ESTIMATION OF PREFERENCE-BASED MEASURES OF HEALTH FROM DISEASE-SPECIFIC CLINICAL OUTCOME MEASURES FOR TOTAL HIP AND KNEE ARTHROPLASTY PATIENTS

by

Susan Marie Odum

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Health Services Research

Charlotte

2013

Approved by:

_____
Dr. Jennifer Troyer

_____
Dr. William P. Brandon

_____
Dr. Jacek Dmochowski

_____
Dr. Sarah Laditka

_____
Dr. Rob Roy McGregor

ABSTRACT

SUSAN MARIE ODUM. Estimation of preference-based measures of health from disease-specific clinical outcome measures for total hip and knee arthroplasty patients. (Under the direction of DR. JENNIFER TROYER)

Transfer to utility (TTU) or mapping methodology allows researchers to estimate a health utility from a disease-specific measure and calculate quality adjusted life years for economic evaluations. The purpose of this study was to develop regression algorithms to map five common disease specific TJA outcome measures to three preference-based health utility scores. An online survey was completed by 438 total hip arthroplasty (THA) patients and 550 total knee arthroplasty patients (TKA). THA patients completed the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC®), Harris Hip Score (HHS), and the Hip Disability and Osteoarthritis Outcomes Score (HOOS). Knee patients completed the WOMAC®, Knee Society Score (KSS), and Knee Disability and Osteoarthritis Outcomes Score (KOOS). All patients completed three preference based questionnaires, the SF-6D, EQ-5D and HUI-3, and responses were used to calculate health utilities. A total of 30 THA mapping models and 30 TKA mapping models were developed and validated. Forecast error measures including ME, MAE, RMSE were defined as our prediction performance criterion. For the THA models, the regression model with HOOS subscores most precisely estimated an EQ-5D health utility. The best performing TKA model mapped the KSS to the EQ-5D. Clinician-researchers can input their disease specific data into these models to estimate health utilities to consider the cost-effectiveness of osteoarthritis-related interventions relative to interventions for very different diseases and conditions.

DEDICATION

I dedicate this dissertation to my parents, Jim and Darlene Odum, for their enduring love and support. They have inspired me and motivated me my entire life. This dissertation is dedicated in loving memory of my father. Completion of this doctoral program would not be possible without the love and support of Dolcie Peake.

ACKNOWLEDGMENTS

Many thanks go to my dissertation chair, advisor and mentor, Jennifer Troyer. Her commitment to my work is very much appreciated. I would also like to thank my committee members for their ongoing support: William Brandon, Jacek Dmochowski, Sarah Laditka and Rob Roy McGregor.

I would like to acknowledge the University of North Carolina at Charlotte Belk College of Business for the grant award that covered the expenses for the license fees required to conduct this research. I would like to thank the OrthoCarolina Research Institute for providing access to potential study participants and access to the Ortech software. Additionally, I am grateful for the assistance of Jeff Guerin and Sean Volkhart, of Ortech, for their help with the online survey and database support. Finally, I would also like to thank the orthopedic surgeons, Doug Dennis and Ray Kim as well as the staff at Colorado Joint Replacement Institute, Kendall Slutzky and Donnis Rafferty, for all of their help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AQoL	Assessment of Quality of Life
CJR	Colorado Joint Replacement
CEA	cost-effective analysis
CUA	cost utility analysis
EQ-5D	EuroQol-5D
HHS	Harris Hip Score
HRQOL	health-related quality of life
HOOS	Hip Disability and Osteoarthritis Outcomes Score
HUI-3	Health Utilities Index 3
ICF	International, Classification of Functioning, Disability and Health
IWQOL-lite	Impact of Weight on Quality of Life-Lite
KOOS	Knee Injury and Osteoarthritis Outcomes Score
KSS	Knee Society Score
MAE	mean absolute error
ME	mean error
NN	neural network
OA	osteoarthritis
OC	OrthoCarolina
QALYs	quality-adjusted life years
QOL	quality of life
RMSE	root mean square error
SD	standard deviation

SF-6D	Short Form-6D
SG	standard gamble
THA	total hip arthroplasty
TKA	total knee arthroplasty
TJA	total joint arthroplasty
TTU	transfer to utility
WHO	World Health Organization
WOMAC®	Western Ontario and McMaster Universities Osteoarthritis Index

INTRODUCTION

Osteoarthritis (OA) of the hip and knee is a degenerative disease that affects nearly 5 million people in the United States (U.S).[1] When conservative treatments fail, total joint arthroplasty (TJA) is the gold standard treatment.[1] The United States annual economic burden of TJA procedures ranges from approximately $3 billion to $6 billion.[1] As the economic burden of TJA is expected to increase it is important to consider the cost-effectiveness of current, as well as, new technology.[2-5]

Pain due to progressive osteoarthritis is a primary reason for performing a total joint arthroplasty.[6] A primary benefit of total joint arthroplasty is restoration of function related to activities of daily living, as well as, recreational activities.[6] Therefore, pain and function are primary outcomes of interest in total joint arthroplasty research. Furthermore, long-term and longitudinal clinical outcomes are of substantial interest to patients, surgeons and manufacturers of total joint implants. When assessing clinical outcomes of TJA, it is common to use disease specific health-related quality of life (HRQOL) measures[6-8] such as the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC®)[9-11], Knee Society Score (KSS)[12,13], Harris Hip Score (HHS)[14], Hip Disability and Osteoarthritis Outcomes Score (HOOS)[15-17] and Knee Injury and Osteoarthritis Outcomes Score (KOOS).[17-19] Generic, preference-based measures of health include the EuroQol-5D (EQ-5D)[20,21], the Health Utilities Index 3 (HUI-3)[21] and the Short Form-6D (SF-6D). [22,23] Table 1 illustrates the subscales for each of these HRQOL instruments. Preference-based instruments allow the researcher to administer a multi-dimensional set of questions regarding overall health to the patient. Based on prior research regarding societal preferences for the health states covered under the

combinations of questionnaire responses, algorithms can be applied to the patient responses to construct a measure a single health index score for each patient ranging from 0 (death) to 1 (perfect health).[24] The health index score represents one's health status at that given point in time and reflects societal preferences for the combination of health attributes identified by the patient. This link to preferences has resulted in the index being referred to as a health utility by economists.[24] The health utility value is used to calculate quality adjusted life years (QALYs), which is the effectiveness measure for cost-effective analysis (CEA).[24,25] QALYs are the amount of time spent in a health state multiplied by the health utility score given to that health state. Figure 1 illustrates a simple example of these calculations.[25] While disease specific measures are clinically useful, they do not provide a preference based measure of health utility necessary for cost-utility analyses.

The purpose of this study is to develop regression algorithms using a transfer to utility (TTU) method to map five common disease specific TJA outcome measures to preference-based health utility scores derived from patient responses to three different multi-dimensional health measurement instruments. The preferred regression models will be compared with outcomes from neural network analysis.

Literature Review

Health Related Quality of Life and Cost-effective Analysis

Conceptual Model. The World Health Organization's (WHO) International Classification of Functioning, Disability and Health (ICF), is a framework for describing health and health-related states across a broad range of diseases and health conditions.[26] The ICF conceptual model (Figure 2) integrates medical and social aspects of health and

provides a global framework for defining and measuring health and disability.[26] The premise of the model is that disability and function are the outcomes of the interaction between a health condition as well as environmental and personal contextual factors.[26] Functioning refers to all body functions, activities and participation, while disability is encompasses activity limitations and participation restrictions.[26] Function is measured at three levels including the specific body part, the whole person and the whole person in a societal context.[26] Environmental contextual factors include the structural characteristics of one's living environment as well as the climate and terrain of the one's external environment.[26] Personal factors include demographic characteristics as well as social support and psychological beliefs of health.[26]

Cost-effective Analysis. Interest in considering the cost-effectiveness of medical care has been growing, as health care cost growth continues to outpace inflation in the U.S. In general, cost-effective analysis (CEA) involves estimating the added costs associated with a new medical intervention relative to the improvements in a stated health objective attributable to the intervention. Cost utility analysis (CUA) is a type of cost effective analysis that assesses the value of an intervention with respect to quantity and quality of life.[3,5,24] Quality-adjusted life years (QALYs) is a common, generic health outcome measure in a CUA that allows comparisons across all areas of healthcare interventions. [3,24,25] Therefore, information obtained through a CUA can be used as a tool to guide healthcare decisions with respect to cost and effectiveness. [3,24,25]

Calculating Health Utilities. Health utility is an individual's preference value for a given health state or health outcome.[3,24] Health utilities range from zero (death) to one (perfect health).[3,24] The health utility is used to calculate QALYs, which is the amount of

time spent in a health state multiplied by the utility score given to that health state. [3,24,25] Health utility measures are designed to reflect population preferences for different health states on a large number of health dimensions.[3,24] This allows for comparisons of gains in health related quality of life across a variety of patient populations, diseases and intervention types.[3,24]

