scale and to the OARSI JSN scale. Again, the most severely-affected TF compartment was assessed. UCOAG JSN and OARSI JSN grades were correlated with the WORMS cartilage morphology score for the worst of the tibial (anterior, central, posterior) and femoral (central, posterior) sub-regions. UCOAG femoral osteophyte grades were correlated with the WORMS osteophyte score for the worst of the femoral (central, posterior) sub-regions. UCOAG tibial erosion grades were correlated with the WORMS bone attrition score for the worst of the tibial (anterior, central, posterior) sub-regions. Finally, UCOAG subluxation grades were correlated with the WORMS meniscal extrusion score. This WORMS feature was chosen because of prior studies that have shown a moderate association of subluxation with lower-limb alignment (Pearson’s $r = 0.51$, $p < 0.001$), which is in turn associated with meniscal extrusion (Pearson’s $r = 0.45$ and $0.62$, $p < 0.001$, for varus alignment associated with medial meniscal subluxation and valgus alignment associated with lateral meniscal subluxation, respectively)\textsuperscript{82,329,330}. 

\textsuperscript{135}
6.3.3 Procedure

6.3.3.1 Concurrent Validity

The KL and OARSI JSN grades and WORMS scores had already been recorded by MOST; therefore only UCOAG grades were required. Each of the 72 images was graded with the UCOAG by a single reader. De-identified images were provided by MOST and the contralateral knee was removed from the radiograph to prevent confusion. Once the images were graded, unblinded data were released by MOST for each participant, including demographic data, KL and OARSI JSN grades, and WORMS composite scores.

6.3.3.2 Sensitivity to Change

KL, OARSI JSN and UCOAG grades and WORMS scores were obtained for the 75 pairs of radiographs used to assess sensitivity to change, as described for concurrent validity. For the UCOAG grades, images were presented in randomized pairs, with the chronological order known to the reader, to be consistent with the procedure followed by MOST for the other scales. Reading films in chronological order has been found to increase the detection of clinically-relevant change without overestimating non-relevant differences \(^{75,331-333}\). This procedure is also more clinically-relevant than using radiographs blinded to chronology \(^{331}\). The unblinded dataset was released from MOST following receipt of the UCOAG readings, as detailed above.
6.3.4 Data Analysis

6.3.4.1 Concurrent Validity

Spearman’s rank correlation coefficients were used to correlate the KL, OARSI JSN and UCOAG grades with the WORMS composite score. Although the OARSI JSN grading scale was primarily created to monitor change over time, rather than to assess OA severity at one point in time, it was included in the validity analysis in order to complete the comparison of radiographic scales to MRI findings. There is no distinct definition of the Spearman’s rank correlation coefficient required to deem a test or scale valid. However, a value of 0.80 or higher is generally considered to show a very high correlation between two features for Pearson’s correlation coefficients. Similarly, a value of 0.60 to 0.80 indicates high validity, 0.30 to 0.60 indicates moderate validity and a value of less than 0.30 indicates low validity. These descriptors were used for all reported correlations. There is no statistical test to compare two Spearman’s rank correlation coefficients. However, for all of the correlations, confidence intervals were used to compare correlation coefficients in a general way.

To investigate the association between the radiographic grading scales and WORMS scores, we did several post-hoc analyses. We calculated the Spearman’s rank correlation coefficient as above, but for knees with medial and lateral TF compartment involvement separately and with right and left knee involvement separately.

To further assess the concurrent validity of the OARSI JSN and UCOAG grades, individual components of these scales were correlated with comparable components of the WORMS composite score using Spearman’s rank correlation coefficients.
6.3.4.2 Sensitivity to Change

To assess the sensitivity to change of the KL, OARSI JSN and UCOAG grades, Spearman’s rank correlation coefficients were used to correlate the change in each grading scale from baseline to 30 months with the change in the WORMS composite score over the same 30-month period.

To further assess sensitivity to change, the change from baseline to 30 months for OARSI JSN grades and the individual features of the UCOAG grading scale were correlated with the change over the same 30-month period for the comparable features of the WORMS composite score, again using Spearman’s rank correlation coefficients.

Finally, the SRM was calculated for the KL, OARSI JSN and UCOAG grading scales. The SRM is a standardized measure of the responsiveness of a scale to change; it is the mean change score divided by the standard deviation of the change scores. In general, SRM values of 0.20 or less represent a trivial response to change, SRM values from 0.20 to 0.50 are small, SRM values from 0.50 to 0.80 are moderate and SRM values greater than 0.80 represent a large response. Analyses were performed using Minitab (version 15.1.30.0, Minitab Inc., State College, PA) and MedCalc (version 12.2.1.0, MedCalc Software, Mariakerke, Belgium). Statistical significance was set at \( \alpha = 0.05 \).
6.4 Results

6.4.1 Participants

The first sample, of 72 PA fixed-flexion baseline radiographs, is described in Table 6-1. Seventy individuals were included; two participants had both right and left knees assessed. A summary of the KL, OARSI JSN and UCOAG grades and WORMS composite scores is found in Table 6-2.

The second sample, of 75 knees with paired PA fixed-flexion radiographs, from baseline and the 30-month follow-up, is described in Table 6-1. Three individuals had both knees included. A summary of the radiographic grades and WORMS composite scores is found in Table 6-2.

Twenty nine knees out of 75 had the lateral TF compartment designated most-affected. Of these, eight showed definite progression of OA on the radiographs in the lateral TF compartment over 30 months. In the remaining 21 knees, there were either definite medial TF compartment changes, or very little change appreciated on the 30-month follow-up radiograph despite there being the required minimum amount of change according to the custom summed WORMS score. This was most likely because of the participant selection criteria. Knees with no discernible JSN and neutral alignment were defined as having the lateral TF compartment most-affected with OA at baseline. These 21 knees were assessed with the UCOAG scale for both the medial and lateral TF compartments. After the results were unblinded by MOST and the WORMS composite scores received, each of these images was reviewed. Only the TF compartment which changed the most on the WORMS scoring was included in the analyses.
Table 6-1: Description of participant samples [mean (standard deviation)]

<table>
<thead>
<tr>
<th></th>
<th>Concurrent Validity Sample</th>
<th>Sensitivity to Change Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of knees</td>
<td>72</td>
<td>75</td>
</tr>
<tr>
<td>Right : Left</td>
<td>40 : 32</td>
<td>46 : 29</td>
</tr>
<tr>
<td>Males : Females</td>
<td>38 : 34</td>
<td>22 : 50</td>
</tr>
<tr>
<td>Age (years)</td>
<td>63.2 (8.0)</td>
<td>62.3 (8.2)</td>
</tr>
<tr>
<td>Medial : Lateral(^1)</td>
<td>40 : 32</td>
<td>55 : 20(^2)</td>
</tr>
<tr>
<td>Body Mass Index (kg/m(^2))</td>
<td>29.7 (4.7)</td>
<td>30.2 (4.8)</td>
</tr>
<tr>
<td>WOMAC Physical Ability Score(^3) (maximum score 68)</td>
<td>15.6 (12.1)</td>
<td>14.6 (11.8)</td>
</tr>
<tr>
<td>WOMAC Knee Pain Score(^3) (affected knee, right or left, maximum score 20)</td>
<td>3.3 (2.9)</td>
<td>3.3 (3.1)</td>
</tr>
<tr>
<td>WOMAC Total Score(^3) (affected knee, right or left, maximum score 96)</td>
<td>20.9 (15.3)</td>
<td>19.6 (14.0)</td>
</tr>
<tr>
<td>20 m walk (average time of 2 trials, seconds)</td>
<td>16.6 (2.5)</td>
<td>16.5 (2.5)</td>
</tr>
<tr>
<td>5 chair stands (average time of 2 trials, seconds)</td>
<td>11.4 (4.7)</td>
<td>11.5 (4.3)</td>
</tr>
</tbody>
</table>

\(^1\) tibiofemoral compartment  
\(^2\) ratio after analysis of most-affected compartment at 30 months completed  
\(^3\) WOMAC – Western Ontario and McMaster Universities Arthritis Index
Table 6-2: KL, OARSI JSN and UCOAG grades and WORMS composite scores for concurrent validity and sensitivity to change.

<table>
<thead>
<tr>
<th></th>
<th>Concurrent Validity</th>
<th>Sensitivity to Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KL</td>
<td>OARSI JSN</td>
</tr>
<tr>
<td>Number of participants</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>Mean (standard deviation)</td>
<td>2.0 (1.6)</td>
<td>1.4 (1.2)</td>
</tr>
<tr>
<td>Range</td>
<td>0-4</td>
<td>0-3</td>
</tr>
<tr>
<td>Median</td>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>KL Baseline</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Follow-up</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>KL OARSI JSN Baseline</td>
<td>1.4 (1.2)</td>
<td>2.3 (1.3)</td>
</tr>
<tr>
<td>Follow-up</td>
<td>1.8 (1.4)</td>
<td>3.1 (2.0)</td>
</tr>
<tr>
<td>KL UCOAG Baseline</td>
<td>0.9 (0.9)</td>
<td>0.6 (0.8)</td>
</tr>
<tr>
<td>Follow-up</td>
<td>2.0 (2.0)</td>
<td>7.0 (7.0)</td>
</tr>
<tr>
<td>KL WORMS composite Baseline</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Follow-up</td>
<td>3.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

KL – Kellgren-Lawrence grading scale
OARSI JSN – Osteoarthritis Research Society International joint space narrowing grading scale
UCOAG – Unicompartmental osteoarthritis grading scale
WORMS – Whole organ magnetic resonance imaging score
If both compartments changed the same amount, the lateral TF compartment was used. This changed the proportion of knees with medial and lateral TF compartments most-affected from 46:29 to 55:20.

6.4.2 Concurrent Validity

Spearman’s rank correlations between the radiographic measures of knee OA severity and WORMS composite scores were high to very high (Figure 6-1 and Table 6-3). The confidence intervals overlapped considerably, showing that no scale was preferred. Correlations of OARSI JSN and UCOAG JSN grades to WORMS cartilage morphology scores were also very high, however correlations for the other UCOAG features were less robust.

Post-hoc analyses demonstrated no differences in Spearman’s rank correlations between right and left knees for any of the comparisons. Also there were no differences between medial and lateral TF compartments, with one exception. The Spearman’s rank correlation coefficient for the association of the UCOAG femoral osteophyte grade with the WORMS femoral osteophyte score was 0.62 (p < 0.0001) for the medial TF compartment and 0.37 (p = 0.0381) for the lateral TF compartment.