Health utilities can be calculated using direct measurement and indirect measurement methods.[3,24] Two common direct methods include the standard gamble approach and the time trade-off approach.[3,24] The standard gamble approach involves presenting individuals with a choice of a compromised health state with a certain probability compared to one better health state and one worse health state both with an uncertain probability.[3,24] The standard gamble approach is a classic way of measuring preferences in economics under conditions of uncertainty and is consistent with standard models of utility maximization. For example, individuals are presented with a choice between a certain state of chronic knee pain compared to a treatment that has an uncertain outcome – it could result in perfect health or it could result in death.[24] They are then asked to determine what probability of perfect health would make them indifferent to remaining in their current, certain state of chronic knee pain.[24] If they decide they are indifferent at a 0.7 probability of perfect health and a 0.3 probability of death, the health utility level for an individual with chronic knee pain is 0.7.[24] The time trade-off method, which is based on value theory, requires individuals to decide how many years of life they would be willing to sacrifice to avoid a certain compromised health state.[3,24] For example, individuals are asked to contemplate how many years of life they are willing to sacrifice to avoid a certain health state of chronic knee pain. For instance, an individual

might be told that they have a choice between ten additional years of life with chronic knee pain or a shorter lifetime without chronic knee painIf one is indifferent between ten years of chronic knee pain and seven years without chronic knee pain, the utility level for an individual with chronic knee pain is 0.7.[24] There is no psychometric evidence suggesting that individuals make health care decisions in this precise way. However, individuals often make choices where they trade off a lower-valued but certain payoff for a higher but uncertain expected payoff (standard gamble); in addition, individuals indicate a willingness to trade off time for increased health. The standard gamble and the time trade-off are two methods that direct people's attention and force them to decide between the presented options. Because people are inherently risk averse, the standard gamble tends to yield higher utility values than the time trade-off method. In other words individuals will choose a higher probability value of perfect health to avoid a higher probability of death.[24,27]

Generic, preference-based instruments, such as the EQ-5D, measure health utility using an indirect method.[24] Each generic, preference-based instrument includes a variety of attributes that are valued by large population samples using direct measurement methods such as the standard gamble approach.[24] Finally, a scoring algorithm is developed to generate unique weighted health states.[24] When disease specific measures have been collected and are available but generic preference-based measures are not available, a mathematical technique called transfer to utility, or mapping, is an option to obtain a health utility when needed for a CUA.

Cost Utility Analysis and Total Joint Arthroplasty. Total joint arthroplasty is the most widely studied orthopedic surgical procedure.[3,4] This field has experienced

tremendous growth with substantial advances in medical technology and total joint implant design.[3,4] While the efficacy and effectiveness research is important for clinical decision making, economic evaluations, such as cost utility analysis, are paramount in determining the true societal value of such advances.[3,4] A well designed cost utility analysis requires a societal perspective, accurate utility measures, discounting of health costs and accurate medical costs.[3,4] The associated medical costs should encompass all indirect medical costs, direct medical costs, opportunity costs and projected medical costs.[3,4]

In a 2004 review of cost-utility analysis, Bozic et al.[4] reported that of the 116 cost-utility analyses published between 1976 and 2001, only 37 were orthopedic related. Of the 37 orthopedic related CUAs, 11 were associated with total joint arthroplasty.[4] All of the TJA studies reported that the procedure was cost-effective with ratios below the threshold value of $50,000 per QALY.[4] However, the methodological quality of these studies was inconsistent and poor.[4] For example, the source of the health utility could not be determined in 24%.[4]

Economic evaluation in orthopedics is in its infancy as illustrated by inadequately designed studies.[3,4] Transfer to utility offers clinicians a practical and immediate method of assessing the cost-effectiveness of various total joint interventions and technological advances. For clinicians that collect disease specific measures to evaluate long-term and longitudinal outcomes, the regression algorithms can be used to map the disease specific scores to a health utility measure. The resulting health utilities can then be used to construct QALYs which are then used to calculate the incremental cost-effective ratio to evaluate the benefits gained from new interventions and technology. Ultimately, the cost-

effectiveness information can facilitate clinical and policy decision making when considering the adoption of new health care interventions. For example, a new total joint implant that potentially provides improved performance and longevity is approved for marketing. This new implant may differ in design from an older, yet similar, implant that the surgeon used previously. Disease specific measures can be used to compare the safety and efficacy of the two implants. The surgeon can further evaluate the cost effectiveness of these two implants using the TTU regression models. Based on the QALYs gained and the cost differential between the two implants, the decision can be made to utilize the most cost-effective implant.

Total Joint Arthroplasty Health Related Quality of Life Measures

Disease Specific Measures

Over the past few decades, several disease specific measures have been used to evaluate outcomes following TJA and there has been wide variation in the reporting of these measures.[6,7,28] Such variation introduces challenges in comparing the literature, which has clinical decision making, research and policy implications. Riddle et al.[28] conducted a meta-analysis to determine the extent of the variation in the use of such measurement tools implemented in randomized clinical trials.[28] The findings indicate that the KSS and WOMAC® instruments are the most commonly reported primary outcome for knee arthroplasty and the HHS and WOMAC® are the most commonly used tools to assess outcomes of hip arthroplasty.[28] Similarly, Ethgen et al.[8] reported that the WOMAC® was the most commonly reported instrument in cohort studies of TJA but this review did not include HHS or KSS in the criteria.

Harris Hip Score. The HHS was developed in 1968 by an orthopedic surgeon, William H. Harris.[14,29] The HHS includes pain, function, range of motion and hip deformity constructs.[14,29] Both patient-reported and provider-reported measures are included and the maximum score is 100 points.[14,29] Higher scores represent better clinical outcomes.[14] Of the total 100 possible points, 44 possible points are allocated for pain and 47 points are allocated for function.[14] The remaining points are assigned to range of motion and absence of deformity.[14] While the Harris Hip Score was initially tested on a small series of 38 hip fracture cases, it was also designed to be used with a variety of hip conditions and treatment options.[14]

Knee Society Score. The KSS was initially developed by the Knee Society in 1989 and further modified in 1993.[12] Based on a panel of surgical and clinical experts, i.e., Knee Society members, three main constructs were included in the score: pain, knee joint stability, and range of motion.[12] Similar to the HHS scoring algorithm, the maximum score is 100 points with higher scores indicating better clinical outcomes. Of the 100 points, a possible 50 points are assigned to pain, 25 points are for stability and 25 points are possible for range of motion.[12] Deformity and misalignment of the native joint and the knee arthroplasty are assessed but are deductions in the overall score.[12] A well-aligned knee with no pain, 125 degrees of motion, and good anteroposterior and mediolateral instability will achieve a KSS of 100 points.[12] The maximum function score is also 100 points and consists of patient reported outcomes that measure walking distance (50 points) and the ability to ascend and descend stairs (50 points).[12] The use of assistive walking devices, such as canes and walkers, are deducted from the total function

score.[12] It is common in the TKA literature to report the pain score, the function score as well as the total score.

Western Ontario-McMaster Osteoarthritis Index. The WOMAC® was developed in 1982 as a patient reported measure of hip and knee osteoarthritis.[9-11] The WOMAC® consists of 24 questions assessing three dimensions of pain, disability, and joint stiffness.[9-11] Because OA patients and physicians were involved in the development of the questionnaire, the items represent aspects of OA that are relevant to both.[9-11] Through numerous validation studies, it has been shown to be valid, reliable and responsive across a number of interventions.[9-11,13] The KOOS was developed in 1998 to assess clinical outcomes related to a variety of treatments for knee injuries as well as osteoarthritis.[17-19] The KOOS is a patient reported clinical outcome measure designed to evaluate clinical change over time. [17-19] The questionnaire includes all WOMAC® questions in their original form as well as questions related to sports and recreational specific activity related difficulties and knee related health related quality of life. [17-19] Subsequently, the HOOS was developed as an analogous clinical measure for hip related conditions. [15-17] Several studies have shown that the HOOS and KOOS are valid, reliable and responsive measures across a number of interventions.[17-19]

A literature search revealed that population values for the disease specific measures are not reported. The reported literature is specific to single studies. While institutions may prospectively collect disease specific measures to evaluate clinical outcomes of treatments there is currently no national repository of total joint clinical outcome data using disease specific or generic, preference-based measures.

Generic, Preference Based Measures

Generic, preference-based HRQOL instruments provide an indirect measure of generic health utility. The EQ-5D[20,21], HUI-3[21,30] and SF-6D[31] are three commonly reported generic, preference based HRQOL measures. These instruments include a set of non-disease-specific health states that are based on a combination of general attributes that have been valued by a sample of the general population. A scoring algorithm is created and patients with any disease complete the questionnaire and the appropriate scoring algorithm is applied to define the generic health state. The calculated single utility score is on an interval measurement scale, ranging from 0 to 1.

Health Utilities Index-3. The HUI-3 consists of eight structurally independent attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain.[21,30] The HUI-3 defines 972,000 unique health states. [21,30] The HUI-3 score is based on community preferences developed using a visual analog scale and a standard gamble approach.[21,30] The utility score is derived using a multiplicative, multi-attribute mathematical utility function. [21,30] The multiplicative algorithm defines the interactions among various health states and accurately predicts average scores for independent samples with a variety of diseases.[21,30] Several studies have tested the reliability and validity of the HUI-3 across diverse populations and a variety of disease conditions and intervention types.[21,30]

EuroQol 5D. The EQ-5D includes five attributes: mobility, self-care, usual activity, pain and discomfort, and anxiety and depression. [20,21] The permutations of the five attributes result in 243 unique health states. [20,21] Preference weights were developed using valuation population sets based on the visual analog scale technique and the time

trade-off method (TTO). [20,21] The EQ-5D utility score is calculated using a scoring algorithm that is based on econometric modeling. [20,21]

Short Form 12 and Short Form 6D. The SF-12 is a 12 item self-reported questionnaire that includes fewer questions of the eight attributes included in the original SF-36.[31] The eight attributes are as follows: physical functioning, role physical, role emotional, social functioning, bodily pain, mental health, and vitality.[31] Using the Quality Metric scoring services, SF-12 responses are converted to the SF-6D health utility score.[31] The SF-6D combines the role physical and role emotional attributes into one attribute defined as role participation.[22,23,31] The SF-6D uses an econometric scoring algorithm with population based preference weights.[22,23] In developing the algorithm, a total of 18,000 unique health states were defined using the standard gamble method.[22,23]

Comparative Utility Studies. Because the EQ-5D, HUI-3 and SF-6D have been shown to generate different utilities based on the condition and population, it is important that the appropriate instrument is selected.[32,33] Barton et al.[32] compared the scores from the EQ-5D and the SF-6D to measure the benefits alleviating knee pain. The study findings showed that both scores had comparable construct validity.[32] However, the SF-6D did not discriminate between those who improved post intervention and those who showed no improvement.[32] Blanchard et al.[33] investigated the construct validity of the HUI-3 in a series of patients with OA of the hip awaiting hip arthroplasty. One hundred and fourteen patients completed the HUI-3, HHS, and WOMAC® questionnaires. [33] The mobility and ambulation attributes of the HUI-3 showed moderate correlations with the HHS total score, and the physical function attribute of the WOMAC®.[33] The HUI-3 pain

attribute indicated a strong correlation with the WOMAC® physical function as well as moderate correlations with the HHS pain score and the WOMAC® pain score.[33]

Transfer to Utility

Transfer to utility is recognized as a valid method of obtaining a health utility to calculate QALYs for economic evaluations. There are two common types of TTU. One method, which is not the type being considered in the dissertation, involves mapping a generic (general health), non-preference based measure, such as Short Form 36 to a generic, preference-based measure, such as the Short Form 6-D. The second type of TTU, which is the type being considered in the dissertation, is to map a disease specific, non-preference based measure, such as the KOOS to the preference-based EQ-5D. To conduct a TTU, two data sets are required to develop the regression model. First, the estimation data set is initially used to develop the regression model. Once the best model is selected, it is then tested on the second data set to estimate the health utility from the disease specific measure. In a review of mapping studies, Brazier et al. included 38 papers that either mapped generic non-preference-based measures or disease specific measures to a generic preference based-measure. A total of 119 models were used across these 38 papers and the EQ-5D was the most common generic preference based-measure used. The sample sizes ranged from 68 to over 23,000 participants. A total of 12 studies were reviewed that used a disease specific measure and four of these were unpublished manuscripts or conference proceedings.

Transfer to utility has been utilized to predict health utilities from disease specific, non-preference based measures for several healthcare populations and disease conditions, including stroke[34], obesity[35], oral health[36], hydrocephalus[37], angina[38], and arthritis[39,40].

Two studies assessing stroke and obesity are described in more detail to illustrate the large variation in the precision of regression models in estimating health utilities from disease specific measures. Mortimer et al. developed regression models to estimate preference-based scores using the Assessment of Quality of Life (AQoL) from the disease specific National Institutes of Stroke Scale. The authors found a significant difference between the observed and estimated AQoL utility scores with mean absolute errors that ranged from 0.12 to 0.31 depending on the severity of the disability. Brazier et al assessed the accuracy of TTU regression models in estimating the SF-6D utility from the Impact of Weight on Quality of Life-Lite (IWQOL-lite) in a study sample of obese patients. The mean absolute error between the observed and the estimated SF-6D scores was only 0.098.

While the use of TTU to map values from disease specific health measurement instruments to utilities derived from generic health measurement instruments is increasing there have been few studies related to OA and TJA.[32,39-42] Furthermore, the WOMAC® is the only disease specific instrument that has been mapped to predict a preference-based health utility. Grootendorst et al.[39] constructed regression models to estimate health utility using the HUI-3 from the WOMAC® subscale scores in 255 knee arthroplasty patients. The best performing regression model included the WOMAC® subscales, the squared and interaction terms of the WOMAC® subscales as well as age, gender and duration of symptoms. The root mean squared error (RMSE) was 0.2065.[39] In a similar study, Marshall et al.[40] used the Grootendorst et al.[39] model in 145 preoperative hip patients and reported a RMSE of 0.1698.[40]

A TTU model to predict a health utility derived from the EQ-5D using elements of the WOMAC® instrument was developed using a sample of 348 patients with knee pain.[32] Barton et al. reported a RMSE of 0.180.[32] The authors also compared the QALYs calculated using the observed utility scores to the estimated utility scores and found that the calculated QALYs using the predicted health utilities were lower compared to the actual utility scores.[32]

In a review of mapping literature, Mortimer et al.[41] addressed the methodological and conceptual concerns with TTU regression and found large variation in the explanatory power of regression models. A broad overview of study results indicate that the explanatory power of models mapping disease specific HRQOL scores, such as the WOMAC®, is generally lower when compared to other models mapping broader HRQOL scores, such as the SF-36.[41] Mortimer cautions that disease specific instruments that are designed to measure detailed, or narrow, constructs may not be appropriate for TTU to a broader utility measure that covers a broad array of constructs.[41] The addition of data that captures a broader clinical picture may improve the models' ability to estimate health utilities.[41]

Mortimer et al.[41] addressed additional methodological concerns with the TTU regression method that remain unanswered in the current body of knowledge and should be considered. First, a series of group specific mappings for different conditions and severities may provide weaker utility predictive power than a single population based mapping.[41] Second, many disease specific instruments generate ordinal level scores which are then mapped to an interval level utility score.[41] Such models may compromise the ability of the predicted utility to maintain equal proportion changes in the calculated

QALY. [41] The extent to which one score maps to another is largely an empirical issue that has been understudied for instruments commonly administered in an orthopedic setting.

METHODS

Design and Data Source

This study was reviewed and approved as an expedited protocol by the Carolinas HealthCare System Institutional Review Board, the University of North Carolina at Charlotte Institutional Review Board and the Porter and Littleton Adventist Hospital Joint Institutional Review Board. Informed consent was waived. Potential participants were presented with a study overview and request for volunteer participation (See Appendix 2). Survey responses were stored in the password protected, 21 CFR Part 11 compliant, patient registry.

The data obtained from this cross-sectional survey of postoperative total hip and total knee arthroplasty patients who had surgery at either OrthoCarolina, P.A. (OC) located in Charlotte N.C or Colorado Joint Replacement Center (CJR) located in Denver, C.O.. The data obtained from this survey were used to develop regression equations using the TTU method to map disease specific outcome measures to generic, preference-based health utility scores. Participants were identified and recruited using the OrthoCarolina, P.A. Patient Registry and the Colorado Joint Replacement Center Patient Registry, which both store longitudinal data for TJA procedures.

The OC Hip and Knee Center and the CJR are both private, tertiary practices of specialty hip and knee replacement surgeons located in metropolitan areas. Eight hip and knee replacement surgeons practice at OC and five surgeons at CJR specialize in hip and knee replacement. On an annual basis, approximately 2,600 total joint surgeries are performed at OC and approximately 1,300 are performed at CJR. While the majority of patients reside in-state, approximately 15% travel from adjoining states and 3% from

other regions. These proportions are similar between each practice. Additionally, the payer mix at each practice is similar. Among the CJR total joint patient population, a total of 41% are Medicare beneficiaries, 54% have private insurance, and 5% are Medicaid. The proportion of OC patients with Medicare is slightly higher at 51%. The proportion of OC patients with private insurance is approximately 46% and the remaining 3% are Medicaid, self pay or workers compensation.