6.4.3 Sensitivity to Change

Spearman’s rank correlation coefficients for the sensitivity to change over 30 months for the knee OA radiographic grading scales relative to the WORMS composite score are presented in Table 6-4 and show moderate sensitivity to change. The wide confidence intervals that overlap
Figure 6-1: Radiographic grade plotted against the WORMS composite score for 72 knees with a range of osteoarthritis severity.

KL – Kellgren-Lawrence grading scale
OARSI JSN – Osteoarthritis Research Society International joint space narrowing grading scale
UCOAG – Unicompartmental osteoarthritis grading scale
WORMS – Whole organ magnetic resonance imaging score
**Table 6-3:** Spearman’s rank correlation coefficients (r) for concurrent validity of several methods of radiographic knee osteoarthritis assessment.

<table>
<thead>
<tr>
<th>Correlates</th>
<th>Spearman’s r</th>
<th>p-value confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>KL WORMS composite score</td>
<td>0.830</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.741 to 0.890</td>
</tr>
<tr>
<td>OARSI JSN WORMS composite score</td>
<td>0.835</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.747 to 0.893</td>
</tr>
<tr>
<td>UCOAG WORMS composite score</td>
<td>0.771</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.657 to 0.851</td>
</tr>
<tr>
<td>OARSI JSN WORMS cartilage morphology</td>
<td>0.817</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.722 to 0.882</td>
</tr>
<tr>
<td>UCOAG JSN WORMS cartilage morphology</td>
<td>0.827</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.736 to 0.888</td>
</tr>
<tr>
<td>UCOAG femoral osteophytes</td>
<td>WORMS femoral osteophytes</td>
<td>0.490</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.291 to 0.648</td>
</tr>
<tr>
<td>UCOAG tibial erosion</td>
<td>WORMS tibial bone attrition</td>
<td>0.634</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.472 to 0.755</td>
</tr>
<tr>
<td>UCOAG subluxation</td>
<td>WORMS meniscal extrusion</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.012 to 0.430</td>
</tr>
</tbody>
</table>

KL – Kellgren-Lawrence grading scale  
OARSI JSN – Osteoarthritis Research Society International joint space narrowing grading scale  
UCOAG – Unicompartmental osteoarthritis grading scale  
WORMS – Whole organ magnetic resonance imaging score
Table 6-4: Spearman’s rank correlation coefficients (r) for sensitivity to change over 30 months of several methods of radiographic knee osteoarthritis assessment.

<table>
<thead>
<tr>
<th>Correlates</th>
<th>Spearman’s r</th>
<th>p-value</th>
<th>confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>KL</td>
<td>WORMS composite score</td>
<td>0.479</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.283 to 0.637</td>
</tr>
<tr>
<td>OARSI JSN</td>
<td>WORMS composite score</td>
<td>0.515</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.326 to 0.664</td>
</tr>
<tr>
<td>UCOAG</td>
<td>WORMS composite score</td>
<td>0.473</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.275 to 0.632</td>
</tr>
<tr>
<td>OARSI JSN</td>
<td>WORMS cartilage morphology</td>
<td>0.425</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.219 to 0.594</td>
</tr>
<tr>
<td>UCOAG JSN</td>
<td>WORMS cartilage morphology</td>
<td>0.363</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.149 to 0.545</td>
</tr>
<tr>
<td>UCOAG femoral osteophytes</td>
<td>WORMS femoral osteophytes</td>
<td>0.260</td>
<td>0.0244</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.349 to 0.460</td>
</tr>
<tr>
<td>UCOAG tibial erosion</td>
<td>WORMS tibial bone attrition</td>
<td>0.331</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.113 to 0.519</td>
</tr>
<tr>
<td>UCOAG subluxation</td>
<td>WORMS meniscal extrusion</td>
<td>-0.388</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.565 to -0.177</td>
</tr>
</tbody>
</table>

KL – Kellgren-Lawrence grading scale
OARSI JSN – Osteoarthritis Research Society International joint space narrowing grading scale
UCOAG – Unicompartmental osteoarthritis grading scale
WORMS – Whole organ magnetic resonance imaging score
considerably suggest that no one scale was more sensitive to change than the others. Change over 30 months for the individual radiographic OA features was moderately associated with the corresponding change in WORMS features, although for the association of UCOAG subluxation and WORMS meniscal extrusion, the association was surprisingly negative ($r = -0.388$, $p = 0.0006$), which suggests that an increase in subluxation is moderately associated with a decrease in meniscal extrusion.

The SRMs for the KL, OARSI JSN and UCOAG grading scales were 0.96, 0.86 and 1.00 respectively, which show a large response to change for all of the radiographic grading scales.

### 6.5 Discussion

The KL, OARSI JSN and UCOAG grading scales were all highly or very highly associated with WORMS composite scores of articular damage due to knee OA. Furthermore, the three radiographic scales can all be considered equally correlated to the WORMS composite scale.

No radiographic grading scale for the assessment of knee OA severity has previously been correlated to the WORMS composite scale used in this study. However, KL grades have been correlated to cartilage defects [Spearman’s $r = 0.55$, $p < 0.01$; Pearson’s $r$ of up to 0.52 (medial femoral condyle), depending on location, $p < 0.05$] and cartilage volume [Pearson’s $r = -0.30$ (medial tibial cartilage) to -0.49 (lateral tibial cartilage) $p < 0.01$] as seen on MRI. The greater association observed in our study may be due to the inclusion of several selected knee OA features in our WORMS composite scale, which suggests that the KL grading scale is able to describe the presence of several articular features of knee OA. KL grades have also been
correlated to osteophytes observed on MRI (Pearson’s $r = 0.66$ in the medial TF compartment, $p < 0.05$) \textsuperscript{336}.

JSN assessed on a radiograph showed a strong association with the related WORMS feature of cartilage morphology. This was an expected finding, since articular cartilage makes up a considerable proportion of the joint space. The greatest amount of articular cartilage wear is often seen on the posterior aspect of the medial femoral condyle and the central aspect of the medial tibial plateau, because these regions experience the greatest load while walking \textsuperscript{75, 143}. These regions articulate in slight knee flexion, and therefore the extent of JSN is best appreciated with radiographic protocols that position the knee in flexion, such as the PA fixed-flexion protocol used in this study \textsuperscript{163}. The meniscus also contributes to the observed joint space, so meniscal subluxation or degeneration may contribute to the variance between the observation of JSN on fixed-flexion radiographs and cartilage morphology as seen on MRI \textsuperscript{143, 235, 236, 337}. While some feel that the meniscus does not contribute to the joint space in flexion, others have reported an association between JSN observed on radiographs taken in flexion and meniscal degeneration or extrusion \textsuperscript{143, 235, 236, 337}. While we could not find prior studies correlating ordinal values of JSN to MRI findings of TF OA severity, four studies correlated medial TF compartment joint space width (JSW) measured in millimetres (mm) with medial compartment cartilage volume, measured in mm\textsuperscript{3} (Spearman’s $r = 0.26$ to 0.46, $p < 0.01$) \textsuperscript{233, 338-340}.

The UCOAG has not previously been correlated with MRI findings; however the UCOAG total score performed similarly to the other radiographic scales. While the other grading scales demonstrated a “ceiling effect” when the severity of knee OA measured by the WORMS custom composite scale was between 12 and 18, the UCOAG grading scale did not, which suggests that it might continue to be sensitive to individuals with more severe presentations of
knee OA. Although composite scores are not meant to be broken down to assess the severity of knee OA on a radiograph, the UCOAG individual feature scores were correlated to corresponding WORMS OA features in order to explore the content validity of the UCOAG grading scale. There was a wide range of association. Surprisingly, the UCOAG JSN grades were more highly correlated to WORMS cartilage morphology scores than the UCOAG total score was to the WORMS composite score; however, this does not give a complete picture of OA change in the TF compartments. UCOAG JSN grades showed a very similar association to the WORMS cartilage morphology scores as OARSI JSN grades. UCOAG femoral osteophyte grades were moderately correlated to the WORMS femoral osteophyte scores and UCOAG tibial erosion grades were highly correlated to the WORMS tibial bone attrition scores. Because radiographs are a two-dimensional representation of the bony structure, osteophytes and bone erosion can often only be appreciated on the edges of the bones. Osteophytes which overlap may also not be appreciated. These differences from a three-dimensional MRI representation of the same bone may contribute to variance between the UCOAG femoral osteophytes and UCOAG tibial erosion grades and the corresponding WORMS scores. Although we expected to see a relationship between subluxation and meniscal extrusion, the UCOAG subluxation grades did not correlate significantly with the WORMS meniscal extrusion scores. It is likely that meniscal extrusion contributes more to JSN than to subluxation. On the other hand, ligament laxity and bony erosion may contribute more to subluxation and should be studied in the future, with the goal of understanding the causes of this deformity.

Correlations between radiographic and MRI OA features for left and right knees were the same, as expected. However, it was expected that radiographic OA features of the medial TF compartment, particularly JSN, would be more-highly associated to WORMS scores than those of
the lateral TF compartment, since the fixed-flexion radiograph protocol emphasizes the positioning of the medial tibial plateau parallel to the x-ray beam. While we did not find this difference for JSN, there was a large difference between the medial and lateral TF compartments for the correlation of the UCOAG femoral osteophyte grade to the WORMS femoral osteophyte score. Anecdotally, the readers reported that osteophytes on the lateral femoral condyle were much more difficult to appreciate than those on the medial condyle.

Changes seen on all three radiographic grading scales were moderately correlated with changes seen with the WORMS composite scale for progression of TF compartment OA severity over 30 months. We did not find previous studies that reported the correlation between ordinal measures of change in radiographic knee OA severity and MRI measures. However, several authors have tested the association between change in JSW measured from a radiograph and change in cartilage volume measured from MRI, and have determined that there was no correlation (Spearman’s rank correlation $r = -0.11$ and 0.19, $p > 0.05$). In a similar study, JSW was moderately associated with WORMS cartilage morphology of the whole knee (Spearman’s rank correlation $r = 0.41$, $p = 0.039$). Although continuous scales measuring JSW are often used for clinical trials of potentially disease-modifying OA drugs, we show that ordinal scales for JSN appear to have a higher association to MRI findings of articular cartilage degeneration than continuous scales. This finding is similar to that of Nevitt et al. and suggests that JSN could be a reasonable alternative for JSW as an outcome measure for change in TF compartment OA severity in these studies. Amin et al. reported a sensitivity of 23% and a specificity of 91% for OARSI JSN grades to detect progression in the WORMS cartilage morphology score in the medial TF compartment over 15 or 30 months.
SRMs are a commonly-used unit-less statistical method of comparing the sensitivity to change of different scales to each other\textsuperscript{334, 343}. While SRMs are most appropriate for continuous data, they have been used in the past for ordinal scales\textsuperscript{72, 145, 344}. The SRMs for all three radiographic scales were over 0.86, which is considered high. (All SRMs reported below are for worsening of OA severity; the sign reported in the original study could be positive or negative depending on the scale and research question however we have changed them to positive scores for comparison purposes.) SRMs previously reported for the KL grading scale were much lower (0.19 to 0.23 over one year, depending on the individual reader)\textsuperscript{344} while SRMs for ordinal scales of JSN were somewhat higher (SRM of 0.34 to 0.75 over one to three years with fixed-flexion images)\textsuperscript{72, 145, 344}. Continuous measurements of change in JSW (SRM of 0.15 to 0.75 over one to three years) showed similar results to JSN\textsuperscript{295, 345, 346}. SRMs tend to be higher in studies of individuals who already have knee OA\textsuperscript{295, 346} and also in studies with a longer follow-up time, since more change is presumed to have occurred\textsuperscript{239, 345}. Radiograph protocols that place the knee in flexion tend to produce higher SRMs as well\textsuperscript{345}. Our results might be somewhat magnified in comparison to previous studies since our participant sample contained only individuals whose knee OA had worsened by a minimum amount. Our “mean change” score, the numerator of the SRM equation, might have been greater than in studies which included individuals who did not change\textsuperscript{72, 344}.