Both centers use the same registry software which has the technological capability to send a secure email to individuals with email addresses on file. Those who volunteered participation were presented with a link to the online survey and provided a general study overview and the specific instructions that accompany each questionnaire. Individuals who had any type of total hip replacement were asked to complete the HHS, WOMAC®, and HOOS disease specific questionnaires. Individuals who had any type of total knee replacement were asked to complete the KSS, WOMAC® and KOOS disease specific questionnaires. For patients who have had more than one joint replacement, the most recent was included in the sample. All patients were asked to complete the EQ-5D, SF-12 and HUI-3 generic, preference based health utility questionnaires. The HOOS and KOOS questions contain all of the WOMAC® questions and individuals were only asked those questions one time. Additionally, only the patient-reported pain and function constructs of the KSS and HHS were asked. The total number of questions was 47 for hip patients and 48 for knee patients and it was estimated to take 30 minutes to complete. To maximize response rate, three email blasts were sent. Between February 2011 and April 2011, three emails were sent to each OC TJA patient. The CJRI patients were each sent three emails between June 2011 and August 2011. Due to the programming of the online

survey, it was not possible for participants to submit responses more than one time. To minimize any order effects, patients from the study sample were randomly assigned to one of six blocks, which defined a different order of clinical measures and utility measures. The blocks are presented in Table 2.

A total of 1,788 total hip replacement patients and 2,458 total knee replacement patients were sent an email. Of the 4,246 emails sent, 3,258 (76.7%) patients did not complete the survey. The reasons for non-responses are as follows: 13 declined participation; 116 initiated the survey but did not answer any questions; 110 email addresses were identified as invalid; and 3,019 patients did not respond in any manner. Therefore, 988 (23.3%) patients were willing to participate. The 110 invalid email addresses were entered into the registry as @none.com or @decline.com and discovered at the time of analysis. It is not possible to determine how many of the 3,019 nonresponders had invalid emails due to either an inactive email account or a misspelled address. Therefore, the 23.3% response rate may be underestimated.

A total of 438 total hip replacement patients and 550 total knee patients participated in the survey. The total number of completed sets of the five disease specific TJA outcome measures and each of the three preference-based health measures is presented in Table 3. Of the 988 total joint replacement patients included in the sample, 504 were female, 484 were male and the average age was 61.0 years (SD 9.8 years).

To determine any differences between patients with respect to having an email address, 2,789 patients who had an email address on file (emailers) were compared to 2,370 patients who did not have a documented email address (nonemailers). Due to HIPAA issues, only data from the OrthoCarolina, P.A. Registry was available to assess

any differences between emailers and nonemailers. Therefore, the sample of nonemailers was obtained from a query of OrthoCarolina, P.A. registry patients who had a postoperative evaluation between January 2010 and December 2011. The emailers sample included only OrthoCarolina, P.A. study patients to minimize any further bias between sites. A significantly ($p=0.0002$) greater proportion of patients with no known email address were females. Of the 2,370 patients without a known email address, 1407 (59%) were females and 963 (41%) were males. Patients with no email address were significantly ($p<0.0001$) older than those with an email address. The average age of patients with no email address was 64.5 years (SD 11.8 years) compared to 61 years (SD 11.0 years) for the group of patients with a documented email address.

To determine any differences between patients who responded to the survey and those that did not respond, the 3,128 nonresponders were compared to the 988 responders with respect to gender, age and time since surgery. A significantly ($p=0.036$) greater proportion of nonresponders were females. Of the 3,128 nonresponders, 1715 (55%) were females and 1,413 (45%) were males. The mean age of nonresponders was 61.2 years (SD 11.2 years) compared to 61.0 years (SD 9.8 years) for responders ($p=.62$). There was also no significant ($p=0.95$) difference in the time since surgery between the two groups. The average time since surgery for both groups was 50 months.

Analysis Plan

Sample Size Estimate

Gatsonis and Sampson developed mathematical formulas to estimate power and sample size estimates for use in observational studies in which the independent variables are not fixed but are the outcome study measures.[43] The proposed regression models will

find significant beta coefficients ranging from 0.15 to 0.20 based on the following inputs: 1) available sample size; 2) number of independent variables in the model; 3) alpha level of 0.05; and 5) 80% power.[43] The available sample size ranges from 399 to 506 (sample sizes can be found in Table 3) based on the number of completed responses on each pair of disease specific measure and preference-based measure. Depending on the regression model, the number of independent variables ranges from 4 to 23. The number of independent variables for each regression model can be found in Table 4. Individual power estimates were derived for each combination of input possibilities. These estimates resulted in a range of 0.15 to 0.20 beta coefficients that will result in statistical significance at an alpha level of 0.05.

Univariate Analysis and Bivariate Analysis

Descriptive statistics of the sample demographics and scores for all questionnaires were calculated. These data were used to compare results to previous TJA research to determine generalizability of the findings. As discussed above, bivariate analyses were used to determine any differences between those who responded (responders) to the survey and those who did not respond to the survey (nonresponders). Additionally, bivariate analyses were used to determine any differences in those patients who had email addresses (emailers) on file and those that did not have email addresses (nonemailers) on file. The Wilks-Shapiro test was used to determine normal distribution of the residuals. Differences in proportions of each gender were assessed using a Chi Square test. . For normally distributed data, the differences in means between two groups were assessed using a Student t-test. A one-way analysis of variance (ANOVA) was used to determine

differences in the means between more than two groups. For data that was not normally distributed, a Wilcoxon two-sample test and Kruskal-Wallis test were used.

In the absence of an accepted threshold for clinically significant differences in the disease-specific measures and health utilities, an expert panel of adult reconstruction surgeons from OC was convened and polled. The HOOS, KOOS and WOMAC are measured using a five-point Likert scale and scored from 0 (worst) to 100 (best). The panel recommended that scores within ten points or a 10 percentage point difference between groups is not clinically meaningful.

Development of Prediction Models

As noted above, the key objective was to develop models that allow for the prediction of non-disease specific utility values from commonly used osteoarthritis-specific measures of pain and functioning. For each osteoarthritis-specific measure, a set of linear regression models were estimated for each of the three derived utility values indicating overall health using various functions of the survey elements from the osteoarthritis-specific measurement tool, demographic variables (*age*, *gender*), and time since surgery (*years*). For instance, using responses from the WOMAC®, which include a composite score (*totalWOMAC®*) and subscales for the degree of pain (*pain*), mobility (*mobility*), and stiffness (*stiffness*), the following models of utility as derived from the EQ-5D (*utility1*) were estimated:

1) $Utility = \beta_0 + \beta_1 totalWOMAC^{®} + \beta_2 age + \beta_3 gender + \beta_4 years + \mu$
2) $Utility = \beta_0 + \beta_1 pain + \beta_2 mobility + \beta_3 stiffness + \beta 4age + \beta_5 gender + \beta_6 years + \mu$
3) $Utility = \beta_0 + \beta_1 totalWOMAC^{®} + \beta_2 age + \beta_3 gender + \beta_4 years + \beta_5 totalWOMAC^{®} * age + \beta_6 totalWOMAC^{®} * gender + \beta_7 totalWOMAC^{®} * years + \beta_8 totalWOMAC^{®2} + \beta_9 age^2 + \beta_{10} years^2 + \mu$
4) $Utility = \beta_0 + \beta_1 pain + \beta_2 mobility + \beta_3 stiffness + \beta 4age + \beta_5 gender + \beta_6 years + \beta_7 pain * age + \beta_8 pain * gender + \beta_9 pain * years + \beta_{10} mobility * age +$

$$\beta_{11}mobility*gender + \beta_{12}mobility*years + \beta_{13}stiffness*age + \beta_{14}stiffness*gender + \beta_{15}stiffness*years + \beta_{16}pain^2 + \beta_{17}mobility^2 + \beta_{18}stiffness^2 + \beta_{19}age^2 + \beta_{20}years^2 + \mu$$

Regression models for the KSS/HHS and KOOS/HOOS instruments are shown in Table 3. Prediction models were constructed using 85% of the data for model development (in-sample), and 15% of the sample was held back and used only for assessing the predictive ability of the models with a sample not used to estimate the models (out-of-sample). Models were estimated using ordinary least squares. Prediction performance of the models was assessed using both in-sample and out-of-sample measures of prediction performance. Figure 3 illustrates a schema of the analytical steps involved in the mapping. The primary criterion used to assess model performance was the mean absolute error (the average absolute prediction error), where the preferred model will have the lowest mean absolute error. In addition, all models were evaluated using the mean error, the root mean squared error (the positive square root of the average squared prediction error) and the proportion of observations with absolute prediction errors above 0.1 for each model. Because there is no consensus in the literature for an acceptable proportion of observations with a large (>.1) absolute predication error, a determination was made that models with 10% or more of the forecast errors greater than 0.1 were not acceptable for individual prediction. Therefore, if 10%, or less of the forecast error was greater than 0.1, the model is deemed appropriate for use in estimating health utility values at the individual observation level in this study.