Knee OA is a slowly progressing disease and not a lot of change in severity is expected over 30 months\textsuperscript{182}. For the sensitivity to change question we explicitly chose knees with at least a minimum amount of disease progression, as seen on MRI. Even so, the radiographic grading scales only identified 43 to 50 of the knees as having worsened (depending on the scale). Sensitivity scores were fairly low (57% for KL, 49% for OARSI JSN and 67% for UCOAG).
Ordinal scales, with levels from zero to three, four or even 13 are not always sensitive enough to pick up small amounts of change over 30 months.

The MOST database includes individuals who have, or are at risk for knee OA, and most have mild to moderate OA. Therefore, the full range of knee OA severity was not well represented in our participant samples. This attenuation might have not allowed the UCOAG grading scale to be fully evaluated throughout the full range of grades. Testing of the UCOAG grading scale on a cohort with more severe knee OA would be recommended. Also, as more data become available from MOST, future studies should include similar sensitivity to change analyses using data on individuals followed at five or more years, as sensitivity of radiographs to change in knee OA severity tends to increase with longer time-lines. The inclusion of individuals who have not changed would allow for the calculation of specificity as well.

Attempts were made by MOST to increase the sensitivity to change of the KL and OARSI JSN grading scales. An additional grade of 3.5 was added to the KL scale. If a knee showed no joint space (i.e. KL grade four) on a PA radiograph but some residual joint space on a lateral view, then a grade of 3.5 was assigned. In the participant sample for sensitivity to change grade 3.5 was only assigned once. If the modified definition of the KL grading scale was used, the correlation with the WORMS composite score was 0.473 (p < 0.0001); therefore the modified grading scale did not provide any increased sensitivity to change. Increased sensitivity to change was greater for the modified OARSI JSN grades. If increased JSN was evident but not enough to increase the score by a full grade, half-grades were given. In our sample, the OARSI JSN grade increased by a half-grade for nine knees. If the modified definition of the OARSI JSN grading scale had been used, the correlation with the WORMS
composite score would have been 0.563 (p < 0.0001) and the SRM would have been 0.98, showing a beneficial effect of the revisions with respect to sensitivity to change.

Accuracy and standardization of radiographic technique are very important to maximize sensitivity to change. Many have commented on the superiority of images taken in flexion compared to those taken in full extension. Some have found greater sensitivity to change when the anterior and posterior margins of the medial tibial plateau are seen parallel to each other, however, others have not noticed significant differences. Vignon et al. studied PA fixed-flexion radiographs and found that the SRM for JSW could range from 0.06 to 0.34 depending on the position of the medial tibial plateau. The MOST fixed-flexion protocol was closely standardized, with the legs placed in a frame that positions the feet a consistent distance apart and in 10° of external rotation. Up to three radiographs could be taken to ensure ideal observation of the medial TF compartment.

One limitation of this study might be the unusually high number of participants in sample one with designated lateral TF OA. This was because individuals with no JSN were determined to have lateral TF compartment OA if the HKA was neutral (i.e. 1° of varus to 1° of valgus) or valgus. The inclusion of individuals with neutral alignment in the lateral TF compartment category may have overinflated their numbers, given that in general approximately 22% of individuals with knee OA are affected primarily in the lateral TF compartment. One effect of this increase in knees with the lateral TF compartment designated most-affected was that the correlations between the radiographic grading scales and the WORMS composite score might have been attenuated because the radiographic protocol favours the assessment of OA features in the medial TF compartment.
A second limitation of this study was the participant selection criteria for the sensitivity to change question. Individuals were chosen based on a minimum amount of change on a WORMS-derived scale scored out of 164. The intent was that this score would give a global sense of the severity of the articular features in both TF compartments and would allow selection of a range of presentations of knee OA. We then correlated change in the three radiographic grading scales against a smaller WORMS composite score, which scored the articular features of OA only in the most-affected TF compartment. Unfortunately when the custom summed WORMS score had picked up “change”, this change was not always in the designated most-affected TF compartment. This occurred most often when there was no noticeable JSN at baseline and neutral alignment; this required that the designated most-affected compartment be changed. A participant selection score that focused on choosing a single TF compartment would have prevented this confusion.

Radiographs are less expensive, faster and require less expertise to perform and assess than MRIs; therefore they remain the standard for the evaluation of knee OA severity.\textsuperscript{135,346} We conclude that since the KL, OARSI JSN and UCOAG grading scales are all highly correlated to OA joint changes seen on MRI, these grading scales are equally valid and may be used in place of WORMS scores for evaluation of the severity of knee OA. Change in OA severity is also commonly assessed with radiographs.\textsuperscript{135} We conclude that since all three radiographic scoring methods are moderately to highly sensitive to change for knee OA severity over 30 months, they can all be used in place of WORMS scores to monitor OA progression. While some tissues seen only on MRI (for example ligaments, synovium and meniscus) might warrant its use for clinical and research purposes, we have shown that standardized PA fixed-flexion radiographs are
sufficient for the assessment of the articular features of mild to moderate TF OA, for clinical and research purposes.
Chapter 7

General Discussion and Future Perspectives

Knee osteoarthritis (OA) is an important health issue. The reported incidence of knee OA ranges from 5.4% to 38%, depending on the population studied\textsuperscript{6-13}, and that incidence is expected to increase as the population in western and eastern countries becomes older and more obese\textsuperscript{16-18}. The costs to Canada’s health care system are large\textsuperscript{38,40}. For example, 5156 total knee arthroplasties (TKA), a common surgical option for end-stage knee OA, were performed in Ontario in 1993/1994 while 11 488 were performed in 2003/2004\textsuperscript{352}, at an estimated cost of $21 000 each (in 2007) including medical care and the cost of the implant\textsuperscript{40}. Current treatments like TKAs are focused primarily on symptomatic relief, although research is ongoing to identify potentially disease-modifying OA drugs (DMOAD)\textsuperscript{353}. If knee OA can be discovered early in the course of the disease there might be greater potential for benefit from interventions such as DMOADs, but also conservative treatments such as weight reduction, bracing and orthotics to change joint loading, strengthening and other exercise\textsuperscript{134,241,354-356}. The evaluation of knee OA needs to be effective, simple and cost effective, since large numbers of individuals will need to be evaluated in the future.

Magnetic resonance imaging (MRI) of the knee is becoming more common for the assessment and monitoring of knee OA for research purposes, particularly for large multicentre studies\textsuperscript{2,97,255}. Even so, MRI is still limited in its clinical use as a diagnostic tool for knee OA, primarily because of its high cost, time constraints and perceived lack of need\textsuperscript{255}. Therefore the assessment of knee OA severity seen on radiographs remains the standard for the clinical diagnosis and monitoring of knee OA. While radiographs are commonly assessed with grading
scales, there are concerns that currently-used scales do not recognize the full spectrum of presentations of knee OA. Therefore, there is a need for a composite scale that includes several features of knee OA. Because knee OA is associated with frontal-plane alignment, the monitoring of alignment can assist in the early detection of knee OA, or highlight individuals at risk. There is a need for a way to measure frontal-plane alignment that is accurate, simple and able to be used in a clinic setting. The overall objective of this thesis was to evaluate tools which may be used to assess for knee OA risk and to monitor the severity and progression of the disease, for clinical decision making and research purposes.

7.1 Estimation of the Hip-Knee-Ankle Angle Using Knee Radiographs

Varus and valgus alignment of the lower extremities (LE) is probably associated with knee OA onset and definitely associated with knee OA progression. It is important that LE alignment in measured accurately, so that interventions can be prescribed appropriately, and research studies which include LE alignment can be compared to one another. In Chapter 3, the ability of the femoral shaft-tibial shaft (FS-TS) angle measured from knee radiographs to estimate the hip-knee-ankle (HKA) angle was evaluated. The FS-TS angle is commonly used for this purpose; however some have argued that the association between the two angles is not strong enough to substitute the FS-TS angle for the HKA angle. Two factors which might influence this relationship are the type and degree of varus or valgus deformity as well as the length of the femoral and tibial shafts used when calculating the FS-TS angle. The purpose of the study presented in Chapter 3 was to determine if the relationship between the HKA and FS-TS angles changed depending on the type and magnitude of frontal-plane LE deformity,
and the femoral and tibial shaft lengths used to determine the FS-TS angle. The lengths of the long-bone shafts visible on a typical knee radiograph were also determined.

One hundred and twenty full-length LE radiographs with a wide range of varus and valgus frontal-plane alignment were selected from the Multicenter Osteoarthritis Study (MOST) database. The HKA angle and five versions of the FS-TS angle (determined using the full length of the long bone shafts, and two thirds, one half and one third of the length of the shafts, and 10 cm shafts) were calculated. The mean offset between the HKA and FS-TS angles was -5.0°; however, varus limbs had a greater offset while valgus limbs had a smaller offset. Therefore, depending on the individual or population, it would be inaccurate to always use -5.0° as the offset. The Pearson’s correlation between the HKA angle and the full-shaft FS-TS angle was high; however the correlations between the HKA angle and the shorter-shaft FS-TS angles were smaller. Since only one-third of the long bone shafts were visible on the knee radiographs, we recommended that the HKA be used when an accurate estimation of LE alignment is required.