For each osteoarthritis-specific measure of pain and functioning, a formula was derived to estimate each of the three overall utility scores based on each osteoarthritis-specific measurement tool, demographic variables, and time since surgery. These

formulas were programmed into a basic SAS program to share with future researchers. In addition to the regression models previously described, neural networks (NN) were used to estimate each heath utility measure from each disease specific measure for each of the best performing regression models. A neural network is a computational data modeling method that is more precise and robust in handling complex, nonlinear data.[44,45] This method has been used in a variety of scientific areas such as medicine, economics, and sociology.[44,45] In general, NNs are modeled to simulate the processes of the human brain.[44,45] The initial layer of computational nodes represents the model inputs, such as the disease specific data.[44,45] These nodes are weighted and a mathematical function is modeled to estimate the output nodes, which in this case is the health utility.[44,45] The general design of the neural network is diagramed in Figure 4. Similar to the mapping process a sample of the data is used to train the NN.[44,45] This training process is an iterative process that compares the error in the estimated output and adjusts the weights to improve the accuracy of the NN.[44,45] Once the NN reaches the minimal error calculation, the training process is complete and the NN design is finalized.[44,45] The remaining data is then input into the NN and the final model estimates the health utility.[44,45] The mean absolute error and the root mean square error of the NN's were compared to the best performing linear regression model to determine the most accurate method of estimating the health utility from the disease specific measure. While a NN with three hidden nodes was used for analysis, additional networks were tested by varying the number of hidden nodes until it was clear that the MAE and RMSE were increasing.

RESULTS

Total Hip Arthroplasty

Results of Univariate Analysis

Sample Characteristics. Of the 987 individuals who participated in the online survey, 437 (44.3%) were THA patients. The majority of the THA patients were female (Table 5) and the mean age of the THA patients was 62.97 years (SD 10.75 years; Range 23.87 – 95.45 years). More than half of THA patients were < 65 years old (242 of 437 (55.4%)), 147 (33.6%) were between 65 years and 75 years old and only 48 (11%) were over the age of 75 years (Table 5). The mean follow-up for the THA cohort was 3.6 years (SD 3.63 years). One hundred and eighty one (41.4%) of the THA patients were less than two years from surgery, 155 (35.5%) patients were between two years and five years from surgery and 101 (23.1%) had greater than five years follow-up.

The demographics of the THA patients included in the study sample were similar to the THA patient population at OrthoCarolina. The mean age of the hip patients in the study sample was 62.97 years (SD 10.75 years) and the mean age for OC hip patients was 62.87 years (SD 12.65 years). This difference was not statistically significant (p=0.88). However, a significantly (P<0.0001) lower proportion of hip patients in the study sample were greater than 75 years of age (11%) than in the OC THA patient population (17%). There was a greater proportion of female THA patients at OC (53%) as compared to the hip patients included in the study (49%). This difference was not statistically significant (p=0.13).

Disease Specific Measures. Out of a maximum score of 44 points, which indicates no pain, the mean HHS pain score was 36 points (SD 9.9 points; Range 0 – 44 points). In

this cohort of THA patients, the mean HHS function score was 40.8 (SD 14.0 points; Range 0 - 47 points) out of a maximum HHS function score of 47 points. The total WOMAC® score is a composite of the function, pain and stiffness subscores. The mean score of 78 points (SD 22.3 points; Range 0 – 100 points) for the domain of stiffness was the lowest subscore. Pain and function were 83 points (SD 19.2 points; Range 0 – 100 points) and 81 points (SD 19.1 points; Range 0 – 100 points), respectively. The mean total WOMAC® score for THA patients was 81 (SD 19.1; Range 0 – 100 points). The HOOS score includes subscores for activity of daily living function (FnADL), sports and recreation activity function (FnSRA), pain, quality of life (QOL), and knee symptoms. While the HOOS questionnaire includes the total WOMAC® score, there is no total HOOS score. Patients rated themselves the lowest with respect to FnSRA (Mean 69 points; SD 26.2 points; Range 0 – 100 points) and QOL (Mean 73 points; SD 23.4 points; Range 0 – 100 points). The HOOS pain subscore, FnADL subscore, and symptom subscore were all between 81 points and 83 points, on average.

Generic, Preference-based Measures. Table 6 illustrates the median values for all three health utility measures. The health utility values derived from the SF-6D were slightly lower (Median 0.86 points; IQR 0.26; Range 0.37 – 1.00) on average than the values derived from the HUI-3 and slightly higher than the EQ-5D. The median EQ-5D value was 0.84 points (IQR 0.20; Range) The median value of 0.91 (IQR 0.19; Range 0.08 – 1.00) points obtained from the HUI-3 was the highest median health utility value.

Results of Bivariate Analysis

Sample Characteristics. A Student T-test was used to determine differences between the mean disease specific scores and between males and females. Using a

standard significance level of 5%, there were no significant differences between males and females in age at the time of follow-up, time since surgery, or pain and function levels. Table 7 illustrates the differences with respect to gender.

Analysis of variance tests were used to compare the mean values of the disease specific measures with respect to the three age groups (Table 8) and three follow-up periods (Table 9). As indicated by lower Harris Hip pain scores, average pain levels were significantly ($p=0.0004$) higher among patients who were less than 65 years of age as compared to those patients in either of the older age groups. While there are no criteria defined for clinically meaningful differences in the HHS subscores, there are categories of pain scores. Therefore, we defined a clinically meaningful difference as average scores that are defined by different categories. Mild pain is defined as a score ranging from 30 to 39. Therefore, the statistically significant difference in HHS pain subscores between age-groups does not represent a clinically meaningful difference.

While there were no significant differences in HHS function scores, the modified total HHS scores were significantly ($p=0.01$) lower among the youngest patients. Because there are no defined categories of the modified total HHS, we used a threshold value of 10 points to define clinical relevance. Thus, we do not denote the statistically significant difference in the modified total HHS as a clinically meaningful difference. As measured by each domain of the WOMAC® and HOOS questionnaires, patients in the youngest age group experience more postoperative pain and decreased function as compared to patients who are greater than 65 years of age at the time of survey. These differences are statistically significant for WOMAC® pain ($p=0.04$), WOMAC® stiffness ($p=0.05$), HOOS pain ($p=0.02$), HOOS QOL ($p<0.0001$), HOOS Symptoms ($p=0.01$). In

each of these subscores, the largest differences were noted between patients less than 65 years old and those greater than 75 years old. There are no formal criteria for interpreting whether these differences in the WOMAC® subscores and HOOS subscores are clinically meaningful. Both of these measures are based on a 5-point Likert scale with scores ranging from 0 (worst) to 100 (best). With that in mind, we posit that any differences in WOMAC® hip scores and HOOS scores that were within ten points are not clinically significant. Therefore, the differences in all of WOMAC® hip subscores, between age groups are not clinically meaningful. Similarly, the only clinically meaningful age-group difference was noted with the HOOS QOL subscore.

Patients who were less than two years from surgery at the time of the survey have higher postoperative pain levels and lower postoperative function levels as compared to those who are either between two and five years or greater than five years from surgery. These differences were statistically significant for the HHS pain ($p=0.03$), HHS function ($p=0.05$) and HHS modified total ($p=0.02$) scores as well as WOMAC® stiffness ($p=0.02$) and HOOS FnSRA ($p=0.01$). As previously described, we defined a clinically meaningful difference using a threshold of a ten point difference in the WOMAC® hip subscores and the HOOS subscores. Therefore, we do not consider any of the above statistically significant differences between followup intervals clinically meaningful.