These results have implications for clinical and research purposes. While estimates of the HKA angle using the FS-TS angle may be adequate for screening purposes, and to monitor change in individuals over time, when an accurate picture of LE alignment is required (for example, in surgical planning for TKA), the HKA angle measured from a full-length radiograph is the only valid choice. While large, multi-centre research studies routinely include full-length LE radiographs, smaller studies often do not. The results of studies where estimates of LE alignment were made using knee radiographs may be inaccurate if alignment is an important part of the research question. When comparing the results from studies which used the FS-TS angle to estimate the HKA angle, the type and degree of malalignment of the participants must be taken into account, in order to estimate the offset required to estimate the HKA angle using the
appropriate regression equation. To illustrate, two studies investigated the relationship between LE alignment and knee OA progression\textsuperscript{83,85}. Brower et al.\textsuperscript{83} used the FS-TS angle to measure LE alignment, adding a 4° offset towards valgus. They reported the odds ratio (OR) for OA progression from Kellgren-Lawrence (KL) grade two to grade three or four, in individuals with valgus alignment compared to those with neutral alignment, to be 1.39 (p > 0.05) and the OR for individuals with varus alignment compared to those with neutral alignment to be 2.90 (p < 0.05)\textsuperscript{83}. Cerejo et al.\textsuperscript{85} determined the HKA angle with full-length radiographs and reported the OR for the same degree of OA progression associated with valgus alignment as 2.46 (p > 0.05) and the OR for OA progression associated with varus alignment as 4.12 (p < 0.05). While both sets of results report progression of knee OA to be related to varus but not valgus alignment, the magnitude of the ORs is considerably different. Some of this difference could be related to the different methods of measuring LE alignment.

### 7.2 Estimation of the Hip-Knee-Ankle Angle Using Pelvis-to-Ankle Photographs

Determining effective, simple and cost effective methods of evaluating alignment in the clinical setting is valuable, as large numbers of individuals will require evaluation in the coming years. Because photography is fast, inexpensive, readily accessible and does not require exposure to ionizing radiation, it may be an ideal alternative to radiography to estimate the HKA angle. Prior research has only tested this possibility on healthy young individuals. Therefore, the purpose of the study presented in Chapter 4 was to assess the reliability and validity of the HKA angle determined from pelvis-to-floor photographs (HKA-P) for the estimation of the HKA angle.
determined from full-length LE radiographs in individuals with a range of ages and body mass index (BMI) scores representative of the general population.

Fifty participants were assessed with one full-length LE radiograph and two pelvis-to-floor photographs, taken 30 minutes apart. The HKA angle was calculated from each radiograph. The range of HKA angles was from -8.9° (varus) to 7.4° (valgus). Points were chosen to estimate the knee and ankle joint centres and a proximal femoral point on a photograph. Several possible points were selected and the points that provided estimated HKA angles with the highest correlations to the actual HKA angle were chosen. Using these points, HKA-P angles were calculated for each photograph. Three readers assessed each photograph from the first testing session twice, at least two weeks apart, and one reader assessed the photographs from the second testing session. The intra-rater, inter-rater and test-retest reliability of the HKA-P angle were very high and the Pearson’s correlation between the HKA angle and the HKA-P angle was also very high. Therefore we can confidently recommend that the HKA-P angle be used to estimate the HKA angle in individuals with up to moderate degrees of malalignment, keeping in mind that the HKA-P angle is an average of 4.5° more varus than the HKA angle, using the recommended points for the knee, ankle and proximal femur.

The results presented in Chapter 4 are important, as they support the use of photographs for the estimation of the HKA angle. Because photography is fast, inexpensive, readily accessible and does not require exposure to ionizing radiation this technique is ideal for screening and monitoring purposes. It may also be applied to individuals for whom radiation exposure is contra-indicated (pregnant women, individuals with cancer or serious health issues, or those subjected to repeated ionizing radiation exposure). The photographic technique presented here can be performed by physiotherapists and other health professionals, in their offices or clinics.
without delay or extra costs to the health care system. The required tools are easily accessed and the resulting image can be archived for longitudinal comparisons.

In order to fully recommend the HKA-P angle to monitor change over time, an evaluation of sensitivity to change should be performed. The change in the HKA-P angle measured from photographs taken several years apart should be compared to the change in the HKA angle measured from LE radiographs over the same period of time. Another useful study would be to repeat the comparison of the HKA-P angle to the HKA angle in children. The HKA-P angle assessed from a photograph is ideally suited for children because of the avoidance of ionizing radiation, and could be used to assess and monitor changes in frontal-plane alignment which occur as children age and to identify individual children at risk for malalignment. Replication of the study with other specific populations such as those with severe knee OA, severe varus or valgus deformity, or those with obesity would also be useful, to ensure that the techniques used, especially the determination of the hip, knee and ankle points, are applicable to the populations that we are most interested in monitoring frontal-plane alignment. As a preliminary analysis, the Bland-Altman plot for the comparison of the HKA and HKA-P angles did not show heteroskedasticity, indicating that the relationship between the two angles was consistent within the range of alignment presented in the participant sample (-8.9° to 7.4°). See Figure 4-6.

The HKA-P angle may be compared to the FS-TS angle for the estimation of the HKA angle. The FS-TS angle measured with 10 cm-long long-bone axes is approximately 5° more valgus than the HKA angle, while the HKA-P angle is approximately 4.5° more varus than the HKA angle and these offsets must be allowed for. Currently the FS-TS angle is accepted in the literature as providing a valid estimate of the HKA angle; however, the research presented

160
in Chapter 3 suggests that this may not be the case. The high levels of intra-rater, inter-rater and test-retest reliability for the HKA-P and the high correlation of the HKA-P with the HKA, along with the simplicity of the acquisition technique suggest that it might be a preferred option for the estimation of the HKA angle. We emphasize that if precise information is required on an individual (for example, for surgical planning) the HKA angle should be determined from a full-length LE radiograph. That said, a photograph may be adequate to monitor the progression of frontal-plane deformity at the knee.

7.3 Psychometric Properties of the Unicompartmental Osteoarthritis Grade for the Assessment of Tibiofemoral Osteoarthritis Severity on a Radiograph

The UCOAG grading scale was created by Cooke et al. in 1999 as a means of grading the severity of knee OA visualized on radiograph. At that time, while inter-rater reliability was reported as excellent when used on knee radiographs taken in full extension, the full psychometric properties of the UCOAG grading scale were not assessed. Therefore, the goal of the studies presented in Chapters 5 and 6 was to study the reliability, validity and sensitivity to change of the UCOAG grading scale and to make comparisons to other commonly-used grading scales.

For the studies evaluating the UCOAG grading scale, samples were selected from the MOST database. To investigate intra-rater and inter-rater reliability (Chapter 5), 100 posteroanterior (PA) fixed-flexion radiographs with a range of knee OA severity were selected. Osteoarthritis Research Society International (OARSI) joint space narrowing (JSN) grades and the HKA angle were used to determine whether the medial or lateral TF compartment was most-
affected by OA and to ensure that the proportions were consistent for each of four OA severity levels. Three readers applied the UCOAG grading scale to each radiographs twice, with at least two weeks between readings. Intra-rater reliability was described as substantial to excellent and inter-rater reliability was described as moderate to excellent, depending on the analysis used. These levels of reliability are similar to or better than those reported for other OA severity grading scales.

To evaluate the test-retest reliability of the UCOAG (also Chapter 5), a second sample of 100 radiograph pairs was selected from those individuals in the MOST database whose TF compartment OA had not changed over 15 or 30 months in terms of severity, as determined using an MRI-based score for change in cartilage morphology. Participants with a range of OA severity were selected as for the intra- and inter-rater reliability analyses, and the ratio of medial and lateral TF compartment OA was preserved in each of four severity strata. One reader applied the UCOAG grading scale to the 200 randomized radiographs. Test-retest reliability was described as substantial to excellent depending on the analysis used. The minimal detectable change (MDC95) in TF OA was 2.61, suggesting that a change of three UCOAG grades would indicate real change in the severity of TF OA. We could not find any previous research which measured the MDC for knee OA radiographic grading scales. A change of one level on the Kellgren-Lawrence (KL) grading scale is commonly used in the literature to indicate a change in OA severity. However, we could not find supporting documentation for this claim. Because the UCOAG has 13 levels of severity, a MDC95 of 2.61 appears clinically reasonable.

To study the validity of the UCOAG (Chapter 6), a third sample of 72 radiographs was selected according to the same criteria used for the reliability studies presented in Chapter 5. Three different grading scales for the severity of TF OA were applied to the radiographs: the
UCOAG, the KL and the OARSI JSN. The results from each grading scale were correlated to a custom MRI-based ordinal scale that assessed the severity of several articular features of knee OA. The most-affected TF compartment was assessed with each scale, except for the KL scale, where the entire knee was graded. Individual OA features were also compared when appropriate. The correlations between the radiographic grading scales and the MRI-based scale were all high to very high, and very similar to each other, suggesting that all three grading scales were valid choices for the assessment of TF OA severity on a radiograph. The correlations between JSN and the corresponding MRI OA feature of cartilage morphology were also very high, although the associations between the other features of the UCOAG and the corresponding OA features seen on MRI were lower.

To evaluate the sensitivity to change of the UCOAG scale (also Chapter 6), a fourth sample of 75 radiograph pairs recorded from individuals who showed a specified minimum level of change in OA severity between the baseline and 30-month follow-up images was selected. Changes in OA severity were determined using a customized score derived from MRIs of the same knees. All three radiographic grading scales and the same MRI-based scale used for validity assessment were applied to the most-affected TF compartment for each radiograph pair. Correlations between the change in OA severity assessed with each radiographic grading scale and the change assessed with the MRI-based scale showed a moderate sensitivity to change. The standardized response mean (SRM) is a unit-less measure which can be used to compare different scales. SRM results for all three grading scales showed a “large” response to change, indicating a low level of variability in the change scores relative to the mean amount of change.

From the results presented in Chapters 5 and 6, the UCOAG grading scale can be confidently recommended for the assessment of TF OA severity on a PA fixed-flexion
radiograph. The UCOAG compares favorably with existing scales, with high reliability and similar validity and sensitivity to change. Further research on the UCOAG scale should include validity and sensitivity to change analyses with a cohort of individuals with moderate to severe TF compartment OA, in order to assess the function of the UCOAG scale at the upper end of its limit of 13. When the KL and OARSI JSN scales had reached their “ceiling” level the UCOAG grades were between four and nine, suggesting that the UCOAG would be more sensitive than these other scales to presentations of severe OA. Further investigation can be done on the validity of the four features included in the UCOAG scale, to determine what MRI-assessed OA features are most-highly correlated with each UCOAG feature. Since meniscal extrusion was not a significant contributor to subluxation, the WORMS features of cartilage morphology, ligament damage and bony erosion should be correlated to the UCOAG feature of subluxation to investigate the causes of deformity. Because the UCOAG scale was created to be sensitive to change in alignment, another suggestion for future research is the investigation of change in UCOAG grades relative to change in alignment. The predictive validity of the UCOAG with respect to its ability to predict the need for total knee arthroplasty could also be studied. Finally, the correlation of the UCOAG scale to measures of pain and function should be investigated. While the association of these measures to the KL and OARSI JSN grading scales is low and mostly not statistically significant, it is possible that the association with the UCOAG scale might be higher because it includes more features of knee OA. Finally, the UCOAG should also be modified and tested for use on other joints, in particular the patellofemoral and hip joints.
7.3.1 Standardization of Radiography Protocols

The importance of using standardized radiographic procedures has been stressed by several authors. Standardization is important to ensure accuracy, sensitivity to change and to enable research studies to be compared. It is particularly important in longitudinal studies where sequential radiographs are taken over many years. The more recent radiographic protocols, such as the semi-flexed, Lyon schuss, metatarsophalangeal (MTP) and fixed-flexion protocols, all describe standardized LE positioning, including knee flexion, LE rotation and equal weight-bearing. The use of positioning frames such as the SynaFlexer™ (Synarc, San Francisco, California) and foot maps also aids in consistent positioning. The reader who assessed the 75 radiograph pairs for the UCOAG sensitivity to change analysis observed very little change in LE rotation between baseline and follow-up images, highlighting the consistency achieved by MOST with the fixed-flexion protocol over 30 months. This consistency enables the radiographic grading scales to be as sensitive as possible to change over time.

Because joint space width (JSW) is often used to measure change in TF OA over time, measurement of this OA feature is particularly important to standardize. To accurately measure JSW, the anterior and posterior margins of the medial tibial plateau should be superimposed, and lie parallel to the x-ray beam. This occurs with the knee in some flexion. Fluoroscopy, a prescribed x-ray beam angle and repeat imaging may be used to position the medial tibial plateau accurately and reliably. Unfortunately, with the emphasis on visualizing the medial tibial plateau during radiographic imaging, if the anterior and
posterior margins of the lateral tibial plateau were not also superimposed and parallel, the grading results for the lateral TF compartment might not be as accurate or sensitive to change as those for the medial TF compartment. Our validity analysis did not find a difference in the correlation of the medial and lateral TF compartment UCOAG grades with the MRI-based scores for the same TF compartment, or the correlation of the medial and lateral TF compartment UCOAG JSN grades with the MRI cartilage morphology scores for the same compartment. However it is possible that a measure of JSW, which is a continuous measure, might reveal differences between the medial and lateral TF compartments.

Application of the radiographic grading scales in a standardized fashion is important as well. Periodic assessments of a reader’s performance should be performed, with re-training done as necessary. In the case of multi-centre studies, it is critical that the reading technique be identical at each centre. Guermazi et al. suggest that because inter-reader variability, even between experts, is consistently greater than intra-reader variability, ideally all of the radiographs should be read by one reader. While our intra-rater reliability results were substantial to excellent, and our inter-rater reliability results were moderate to excellent, it remains essential to maintain vigilance with respect to training and adherence to the grade level descriptions to ensure ongoing reliability.

7.3.2 Most-Affected Tibiofemoral Compartment, Medial or Lateral?

The definition of the “most-affected” TF compartment used by MOST caused some confusion in the UCOAG studies. MOST stipulated that the TF compartment with the greatest OARSI JSN grade be designated most-affected. For knee radiographs with equal medial and
lateral OARSI JSN grades, if the HKA angle was neutral or valgus (i.e. greater than -1°) the lateral TF compartment was designated most-affected. This follows prior research suggesting that the risk of medial TF JSN was decreased in individuals with neutral and valgus frontal-plane alignment ⁹¹. For the UCOAG reliability studies, the readers were not told which compartment MOST considered to be the most-affected. The three readers agreed on the most-affected compartment 97% to 99% of the time, confirming the consistency with which the UCOAG grading scale was applied. However, Table 7-1 shows that for participant samples one and two that for 10% to 12% of the sample MOST designated the lateral TF compartment as the most-affected while the readers designated the medial TF compartment. For a smaller percent of the time (1% to 3%) MOST designated the medial TF compartment as the most-affected while the readers designated the lateral compartment.

For the validity and sensitivity to change studies, MOST provided information on which TF compartment was designated as most-affected so that the radiographic grading scales would all be applied to the same compartment. The reader did notice, however, that for some of the images the opposite TF compartment appeared more affected, at least with respect to the application of the UCOAG grades. For the sensitivity to change study, as the paired radiographs were being assessed with the UCOAG, the reader noted that for 28% of the image pairs, the TF compartment opposite to the one identified by MOST as the most-affected had obviously changed, or both TF compartments appeared unchanged. Therefore the decision was made to assess both TF compartments and the compartment with the greatest change in the MRI-based score was kept as the most-affected. If both TF compartments had changed equally then the lateral TF compartment was designated most-affected. This led to a change in the proportion of medial and lateral TF compartments designated most-affected from 46:29 to 55:20 which meant
Table 7-1: Frequency of radiographs with medial and lateral tibiofemoral compartments designated as “most-affected”, by the Multicenter Osteoarthritis Study (MOST) and by the readers.

MOST used a combination of joint space narrowing and the hip-knee-ankle angle to determine the “most-affected” compartment while the readers used the unicompartmental osteoarthritis grade (UCOAG).

a) Participant sample 1 (for intra-rater and inter-rater reliability).

<table>
<thead>
<tr>
<th>Readers</th>
<th>MOST</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>medial</td>
<td>lateral</td>
<td>total</td>
<td></td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>67</td>
<td>12</td>
<td>79</td>
</tr>
<tr>
<td>lateral</td>
<td></td>
<td>3</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>70</td>
<td>27</td>
<td>97</td>
</tr>
</tbody>
</table>

b) Participant sample 2 (for test-retest reliability)

<table>
<thead>
<tr>
<th>Readers</th>
<th>MOST</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>medial</td>
<td>lateral</td>
<td>total</td>
<td></td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>71</td>
<td>10</td>
<td>81</td>
</tr>
<tr>
<td>lateral</td>
<td></td>
<td>1</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>72</td>
<td>25</td>
<td>97</td>
</tr>
</tbody>
</table>
that the distribution of damage relative to compartment no longer matched that described for the sample after selection.

Therefore, the definition of “most-affected TF compartment” should be reconsidered. It is obvious that some knees that are neutral at baseline can progress to medial TF OA and some to lateral TF OA. The designation of knees with neutral alignment to only the lateral TF compartment category, as done by MOST, may be somewhat misleading. It is likely that this definition has arbitrarily increased the percentage of knees with designated lateral TF compartment OA in the MOST database. Merle-Vincent et al. 348 made a similar conclusion for radiographs showing knees with mild OA, and found that the radiographic protocol impacted the distribution of medial and lateral TF compartment OA. Some studies have included knees with 1° of valgus to 1° of varus as a separate “neutral” category 35, 53, 91. This approach might be more appropriate for the MOST database, where many individuals are at risk for knee OA but do not actually have radiographic changes, at least at baseline and therefore should not be included in the lateral TF compartment group. In future studies, if the designation of a most-affected TF compartment is required and radiographs do not show greater JSN in one compartment then images with neutral alignment should not be categorized as at risk for OA at all.

7.3.3 Natural History of Tibiofemoral Osteoarthritis and its Relationship to Radiographic Grading Scales

The development of knee OA has been described to progress in a stereotypical fashion. Osteophytes develop first (although small osteophytes are not necessarily precursors of the disease 197) and JSN occurs subsequently 144. The presence of osteophytes has traditionally been
considered the best method of defining the presence of radiographic knee OA \textsuperscript{73,144}. The KL scale, with its initial emphasis on osteophytes and inclusion of definite JSN only at grade three has exploited this common presentation of OA \textsuperscript{81,170}. On the other hand, the progression of JSN is associated with OA progression \textsuperscript{74} and so JSN scales and JSW scores are often used to monitor progression \textsuperscript{22,317-320}. We found that the KL, OARSI JSN and UCOAG grading scales were equally well-correlated to the MRI-based scores of TF OA, and also equally sensitive to change. Therefore, all three scales would be equally suitable for identifying TF OA and monitoring change over time.

7.4 Concluding Statements

Because of the prevalence of knee OA in the population of older adults, the assessment of OA risk, incidence, and progression is a priority, to enable monitoring and early treatment. The assessment of LE frontal-plane malalignment, which is a significant risk factor for OA progression, is typically performed by measuring the HKA angle on a full-length radiograph. The FS-TS angle measured from knee radiographs has been used in the past but we have shown that this is not recommended. On the other hand, pelvis-to-floor photographs show promise for the estimation of the HKA angle.

Grading of knee radiographs in terms of the severity and progression of knee OA has traditionally been performed with the KL and OARSI JSN scales, amongst others. We have shown that the UCOAG grading scale is a viable alternative, with moderate to high reliability, high validity and moderate sensitivity to change.
Chapter 8

References


172


(97) OsteoArthritis Initiative, a knee health study. 2011. Coordinating Center, University of California San Francisco.


179


183


Vignon E. Radiographic issues in imaging the progression of hip and knee osteoarthritis. *J Rheumatol Suppl* 2004;70:36-44.


Dempster WT. Space requirements of the seated operator. Dayton, OH: Wright-Patterson Air Force Base; 955 Jan.


Middel B, van SE. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care* 2002;2:e15.


(305) Bland JM, A. How can I decide the sample size for a study of agreement between two methods of measurement? John Martin Bland 2004 December 1;Available at: URL: http://www-users.york.ac.uk/~mb55/meas/sizemeth.htm.


(308) Project 15 Test-retest reliability of semi-quantitative readings from knee radiographs. Osteoarthritis Initiative; 2012 Apr 16.


(332) Bruynesteyn K, Van Der HD, Boers M et al. Detecting radiological changes in rheumatoid arthritis that are considered important by clinical experts: influence of reading with or without known sequence. *J Rheumatol* 2002;29(11):2306-2312.


Appendix A
Approvals from Multicenter Osteoarthritis Study for Chapter 3

MEMO #0947

December 22, 2006

To: Lisa Sheehy

From: Jean Hietpas and Michael Peterson

Re: Approval of Ancillary Study Proposal AS06-03 by Lisa Sheehy entitled

"Comparison of Mechanical Axis and Anatomical Axis Measurements on Lower Limb X-rays"

Cc: MOST Executive Committee

AS06-03 Co-Investigators

Congratulations; your modification proposal for the ancillary study AS06-03 entitled "Comparison of Mechanical Axis and Anatomical Axis Measurements on Lower Limb X-rays" has been conditionally approved by the MOST Executive Committee.