Generic, Preference-based Measures. Wilcoxon two-sample tests and Kruskal-Wallis tests were used to determine statistical differences in health utilities with respect to gender, age-group, and time since surgery. There were no significant differences between males and females in the SF-6D ($p=0.87$), the EQ-5D ($p=.84$) or the HUI-3 ($p=0.21$). There were also no statistically significant differences between age groups for

the SF-6D (p=0.17), the EQ-5D (p=0.17) or the HUI-3 (p=0.26). For the SF-6D and the HUI-3, the lowest health utility values were measured for the oldest age group of patients that were greater than 75 years old. THA patients that were between the ages 65 years and 75 years of age reported the highest health utility values as derived by the SF-6D and the EQ-5D. The highest health utility values across the three age-groups were found with the HUI-3. There were no significant differences in health utility values between the time since surgery intervals as measured by the SF-6D (p=0.19), the EQ-5D (p=0.32) or the HUI-3 (p=0.06). For all three generic, preference-based measures, the highest health utility values were found in the group of THA patients that were between two and five years from surgery at the time of the survey. The highest median health utility value was measured using the EQ-5D for THA patients between two and five years follow-up. The standard gamble approach was used to determine the health utilities for the SF-6D and the HUI-3. Because the standard gamble approach is dependent upon the level of one's risk aversion, the health utilities are typically higher when elicited using the standard gamble approach. In contrast, the time trade off method does not include risk of death as an alternative and it typically yields lower utility than those produced using SG methods.[24,27] Therefore, it is unexpected to find higher health utility values as measured by the EQ-5D which used the time trade-off method to determine the possible health utilities. [24,27]

Total Knee Arthroplasty

Results of Univariate Analysis

Sample Characteristics. Of the 987 individuals who participated in the online survey, 548 (55.7%) were TKA patients. Fifty-three percent (291 of 550) of the TKA

patients were female (Table 10) and the mean age of the TKA cohort was 65.3 years (SD 8.28 years; Range 33.9 – 87.6 years). Nearly half of TKA patients were < 65 years old (270 of 550 (49.1%)), 211(38.4%) were between 65 years and 75 years old and only 69 (12.5%) were over the age of 75 years (Table 13). At the time of the survey, the mean time since surgery (follow-up) was 2.89 years (SD 2.5 years) for the TKA cohort. Two hundred and sixty four (48%) of the TKA patients were less than two years from surgery at the time of the survey, 192 (34.9%) patients were between two years and five years from surgery and 94 (17.1%) were greater than five years from surgery at the time of the survey.

The demographics of the TKA patients included in the study sample differed from the TKA patient population at OrthoCarolina. Knee patients at OC were significantly ($p<0.0001$) older than the knee patients included in the study. The mean age of the knee patients in the study sample was 65.3 years (SD 8.28 years) and the mean age for OC knee patient was 67.4 years (SD 10.63 years). There was also a lower proportion of patients in the study sample that were greater than 75 years of age (12.55%) than in the OC TKA patient population (23.43%). There was a greater proportion of female TKA patients at OC (61%) as compared to the knee patients included in the study (53%). These differences in age-group proportions ($p<0.0001$) and gender proportions ($p=0.0007$) were statistically significant.

Disease Specific Measures. Out of a maximum score of 50 points, the mean KSS pain score was 40.8 points (SD 13.8 points; Range 0 – 50 points). In this cohort of TKA patients, the mean KSS function score was 83.1 (SD 20.6 points; Range 0 – 100 points) out of a maximum score of 100 points. The total WOMAC® score is a composite of the

function, pain and stiffness subscores. The mean WOMAC® stiffness score was 78.9 points (SD 20.5 points; Range 12.5 – 100 points). The WOMAC® pain and function scores for the TKA cohort were similar. The mean pain score was 87.7 points (SD 16.2 points; Range 5 – 100 points) and the mean function score was 87 points (SD 16.6 points; Range 4.4 – 100 points). For the TKA patients, the mean total WOMAC® score was 85.8 (SD 16.1; Range 9 – 100 points). The KOOS score includes subscores for activity of daily living function (FnADL), sports and recreation activity function (FnSRA), pain, quality of life (QOL), and knee symptoms. While the KOOS questionnaire includes the total WOMAC® score, there is no total KOOS score. TKA patients also rated themselves the lowest with respect to FnSRA (Mean 64.9 points; SD 27.5 points; Range 0 – 100 points) and QOL (Mean 70.6 points; SD 24.3 points; Range 0 – 100 points). On average, the KOOS symptom subscore, pain subscore, and FnADL subscore were all between 81 points and 87 points.

Generic, Preference-based Measures. Table 11 illustrates the median values for all three health utility measures. The health utility values derived from the SF-6D were slightly lower (Median 0.80 points) and a little more variable (IQR 0.24 points; Range 0.4200 – 1.0000) on average than the values derived from the EQ-5D and HUI-3. A total of 506 (92%) TKA patients completed the EQ-5D. The median EQ-5D value was 0.83 points (IQR 0.20 points; Range 0.3078 – 1.0000). The highest median utility value for the TKA patients was the HUI-3.The median HUI-3 health utility value was 0.85 points (IQR 0.23; Range -0.19 – 1.00. Because the standard gamble approach is dependent upon the level of one's risk preference, it is expected that the HUI-3 and SF-6D would yield higher utility than the EQ-5D which was developed using the TTO method.[24,27]

Results of Bivariate Analysis

Sample Characteristics. A student T-test was used to determine differences between the mean disease specific scores for males and females. Using a standard significance level of 5%, there were no significant differences between males and females in age at the time of the survey or years since surgery at the time of survey. Table 12 illustrates the differences with respect to gender. There were statistically significant differences in pain and function levels between male and female TKA patients. Females reported significantly more pain and lower function levels as measured by the KSS, and the WOMAC®. Additionally, females reported significantly lower function and more pain on the relevant KOOS subscores as compared to male TKA patients. Using our threshold of 10 point difference in these scores as the clinically meaningful threshold, none of these differences were clinically meaningful. Analysis of variance tests were used to compare the mean values of the disease specific measures and the generic, preference-based measures with respect to the three age groups (Table 13) and three follow-up periods (Table 14). There were no statistically significant differences in KS pain scores ($p=0.38$) or KS function scores ($p=0.07$) between the three age groups. As measured by each domain of the WOMAC®, mean WOMAC® scores were nearly identical across all three age groups. There was a statistically significant ($p=0.04$) difference in the symptoms subscore across age groups. The youngest age group experienced more pain (Mean 78.8 points, SD 17.5 points), on average, than patients over the age of 65 years at the time of the survey. In addition to age effects, patients whom are less than two years from surgery at the time of survey report higher postoperative pain levels and lower postoperative function levels as compared to those who are either between two and five years or greater

than five years from surgery. These differences were statistically significant for all subscores of the Knee Society, WOMAC®, and KOOS questionnaires. There was a clinically meaningful difference in WOMAC® stiffness subscore with those less than two years from surgery experiencing significantly more knee stiffness. TKA patients less than two years from surgery also reported statistically and clinically lower sports and recreation function levels as well as lower QOL compared to those with more time between the dates of surgery and survey.

Generic, Preference-based Measures. There were significant differences between males and females in health utility values for the SF-6D and the EQ-5D. Male TKA patients reported a significantly (p=.0009) higher median SF-6D health utility of 0.84 (IQR 0.20) compared to the median SF-6D value of 0.80 reported by female TKA patients. Similarly, the median EQ-5D health utility was 0.84 for males compared to .83 for females (p=0.03). There were no statistically significant differences in health utility measures across age groups for either the SF-6D (p=0.37), EQ-5D (p=0.19), HUI-3 (p=0.10). The lowest health utility measures for the TKA patients was 0.77 (IQR 0.18) as derived by the SF-6D. The highest health utility of 0.91 (IQR 0.22) was reported by the youngest TKA patients (less than 65 years) There were also no statistically significant differences across follow-up periods between the SF-6D (p=0.35), EQ-5D (p=0.34), or HUI-3 (p=0.44). Although not statistically significant, the highest health utility value was reported by TKA patients that were between two and five years from surgery at the time of the survey as derived using the HUI-3. Because standard gamble methods tend to yield higher utility values, it is expected that the health utilities derived from the SF-6D and HUI-3 will be higher on average than the values obtained using the EQ-5D.[24,27]

Results of Multivariate Analysis: Prediction Models

A total of 60 prediction models were developed and tested for each combination of generic, preference based measure (health utility) and disease specific measure. The models are presented in Table 3. Eighty-five percent of the sample was used to develop the prediction models and 15% of the sample was used to test, or validate the models. Four criterion measures were used to evaluate the performance of the prediction models: mean error (ME), mean absolute (MAE), root mean square error (RMSE) and the percentage of errors greater than 0.1. The ME is the average forecast error, which is the difference between the estimated, or predicted, health utility value and the actual, observed, health utility value. The ME is used to assess the prediction models' accuracy in estimating the health utility value at the group level because underestimated values and overestimated values cancel each other. The MAE is the average forecast error between the estimated health utility value and the actual health utility value without regard to the direction, or sign, of the error. The RMSE is the square root of the mean of the squared forecast errors. The MAE and the RMSE are used to assess the ability of the model to accurately predict health utility values at the individual level. The percentage of errors greater than 0.1 was also used to assess the model's precision in predicting the health utility value at the individual observation level. Because there is no guidance in the literature for an acceptable threshold, a value of 10% of the forecast error greater than 0.1 was selected. Moreover, if 10%, or less of the forecast error was greater than 0.1, the model is deemed appropriate for use in estimating health utility values at the individual observation level.