Please provide the committee with a revised ancillary study proposal addressing the conditions, comments, and recommendations provided below and a brief cover memo describing the changes.

Conditions of approval:
Please submit a revised proposal addressing the analytical questions raised by Reviewers #1 and #2.

Please submit a budget addressing the funding issues raised by Reviewer #3.

Comments and Recommendations:

Reviewer #1:

This is a very interesting proposal addressing important questions about knee alignment that can be well answered with data from MOST. Several points need clarification:

1) Will the anatomic alignment measures be taken from the full limb radiograph or the fixed flexion radiograph? Since the full limb is obtained in full extension, using it will limit the generalizability of the study since the current standard for knee radiographs in OA studies is to obtain a flexed view. If it is necessary to obtain a separate extended view for anatomic alignment this would reduce its advantage over getting a full limb. This study should be designed to tell us whether the fixed flexion view can be used to measure anatomic axis. This may require obtaining anatomic alignment data from both the full limb and the fixed flexion view and comparing the relationship between each and mechanical axis.

2) RQ1. The most important question would seem to be a) How much variation is there in the offset between the alternative views and measures? b) What are the determinants of that variation? and c) How accurately can we predict mechanical axis from anatomic axis (from a regression model) and what characteristics are associated with inaccurate prediction?

3) The repeated measures ANOVA to determine if there is an offset between the two measures seems trivial. There will be. The more important question is what is the offset and how accurately can mechanical be predicted from anatomic axis and other participant and skeletal characteristics?

4) RQ2. The "ideal anatomic axis" needs to be defined.
Reviewer #2:

1) Shouldn't the variation in marking the femoral and tibial shaft positions be done first, i.e. currently research question #2. How will the "best" FSTS angle be determined?

2) How will the dominant leg and severity of knee OA effect be considered in the analysis?

Reviewer #3:

The proposal states: The image assessment and analysis would therefore cost US $800, based on a rate of US $20 an hour and is requested from the Multicenter OA study.” There is no specific estimate given to cover any costs for data management or analysis, including costs of providing the investigators with an analytic dataset of MOST variables on the participants included in their analysis. MOST as an entity has no monetary resources to support ancillary studies; resources reside at the individual centers. Ancillary studies by definition are conducted using funds obtained outside of the study. Funding needs to be clarified.

2) There appears to be nothing in the original agreement with the PI of the Laxity and Alignment Study that would preclude use of the images for other purposes.

Please note that MOST ancillary study investigators are required to follow the policies governing ancillary studies and publications as posted on the MOST study website. Please contact Michael Peterson (mpeterson@psg.ucsf.edu, 415-514-8178) for website access permissions. As stated in the Ancillary Study Guidelines: "The Executive Committee must review and approve a draft of the funding application and budgets prior to submission. This should be in the hands of the Executive Committee at least 4 weeks prior to the submission deadline to allow time for review and revisions." Please contact MOST Project Director, Jean Hietpas (hietpas@psg.ucsf.edu, 415-514-8089), if you have questions about the policies or procedures.
We look forward to working with you on this ancillary study. As you move forward, please keep us informed of the status and progress of the study.

Thank you.
MEMO #1808

October 30, 2009

To: Lisa Sheehy

From: MOST Publications Committee

Re: Review of MOST Analysis Plan Proposal by Lisa Sheehy (AP09-03)
entitled “Does Measurement of the Anatomic Axis Consistently Predict Hip-Knee-Ankle Angle (HKA) for Knee Alignment Studies in Osteoarthritis?”

Cc: John Lynch
Charles McCulloch
Jingbo Niu
Yuqing Zhang
Jean Hietpas
Peggy Rasmussen

Congratulations! Your analysis plan proposal entitled “Does Measurement of the Anatomic Axis Consistently Predict Hip-Knee-Ankle Angle (HKA) for Knee Alignment Studies in Osteoarthritis?” (AP09-03) has been approved by the MOST Publications Committee. The following comments were submitted by reviewers.

Reviewer Comments/Recommendations:

205
Reviewer #1:

I think that this study will collect a lot of useful data regarding the relationship between limb alignment (HKA) angle from full limb radiographs and from the FS-TS angle measured from semi-flexed knee radiographs. By addressing the effects of different methods for placing landmarks for FS-TS, the proposal will be able to address when and how HKA differs from FS-TS. It will be particularly interesting to find out whether there are particular situations (i.e.: types of knees/participants) where HKA and FS-TS angle do not agree well. That data could be usefully applied to other studies such as OAI.

The statistical methods for analyzing the data, and determining the degree of agreement (and whether it is poor, or good agreement) seem slightly vague, but since Yuqing Zhang is an author on the proposal, I feel confident that the statistical methods used will be appropriate.

Reviewer #2:

This is a very well conceived plan. The abstract findings are interesting. Please consider the possibility that there may be value in assessing anatomic alignment independently of trying to use it to estimate HKA and the value of AA needs to be determined in studies of its association with outcomes of knee OA (e.g. Felson 2009 publication comparing HKA and AA).

Abstract Development and Review: Development of your abstract must follow instructions in the Publications Guidelines, Section J. Abstracts must be circulated among your co-authors prior to the abstract submission deadline. Also submit the draft abstract to MOSTPublications@psg.ucsf.edu.

<<Publications Guidelines v1.4_01.16.09.pdf>>

Abstract Approval: When co-author and Publications Committee recommendations have been incorporated, obtain final approval from the senior MOST investigator co-authoring your abstract. Submit a signed copy of the enclosed Abstract/Poster/Presentation Approval Form to MOSTPublications@psg.ucsf.edu with a copy of the final abstract and an email or tracking form confirming that you submitted the abstract.

<<AbstractPosterPresent Approval form v1.5_05.05.09.pdf>>
Thank you for submitting a MOST analysis plan and good luck with your analyses!
Appendix B
Letter of Information and Consent, and Ethics Approval for Chapter 4
Letter of Information
& Consent Form

**Standardized standing pelvis-to-floor photographs for the assessment of mechanical alignment of the lower extremity.**

Student Principal Investigator: Lisa Sheehy, PT, MSc  
Supervisors: Dr. Linda McLean, PhD; Dr. Elsie Culham, PhD  
Advisor: Dr. Derek Cooke, MD

You are being invited to participate in a research study directed by Lisa Sheehy to investigate whether we can adequately assess the alignment of the legs using a photograph. This is usually measured with a full-length x-ray of the leg and the two methods will be compared.

Your participation is entirely voluntary, and participating or choosing not to participate will in no way affect any present or future medical treatment or health care that you may require. This consent form will be reviewed with you and the procedures described in detail. Please feel free to ask questions at any time. You will be given a copy of this form to take home. This study has been approved by the Queen’s University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board.

**Purpose of the Study:**

Many people are bow-legged or knock-kneed; these deformities are risk factors for developing knee osteoarthritis (OA) in the future and for the progression of osteoarthritis. To measure these deformities, leg alignment is ideally determined by drawing lines on an x-ray of the whole leg with one line from the hip to the knee and another from the knee to the foot. The resulting angle is then measured. We wish to see if leg alignment can be accurately estimated from a photograph of the whole leg. If alignment measured from a photograph is reliable (i.e. we get almost the same result over and over again) and gives a very similar angle to alignment measured from an x-ray, then photographs may be used by physiotherapists and physicians to identify individuals at
risk for knee OA and who might benefit from therapeutic interventions like exercise, orthotics, bracing or surgery. Photographs might also be used to monitor change in alignment over time. While other groups have investigated photographs used in this way, they tested them only on young, healthy individuals with a limited range of leg alignment and body type. We are recruiting individuals of all ages and leg alignments, including those with prior knee injury or OA.

**Screening and Exclusions:**

Adults of all ages and with any leg alignment are eligible for this study. You must be able to stand for at least 20 minutes without assistance to be in the study. Because an x-ray will be taken, you will not be accepted for this study if you are pregnant or have a serious illness like cancer. Individuals who have had a recent traumatic injury to the knees or ankles will also not be accepted because swelling may make it difficult to identify landmarks on the joints.

**Testing Procedures:**

You will be required to come to the Kingston General Hospital Department of Diagnostic Radiology for a single testing session, which should last for approximately 90 minutes. First you will be asked these questions: your age, sex, height, weight and any history of injury or surgery to your legs, hips, knees and ankles.

You will then be asked to undress from the waist down, but leave your underwear on. A sheet will be provided for modesty. Sticky dots will be placed on your skin on both sides of your pelvis, the centre of the pelvis, and both hips, knees and ankles. You will then stand on a small raised platform with your legs straight and knees facing to the front. An x-ray and a photograph will be taken of you in this position. If the x-ray is poor you will be asked for permission to have another one taken.

After these are done, the dots will be removed and you will change into your street clothes. A $4 coupon for a snack will be provided and you will be given a 30-minute break. There are no restrictions as to what you do during this break. After 30 minutes you will return, undress again and the sticky dots will be replaced. A second photograph will be taken, but in a different, nearby room.

**Risks of Participation:**

There are no risks to answering the questions. There are minimal risks to having a photograph taken; the greatest risk is tripping on or falling from the platform, which will be approximately
eight inches (20 cm) above the ground. You will be assisted on to and off of the platform. The required standing position should not cause discomfort.

This research study involves exposure to radiation from an x-ray of the pelvis and both legs. This radiation exposure is not necessary for your medical care and is for research purposes only. The total amount of radiation that you will receive in this study is about 0.82 millisieverts (mSv), and is approximately equivalent to a uniform whole body exposure of 100 days of exposure to natural background radiation. This use involves minimal risk and is necessary to obtain the research information desired.

There may be psychological risks involved in learning about being bow-legged or knock-kneed. These deformities of alignment do not mean that you will experience pain in the future. If you experience any side-effects from the testing, please let any of the research investigators know. Their contact numbers are at the end of this form.

Benefits of Participation:

There are no direct benefits from participating in this research. You will have a chance to participate in an area of investigation that has not previously been thoroughly investigated. If you wish to see your photographs you can do so. If you wish to obtain your alignment results you can provide contact information to the principal investigator for this purpose only. Your contact information will be destroyed once the results have been sent out. Results will be available within a few months.

Confidentiality:

All data collected in the course of this study is strictly confidential and your anonymity will be protected at all times. Photographs will not include your head or face. Photographs, x-rays and subject information will be identified using coded subject numbers only, not your name or birthdate. Data will be kept on the principal investigator’s computer and as the code will be kept separate in a locked cupboard, there will be no way for you to be identified based on the stored data. This computer is password-protected and files will only be available to the principal investigator and to designated research assistants (Mary Lucas and Mike Brean). The computer is backed-up regularly. Any hard copies or handwritten data will be kept in a locked file cabinet. Any reports, thesis, journal submissions, presentations, or posters that use the data from this study will not use the names of any of the research participants. All investigators involved in this study are trained to keep personal information confidential and safe. The identifying code will be destroyed after five years.