A total of 30 THA models were developed and validated to map each of the three hip disease specific measures (WOMAC®, HHS, HOOS) to each of the three healthy utility measures (SF-6D, EQ-5D, HUI-3). Similarly, 30 TKA models were developed and validated to map each of the three of knee disease specific measures (WOMAC®, KSS, KOOS) to each of the three healthy utility measures (SF-6D, EQ-5D, HUI-3). The ME for all of the 60 mapping models was nearly zero, with non-zero values only when one was willing to report out to 15 decimal places. The additional prediction performance measures (MAE, RMSE, and proportion of error over 0.1) for the THA regression models and neural networks are presented in Tables 15-21. Table 15 illustrates the forecast errors for the mapping of the WOMAC, the HHS, and the HOOS to the SF-6D. Similarly, the performance measures for 10 EQ-5D mapping models are reported in Table 16 and the 10 HUI-3 mapping models are reported in Table 17. The prediction measures for the TKA data are presented in the same manner in Tables 22-24. The best performing THA models and the recommendations for the most accurate mapping combinations as well as the best performing disease specific measure and generic preference-based measure are presented in Table 18. Table 25 reports the recommendations for the top performing pairs and measures for the TKA sample. The coefficient estimates for the 9 most accurate THA prediction models are reported in Table 19 (SF-6D), Table 20 (EQ-5D) and Table 21 (HUI-3). For the TKA cohort, the coefficient estimates of the 9 top performing models are presented in Tables 26-28.

For each of the combinations, or pairs, of the health utility measures and disease specific measures one model was chosen as the most accurate prediction model. Of the 30 THA models developed and tested for each of the 9 pairs, 9 models were selected as

the best performing models to predict each health utility from each THA disease specific measure. From those 9 prediction models, the overall best performing model to predict a health utility from a THA disease specific measure was selected. The validation models with the lowest forecasting errors were selected as the most accurate prediction model. While the validation dataset was used for decision making, the performance measures for the development, or analysis, dataset were also evaluated. The same selection process of the most accurate mapping models was used for the TKA sample. In every case, the models that included the WOMAC®, the HHS/KSS, or the HOOS/KOOS subscores as well as demographic data (age, gender, follow-up), interaction terms, and squared terms were selected as the best performing prediction models. Additionally, these best performing models had lower MAE and RMSE than each of the corroborating neural networks.

Total Hip Arthroplasty Mapping Models

SF-6D Models. The performance criterion measures for the models mapping the SF-6D health utility value are reported in Table 15. Of the four models developed and validated to map the SF-6D to the WOMAC®, Model 4 was the most accurate. Model 4 included the subscores of the WOMAC®, demographic factors and interaction terms. For the estimation sample, Model 2 had a lower percentage of large forecast errors, and the MAE and RMSE were lowest for Model 4. With respect to the validation dataset, or holdout sample, Model 4 had the lowest values for MAE and RMSE as well as the percentage of forecast errors greater than 0.1. Because 26% of the forecast errors are greater than 0.1, Model 4 should only be used at the group level to predict the SF-6D

from the WOMAC®. If one only has WOMAC® scores, these data indicate that SF-6D is the preferred choice for mapping health utilities.

Two models were developed for mapping the SF-6D and the Harris Hip Score. While only the pain and function subscores were available for this study, the model that included the HHS subscores, demographic data, interaction terms and squared terms provided the lowest forecasting errors. For Model 6, 15% of the errors were greater than 0.1. While the proportion of errors greater than 0.1 is lower as compared to Model 4, Model 6 is also not appropriate to estimate the SF-6D from the HHS at the individual level. However, this model is appropriate for use at the group level.

For the mapping combination of SF-6D and HOOS, four models were tested. Model 10 was selected as the most accurate of these four prediction models and it includes the factors of the HOOS subscores, demographic data, interaction terms and squared terms. Model 10 had the lowest values on all four criterion measures for the estimation sample. For the validation sample, the MAE, the RMSE and the percentage of large errors were all considerably lower for Model 10. Model 10 is appropriate for use at the group level. However, the proportion of errors greater than 0.1 is greater than the 10% threshold which indicates it may not accurately predict the SF-6D from the HOOS at the individual level. The coefficient estimates for the best performing models developed to map the WOMAC®, the HHS, and the HOOS to the health utility derived from the SF-6D are reported in Table 19. These coefficient estimates can be used by researchers to apply Model 4 (WOMAC®), Model 6 (HHS) and Model 10 (HOOS) to their datasets to estimate SF-6D health utilities. For example if one has access to HHS, then Model 6 is the appropriate model to use.

EQ-5D Models. The performance criterion measures for the models mapping the EQ-5D health utility values are reported in Table 16. There was some variability among the performance criterion measures for selecting the best performing WOMAC® mapping model. Models 3 and 4 were very close. Model 3 had the lowest percentage of large errors while Model 4 had the lowest MAE and RMSE. Researchers could be confident using either model at the group level only.

Model 6 (HHS) and Model 10 (HOOS) had similarly low forecast errors for the holdout sample. For both of these models, only 12% of the sample had forecast errors greater than 0.1. However, this is greater than the 10% criterion and thus Model 6 and Model 10 should be used to estimate EQ-5D health utility values at the group level. Given that a researcher only has access to HHS or HOOS data, the EQ-5D provides the best option for mapping health utility values. Table 20 illustrates the coefficient estimates for the most accurate models to predict the EQ-5D health utilities are presented. Depending on which disease specific measures are available, researchers can use the coefficient estimates from either Model 4, Model 6 or Model 10 to estimate health utilities derived from the EQ-5D from their respective THA cohorts.

HUI-3 Models. The performance criteria measures for the models mapping the HUI-3 are presented in Table 17. Of the models mapping the WOMAC® to the HUI-3, Model 4 had the lowest MAE, RMSE and percentage of large errors for the holdout sample. For Model 4, a total of 30% of the errors were greater than .1 so this model should only be used to estimate the HUI-3 at the group level.

Model 6 and Model 10 are again, the best performing models to predict the HUI-3 from the HHS and the HOOS, respectively. The percentage of large errors was 18% for

Model 6 and 17% for Model 10. We would recommend that these models were only used to estimate the HUI-3 at the group level. The models used to map the THA related disease specific measures to the HUI-3 yielded the highest forecast errors among all of the mapping pairs for the THA cohort. Therefore, we would not recommend using the HUI-3 for THA patients. With that recommendation in mind, the coefficient estimates for mapping the HUI-3 from either of the 3 disease specific measures are presented in Table 21.

Summary Total Hip Arthroplasty Models.

The results from the data indicate that the single, best performing model for THA patients is Model 10 which mapped the HOOS to the EQ-5D. Using the holdout sample, the MAE was 0.0522 and the RMSE was 0.0649. Only 12% of the forecast errors were above the threshold criterion of 0.1. It is important to note that Model 6 that mapped the HHS to the EQ-5D had very similar forecast errors. For Model 6, the MAE was 0.0551 and the RMSE was 0.0705. Only 12% of the holdout sample for Model 6 had large errors, which is equivalent to proportion of large errors found using Model 10. With such small differences in the performance criterion values between Model 6 and Model 10 one could debate a tie for the best performing model for THA mapping pairs.

Total Knee Arthroplasty Mapping Models.

SF-6D Models. The performance criterion measures for the models mapping the WOMAC®, the KSS, and the KOOS to the SF-6D are illustrated in Table 22. Four models were developed and validated to map the WOMAC® to the SF-6D health utility. Of these four models, Model 4 had the lowest MAE and RMSE. As previously described,

Model 4 included factors for the WOMAC® subscores, demographic data, interaction terms and squared terms. Because 22% of the forecast errors were greater than 0.1, the SF-6D and WOMAC® should only be mapped at the group level.

Model 6 had the lowest forecast errors of the two models that mapped the KSS to the SF-6D. While the forecast errors indicate that Model 6 is also an acceptable mapping option, health utilities derived using the SF-6D should only be estimated at the group level. Model 10 had the lowest MAE, and RMSE values and was the most accurate of the four prediction models that mapped the KOOS to the SF-6D. Model 10 resulted in a total of 25% of the forecast errors above 0.1. Therefore, this model is also not appropriate to be used to estimate the SF-6D health utilities of individual TKA patient level.