Voluntary Nature of the Study:
Your participation in this study is completely voluntary and participating or choosing not to participate will in no way affect any present or future medical treatment or health care that you may require. You may withdraw from this study at any time without penalty or coercion. Your data will be removed if you wish them to be withdrawn.

**Liability:**

In the event that you are injured as a result of the study procedures, medical care will be provided to you until resolution of the medical problem. By signing the consent form, you do NOT waive your legal rights nor release the investigators from their legal and professional responsibilities.

**Payment:**

You will receive $20 to compensate for your time spent while participating in this study. In addition, you will be given a $4 coupon for a snack in the Hospital coffee shop while you have your break. Parking, in the L.D. Acton Building parking lot on the Queen’s University campus, will be provided for free. If you choose to take the bus, your costs (2 bus tickets; $4.30) for this will be reimbursed.

**Subject Statement and Signature:**

As a volunteer participant, I have read and understand the information on this letter of information and consent form for this study. The purposes, procedures and technical language have been explained to me. I have been given sufficient time to consider the information and to withdraw if I choose to do so. I have had the opportunity to ask questions which have been answered to my satisfaction. I understand that I can withdraw at any time. I understand that my participation is in confidence to the researchers only and that my data will be used for scientific purposes only. I am voluntarily signing this consent form and will receive a copy of the form for future reference.

If I am dissatisfied with any aspect of the study, or have questions, concerns or adverse events, I am encouraged to contact the principal investigator or her faculty supervisors:

Principal Investigator  Lisa Sheehy  4lms@queensu.ca  (613) 744-6517

Faculty Supervisor and  Dr. Linda McLean  mcleanl@queensu.ca  (613) 533-6101
Chair, Graduate Programme (Rehabilitation Science)
If I have questions regarding my rights as a research subject I can contact:

Dr. Albert Clark, Chair, Queen’s University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board at (613) 533-6081.

By signing this consent form, I am indicating that I agree to participate in this study:

_________________________________  ____________________________
Signature of Subject                  Date

_________________________________
Signature of Person Conducting Consent Process  Date

By signing this consent form, I confirm that I have carefully explained the nature of the above research study to the subject. I certify that, to the best of my knowledge, the subject understands clearly the nature of the study and the demands, benefits, and risks involved to participants in this study.

_________________________________
Signature of Principle Investigator  Date
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD-DELEGATED REVIEW
March 23, 2012

Ms. Lisa Sheehy
School of Rehabilitation Therapy
Queen's University

Dear Ms. Sheehy
Study Title: REH-511-12 Standardized standing pelvis-to-floor photographs for the assessment of mechanical alignment of the lower extremity.
File # 6066638
Co-Investigators: Dr. L. McLean, Dr. E. Culham, Dr. T.D.V. Cooke

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol, internal review, subject recruitment process, subject recruitment advertisements and information/consent form for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following listing of ethics requirements you must fulfill over the course of your study.

Reporting of Amendments: If there are any changes to your study (e.g., consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. Please use event form: WESEB Multi-Use Amendment/Full Board Renewal Form associated with your post-review file # 6066638 in your Researcher Portal (https://services.queenu.ca/researcher/)

Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days of the event or 15 days after becoming aware of the event. All other serious adverse events must be reported within 15 days of becoming aware of the information. Serious Adverse Event Forms are located with your post-review file # 6066638 in your Researcher Portal (https://services.queenu.ca/researcher/)

Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the context information for the Research Ethics Board.

Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]

Chair, Research Ethics Board
March 23, 2012

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD

The membership of this Research Ethics Board complies with the membership requirements for Research Ethics boards and operates in compliance with the Tri-Council Policy Statement, Part C, Division 5 of the Food and Drug Regulations, OHRP, and US DHHS Code of Federal Regulations Title 45, Part 46 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalwide Assurance Number: FWA00004184, HIRB0000173

Current 2012 membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board:

Dr. A.F. Clark, Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)
Dr. H. Abdallah, Professor, Department of Medicine, Queen's University
Dr. R. Britton, Professor, Department of Emergency Medicine, Queen's University
Dr. M. Evans, Community Member
Dr. S. Horgan, Manager, Program Evaluation & Health Services Development, Geriatric Psychiatry Service, Providence Care, Mental Health Services, Assistant Professor, Department of Psychiatry
Ms. J. Hudacik, Community Member
Mr. B. McNaughton, Community Member
Ms. P. Newman, Pharmacist, Clinical Care Specialist and Clinical Lead, Quality and Safety, Pharmacy Services, Kingston General Hospital
Dr. W. Racine, Emeritus Professor, Department of Pharmacology & Toxicology, Queen's University
Ms. S. Redmond, Privacy Officer, ICES-Queen's Health Services Research Facility, Research Associate, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute
Dr. B. Stinchcombe, Assistant Professor, Department of Anaesthesia and Perioperative Medicine, Queen's University
Dr. A.N. Singh, WHO Professor in Psychosomatic Medicine and Psychopharmacology
Professor of Psychiatry and Pharmacology, Chair and Head, Division of Psychopharmacology, Queen's University, Director & Chief of Psychiatry, Academic Unit, Queen's Health Care, Belleville General Hospital
Dr. E. Tai, Associate Professor, Department of Pediatrics and Office of Bioethics, Queen's University
Dr. E. Van Den Bergh, Professor, School of Nursing and Department of Anesthesiology and Perioperative Medicine, Queen's University

215
Appendix C
Approvals from Multicenter Osteoarthritis Study and Ethics Approval for Chapter 5

MEMO #2495
January 23, 2012

To: Lisa Sheehy
From: MOST Executive Committee
Re: Approval of Modified MOST Ancillary Study Proposal by Lisa Sheehy (AS11-01) entitled “Reliability, Validity and Sensitivity to Change of the Uni-Compartmental OsteoArthritis Grading Scale (UCOAG)”

Congratulations; your modified MOST ancillary study proposal (AS11-01) entitled “Reliability, Validity and Sensitivity to Change of the Uni-Compartmental OsteoArthritis Grading Scale (UCOAG)” has been approved by the MOST Executive Committee. The following comment was submitted by the MOST Executive Committee.

Reviewer Comments/Recommendations:
These are all good changes and should improve the study. Nice work!

The approved ancillary study is enclosed for your documentation.

<<AS11-01 Sheehy Modified Ancillary Prop_Uni-Comp OA 01.03.12_Approved.pdf>>

As you proceed with the ancillary study, be sure to follow required policies governing ancillary studies and publications. You’ll find the Ancillary Studies Guidelines (Version 1.2, June 2010) and Publications Guidelines (Version 1.6, March 2011) on the MOST study website (www.keeptrack.ucsf.edu). As your ancillary study requires the release of MOST images to you, it is required that you complete a Request for MOST Research Image Set and Data Use Agreement for Research Image Set. We will send these documents to you for completion in a follow-up email. Please send an email to MOSTCoordinatingCenter@psg.ucsf.edu if you have any questions.

<<Ancillary Study Guidelines v1.2_06.18.10.pdf>> <<MOST Publications Guidelines v1.6_03.18.11.pdf>>
Now that the ancillary study is approved, the MOST Publications Committee is looking forward to the submittal of your revised analysis plan proposal (AP11-10). Please send your revised proposal and questions/comments to MOSTPublications@psq.ucsf.edu. We look forward to working with you on this ancillary study. As you move forward, please keep us informed of the status and progress of the study.

Thank you.
MEMO #2512

February 9, 2012

To: Lisa Sheehy

From: MOST Publications Committee

Re: Approval of Revised MOST Analysis Plan by Lisa Sheehy (AP11-10) entitled “Uni-Compartmental Osteo Arthritis Grading (UCOAG)"

Cc: MOST Publications Committee

Congratulations! Your revised analysis plan proposal entitled “Uni-Compartmental Osteo Arthritis Grading (UCOAG)” (AP11-10) has been approved by the MOST Publications Committee.

Reviewer Comments/Recommendations:

Reviewer #1:

Nice work on the revisions. Thank you.

Reviewer #2:

Not reviewed by Reviewer #2. If the reviewer responds with comments or recommendations, they will be forwarded to you at a later time.

The approved analysis plan is enclosed for your documentation.

<<AP11-10 Sheehy Revised AP Proposal_Uni-compart OA_Approved.pdf>>

Analysis: Please contact Yuqing Zhang (yuqing@bu.edu) or Jingbo Niu (niujp@bu.edu) at the BU Analysis Center to discuss your analysis plan.
Abstract Development and Review: Development of your abstract must follow instructions in the Publications Guidelines, Section J. Abstracts must have co-authors from each MOST grant (BU, UAB, U-Iowa and UCSF) and be circulated among your co-authors for review at least 10 working days prior to the meeting abstract submission deadline. Also submit the draft abstract to MOSTPublications@psg.ucsf.edu at least 10 working days prior to the abstract deadline. The draft abstract will be posted on the study website for optional MOST Publications Committee review.

<<MOST Publications Guidelines v1.6_03.18.11.pdf>>

Abstract Approval: When co-author and Publications Committee recommendations have been incorporated, obtain final approval from David Felson, the senior MOST investigator co-authoring your abstract. Submit a signed copy of the enclosed Abstract/Presentation Approval Form to MOSTPublications@psg.ucsf.edu with a copy of the final abstract and an email or tracking form confirming that you submitted the abstract to the meeting.

<<MOST AbstractPosterPresent Approval form v1.7_11.08.11.pdf>>

Including MOST Online Slide When Presenting Results from your Analyses: Whenever you have an opportunity to give a podium presentation of the results of your analyses, please include a slide with information about MOST Online (http://most.ucsf.edu), the public data sharing website. The NIH requires that federally-funded datasets be available to the public and that information about how to access these datasets is widely distributed. Therefore, we ask that you include the information on the enclosed MOST Online presentation slide when giving podium presentations of MOST data.

<<Sample MOST presentation_2011.ppt>>

Thank you for submitting a MOST analysis plan and good luck with your analyses.
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD-DELEGATED REVIEW
November 03, 2011

Ms. Lisa Sheehy
School of Rehabilitation Therapy
Queen's University

Dear Ms. Sheehy,

Study Title: REH-507-11 Uni-Compartmental Osteo Arthritis Grading (UCOAG) analysis project: Question 1 - Reliability
File # 6006389
Co-Investigators: Dr. T.D.V. Cooke, Dr. D. Felson, Dr. J. Niu, Dr. L. McLean, Dr. E. Culham

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol, budget, approval from MOST Committee, data use agree with MOST for use of radiographs for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following listing of ethics requirements you must fulfill over the course of your study:

Reporting of Amendments: If there are any changes to your study (e.g. consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. Please use our online forms (https://research.queensu.ca/research/ethics/amendment) associated with your protocol file # 6006389 on the Research Portal.

Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information. Serious Adverse Event forms are located with your protocol file # 6006389 on the Research Portal (https://research.queensu.ca/research/ethics).

Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]

Chair, Research Ethics Board
November 03, 2011

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD

The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards as defined by the Tri-Council Policy Statement, Part C, Division 5 of the Food and Drug Regulations, CHIRP, and U.S. DHHS Code of Federal Regulations Title 45, Part 66 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalwide Assurance Number: #FWA/00004184, HHRB00001173

Current 2011 membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board:

Dr. A.F. Clark, Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)

Dr. H. Abdollah, Professor, Department of Medicine, Queen's University

Dr. R. Biron, Professor, Department of Emergency Medicine, Queen's University

Dr. M. Evans, Community Member

Dr. S. Horgan, Manager, Program Evaluation & Health Services Development, Geriatric Psychiatry Service, Providence Care, Mental Health Services, Assistant Professor, Department of Psychiatry

Dr. B. S. Khilevsky, Professor, School of Nursing, Department of Psychology and Obstetrics & Gynaecology, Queen's University

Ms. B. Morales, Community Member

Ms. P. Newman, Pharmacist, Clinical Care Specialist and Clinical Lead, Quality and Safety, Pharmacy Services, Kingston General Hospital

Dr. W. Rees, Emeritus Professor, Department of Pharmacology & Toxology, Queen's University

Ms. S. Reithland, Privacy Officer, ICES-Queen's Health Services Research Facility, Research Associate, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute

Dr. B. Simchen, Assistant Professor, Department of Anaesthesia, Queen's University

Dr. A.N. Singh, WHO Professor in Psychosomatic Medicine and Psychopharmacology

Professor of Psychiatry and Pharmacology, Chair and Head, Division of Psychopharmacology, Queen's University, Director & Chief of Psychiatry, Academic Unit, Quinte Health Care, Belleville General Hospital

Dr. E. Turk, Associate Professor, Department of Pediatrics and Office of Bioethics, Queen's University

Rev. J. Warren, Community Member
Memo #2495
January 23, 2012

To: Lisa Sheehy
From: MOST Executive Committee
Re: Approval of Modified MOST Ancillary Study Proposal by Lisa Sheehy (AS11-01) entitled “Reliability, Validity and Sensitivity to Change of the Uni-Compartmental OsteoArthritis Grading Scale (UCOAG)”

Congratulations; your modified MOST ancillary study proposal (AS11-01) entitled “Reliability, Validity and Sensitivity to Change of the Uni-Compartmental OsteoArthritis Grading Scale (UCOAG)” has been approved by the MOST Executive Committee. The following comment was submitted by the MOST Executive Committee.

Reviewer Comments/Recommendations:
These are all good changes and should improve the study. Nice work!

The approved ancillary study is enclosed for your documentation.

As you proceed with the ancillary study, be sure to follow required policies governing ancillary studies and publications. You'll find the Ancillary Studies Guidelines (Version 1.2, June 2010) and Publications Guidelines (Version 1.6, March 2011) on the MOST study website (www.keeptrack.ucsf.edu). As your ancillary study requires the release of MOST images to you, it is required that you complete a Request for MOST Research Image Set and Data Use Agreement for Research Image Set. We will send these documents to you for completion in a follow-up email. Please send an email to MOSTCoordinatingCenter@psg.ucsf.edu if you have any questions.

<<Ancillary Study Guidelines v1.2_06.18.10.pdf>> <<MOST Publications Guidelines v1.6_03.18.11.pdf>>
Now that the ancillary study is approved, the MOST Publications Committee is looking forward to the submittal of your revised analysis plan proposal (AP11-10). Please send your revised proposal and questions/comments to MOSTPublications@psq.ucsf.edu. We look forward to working with you on this ancillary study. As you move forward, please keep us informed of the status and progress of the study.

Thank you.
MEMO #2591

June 4, 2012

To: Lisa Sheehy

From: MOST Publications Committee

Re: Approval of MOST Analysis Plan by Lisa Sheehy (AP11-21)

entitled “Uni-Compartmental OsteoArthritis Grading (UCOAG):
Validity and Sensitivity to Change, A Comparison of Three
Scales for the Radiographic Assessment of Knee OA”

Cc: MOST Publications Committee

Congratulations; your analysis plan proposal entitled “Uni-Compartmental OsteoArthritis Grading (UCOAG): Validity and Sensitivity to Change, A Comparison of Three Scales for the Radiographic Assessment of Knee OA” (AP11-21) has been approved by the MOST Publications Committee. Although a revised version of the analysis plan is not required, please take into consideration the following reviewer comments.

Reviewer Comments/Recommendations:

Reviewer #1:

Nice plan. Very thorough and well thought out. A few comments:

1. Why just femoral, and not also tibial, osteophytes are used?

2. Sample size calculation is not clear. You mention an effect size of 0.30. Is this a correlation coefficient? If not, what does it represent (i.e. what groups are you comparing for differences and what is the dependent variable)? Also, isn’t the research question about whether the UCOAG has a higher correlation with WORMS than KL? What is your power for detecting differences in correlations?
3. You need to keep in mind that by selecting knees based on strata of WORMS score, some knees will not have OA at baseline.

4. For the UCOAG measurements, will you use the same algorithm as described on page 4 paragraph 2 for determining which compartment to assess? If JSN is 0 in both compartments, is a degree or two of varus or valgus a valid indicator of compartment involvement?

5. One of the potential advantages suggested for the UCOAG is better discrimination of early OA vs. KL. So you probably should consider not combining KL grades 0-1 in analyses.

6. Note that JSN and OST grades at baseline are all full grades (no partials).

7. For the change analyses, how will you ensure that there are enough knees in the sample with changes in KL grade to give adequate power for the comparison?

The approved analysis plan is enclosed for your documentation.

<<AP11-21 Sheehy AP Validity Uni-Compartmental OA Grading 05.15.12 Approved.pdf>>

Analysis: Please contact Jingbo Niu (niujp@bu.edu) or Yuqing Zhang (yuqing@bu.edu) at the BU Analysis Center to discuss your analysis plan.

Abstract Development and Review: Development of your abstract must follow instructions in the Publications Guidelines, Section J. Abstracts must be circulated among your co-authors for review at least 10 working days prior to the meeting abstract submission deadline. Also submit the draft abstract to MOSTPublications@psg.ucsf.edu. The draft abstract will be posted on the study website for optional MOST Publications Committee review.

<<MOST Publications Guidelines v1.6_03.18.11.pdf>>

Abstract Approval: When co-author and Publications Committee recommendations have been incorporated, obtain final approval from David Felson, the senior MOST investigator co-authoring your abstract. Submit a signed copy of the enclosed Abstract/Presentation Approval Form to MOSTPublications@psg.ucsf.edu with a copy of the final abstract and an email or tracking form confirming that you submitted the abstract to the meeting.
Include MOST Online Information When Presenting Results from your Analyses: Whenever you have an opportunity to give a podium presentation of the results of your analyses, please include a slide with information about MOST Online (http://most.ucsf.edu), the public data sharing website. The NIH requires that federally-funded datasets be available to the public and that information about how to access these datasets is widely distributed. Therefore, we ask that you include the information on the enclosed MOST presentation slide when giving podium presentations of MOST data.

<<Sample MOST presentation_2011.ppt>>

Thank you for submitting a MOST analysis plan and good luck with your analyses.
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD-DELEGATED REVIEW
October 11, 2012

Ms. Lisa Sheehy
School of Rehabilitation Therapy
Queen's University

Dear Ms. Sheehy,

Study Title: REH-540-12 Uni-Compartmental Osteo Arthritis Grading (UCOAG) analysis project: Question 2 - Validity and Sensitivity to Change
File # 6097404

Co-Investigators: Dr. J. Nia, Dr. T.D.V. Cooke, Dr. D. Felson, Dr. E. Culham, Dr. L. McLean

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol, budget, approvals from MOST for proposal and analysis plan and the data use agreement for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following listing of ethics requirements you must fulfill over the course of your study:

- **Reporting of Amendments**: If there are any changes to your study (e.g., consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. Please use event form HSERB Multi-Use Amendment/Full Board Renewal Form associated with your post-review file # 6097404 in your Researcher Portal (https://services.queensu.ca/researcher)

- **Reporting of Serious Adverse Events**: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information. Serious Adverse Event forms are located with your post-review file 6097404 in your Researcher Portal (https://services.queensu.ca/researcher)

- **Reporting of Complaints**: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

**Annual Renewal**: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]

Chair, Research Ethics Board
October 11, 2012

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD

The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards and operates in compliance with the Tri-Council Policy Statement, Part C Division 5 of the Food and Drug Regulations, OHRP, and U.S. DHHS Code of Federal Regulations Title 45, Part 46 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalside Assurance Number: #FWA00004184, 01RB0001173

Current 2012 membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board:

Dr. A.F. Clark, Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)

Dr. H. Abdullah, Professor, Department of Medicine, Queen's University

Dr. B. Berison, Professor, Department of Emergency Medicine, Queen's University

Dr. C. Cline, Assistant Professor, Department of Medicine, Director, Office of Biobanks, Queen's University, Clinical Ethicist, Kingston General Hospital

Dr. M. Evans, Community Member

Dr. S. Horgan, Manager, Program Evaluation & Health Services Development, Geriatric Psychiatry Service, Providence Care, Mental Health Services, Assistant Professor, Department of Psychiatry

Ms. J. Hudac, Community Member

Dr. B. Kislinsky, Professor, School of Nursing, Departments of Psychology and Obstetrics and Gynecology, Queen's University

Dr. J. MacKinnon, Pediatric Geneticist, Department of Pediatrics, Queen's University

Mr. D. McNaughton, Community Member

Ms. P. Newman, Pharmacist, Clinical Care Specialist and Clinical Lead, Quality and Safety, Pharmacy Services, Kingston General Hospital

Ms. S. Rohland, Privacy Officer, ICES-Queen's Health Services Research Facility, Research Associate, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute

Dr. R. Simchson, Assistant Professor, Department of Anesthesiology and Perioperative Medicine, Queen's University

Dr. J. Tang, Medical Resident, Department of Emergency Medicine, Queen's University