The forecast error values were all very similar among these three top performing SF-6D models. The MAE ranged from 0.0709 to 0.0750 and the RMSE ranged from .0876 to .0974. These data suggest that these models can be used to map to the group level but we do not recommend that any of the SF-6D models be used to map at the individual TKA patient level. The coefficient estimates for Model 4, Model 6 and Model 10 are reported in Table 26. Researchers can use these coefficient estimates to estimate the SF-6D health utility of their TKA cohort.

EQ-5D Models. Table 23 reports the performance criterion measures of the models that map the WOMAC®, the KSS, and the KOOS to the health utilities derived from the EQ-5D. The forecast errors were very similar for the four models developed and tested to predict the EQ-5D from the WOMAC®. Model 3 had the lowest ME. However, Model 4 showed the lowest MAE and RMSE values and is deemed as the best

performing model. Only 10% of the errors noted in Model 4 were greater than 0.1 so it can confidently be applied at the individual patient level and the group level.

Model 6 was clearly the best performing model to estimate the EQ-5D from the KSS. Model 6 had the lowest MAE and RMSE. Additionally, 12% of the errors were larger than 0.1, which indicates that it performs well at the the group level, only. Model 10 showed the lowest MAE and RMSE values and was the most precise model in mapping the KOOS to the EQ-5D health utility. While only 13% of the errors were greater than 0.1, Model 6 can only be used to estimate the EQ-5D from the KSS at the group level. The coefficient estimate for the best performing models in mapping each of the three disease specific measures to the EQ-5D derived health utilities are presented in Table 27. Researchers can apply these models using these coefficients to estimate the EQ-5D health utility of a group of TKA patients. Because the percentage of forecast errors are greater than 15%, we do not recommend applying any of these EQ-5D mapping models to individual TKA patients.

HUI-3 Models. Table 24 reports the performance criteria measures for the models mapping to the HUI-3 health utility values. Of the three health utility measures, the HUI-3 results in the highest forecast errors when mapped to either the WOMAC®, the KSS or the KOOS. Nevertheless, the models that include the subscores, demographic data, interaction terms and squared terms were the most precise of the HUI-3 mapping models. This includes Model 4 for the WOMAC®, Model 6 for KSS and Model 10 for the KOOS. All of these models had more than 30% of the forecast errors greater than 0.1. Therefore, if one must estimate the health utility derived using the HUI-3 it should only be

undertaken at the group level. To apply any of these HUI-3 mapping models to another TKA dataset, one can use the coefficient estimates reported in Table 28.

Summary Total Knee Arthroplasty Mapping Models.

These results indicate that Model 6 produces the most accurate health utility values of all of the TKA mapping models. Model 6 maps the KSS to health utility values derived from the EQ-5D. For Model 6, the MAE was 0.051 and the RMSE was .064. However, the percentage of large errors was 10% for Model 4 compared to 12% for Model 6. Ultimately model 6 was favored because the other measures were slightly higher for Model 4. The MAE for Model 4 was 0.056 and the RMSE was 0.07.

DISCUSSION

The safety and efficacy of TJA is most commonly evaluated and reported in the literature using disease specific measures of health-related quality of life.[3,4,6,8,12,14,16-19] To assess the outcomes of THA, disease specific measures including the WOMAC®, the HHS and the HOOS are used. To assess clinical outcomes of TKA, the WOMAC®, the KSS and the KOOS are the most commonly reported disease-specific measures. While these tools provide meaningful data, these disease specific measures cannot directly be used in economic evaluations involving CUA. Best practices in cost-effectiveness analysis involve the use of quality adjusted life years as the outcome of interest in an intervention. Researchers must have information on the patient's level of health and the duration of time spent in that health state to calculate QALY's. Multi-dimensional HRQOL questionnaires, such as the SF-6D, EQ-5D, and HUI-3, can be administered to obtain the health utility level. To date, these general, preference-based HRQOL instruments have not been routinely administered in an orthopedic practice setting.

In order to meet the current and future demands of defining value-based medicine, or cost-effective treatments, researchers need the tools to conduct economic evaluations in a timely manner by utilizing data that are routinely collected on patients in an orthopedic setting. Therefore, we sought to generate regression models that could be used by researchers to accurately predict, or map to, a health utility from a common disease specific measure. To that end we conducted a multi-center, cross-sectional survey of total joint arthroplasty patients. Total hip patients completed the WOMAC®, HHS, and HOOS and total knee patients completed the WOMAC®, KSS, and KOOS. All patients completed the SF-6D, EQ-5D, and HUI-3 preference-based measures.

To our knowledge, there are no published studies that present utility values for relatively representative TKA or THA patient populations. The health utility values reported in the literature for total hip replacement are very specific to intervention and limited by small sample sizes. Therefore, the health utilities reported in this dissertation provide useful information for researchers who wish to use average values from the literature as model inputs for future cost-effectiveness model studies. The median health utility level for the total hip replacement cohort ranged from 0.86 (SF-6D) to 0.91 (HUI-3). We found that these values remained consistent regardless of gender and the age-group of THA patients. Similarly, we did not find any significant differences in utility values derived from any of the three indirect measures across the time intervals between surgery date and survey date for THA patients. For the TKA cohort, the health utility values were 0.80 (SF-6D), 0.83 (EQ-5D) and 0.85 (HUI-3). While we found statistically significant differences in the SF-6D and EQ-5D health utilities between male and female TKA patients, the differences are not clinically meaningful. Furthermore, the health utilities remained consistent across age groups and time intervals between date of surgery and date of survey.

Considering that the EQ-5D yielded the most precise mapping models, we recommend that the EQ-5D health utility values reported here are used by researchers. We posit that the EQ-5D was most accurately estimated by most of the disease-specific measures because the domains between these instruments are most closely related. The domains and subscales of the EQ-5D include mobility, self-care, usual activity, pain and discomfort, anxiety and depression. As compared to the SF-6D and HUI-3, a greater proportion of the EQ-5D questions relate to pain and function which are most relevant to

orthopedic conditions in general. More specifically, the TJA disease specific measures are also predominantly measuring pain and function.

The purpose of the study was to develop and test regression models using multiple pairs of disease, specific measures and generic, preference-based measures. A total of 30 THA mapping models and 30 TKA mapping models were developed and validated. Forecast errors including ME, MAE, RMSE and the percentage of errors greater than 0.1 were defined as our prediction performance criterion. The models that had the lowest forecast error measures on the validation dataset, or hold-out-sample, were selected as the best performing or most accurate prediction models. Furthermore, we posited that that a model should only be used to estimate health utilities at the individual observation level 10% or fewer of the observations from the hold-out-sample had forecast errors greater than 0.1.

Grootendorst et al.[39] were among the first researchers to develop a mapping model to predict a HUI-3 health utility from an OA disease specific measure, the WOMAC® in patients with OA of the knee. Marshall et al. subsequently repeated the Grootendorst et al.[39] investigation, developing and testing four regression models to map the WOMAC® to the HUI-3 using a dataset of 145 patients with hip OA.[40] The model that included the WOMAC® subscores, age, gender, OA duration, interaction terms and squared terms performed marginally better than the other models. Marshall et al. reported a MAE of 0.1698, a RMSE of 0.1684 and a ME of 0.0120 for the most accurate model. Our findings are consistent with Marshall et al.[40] We noted that our Model 4 was the most accurate mapping the WOMAC® and the HUI-3, with a MAE of 0.0901, a RMSE of 0.1177 and a ME of 1.5231xE-15. The forecast error measures we reported are lower

than the forecast error measures reported by Marshall et al. Consistent with Marhsall et al., we also do not recommend that the model is used at the individual patient level.[40] While the model provides reasonably accurate health utility estimates, we found that all of the models that estimate the HUI-3 from any disease specific measure are associated with higher forecast error measures as compared to the SF-6D and the EQ-5D.

Compared to the forecast errors from the Marshall et al. study, the forecast errors for the single best performing hip Model 10 in this study, which mapped the HOOS to the EQ-5D , were considerably lower. The MAE was 0.0522, the RMSE was 0.0649 and the ME was 2.381xE-16. Additionally, only 12% of the errors were greater than 0.1. Therefore, we confidently recommend that this model provides accurate estimates of EQ-5D health utilities at the group level in the situations when direct elicitation of preferences is not possible.

In a sample of 255 patients with OA of the knee, Grootendorst et al. developed and validated four prediction models to map the WOMAC® to the HUI-3.[39] The best performing model in their study was the one that included the WOMAC® subscores, age, gender, years since OA onset, squared terms and interaction terms. This finding is consistent with the present study in that all of the best performing models included disease specific subscores, age, gender, time form surgery to survey, squared terms an interaction terms. Grootendorst et al. reported a MAE of 0.1628, a RMSE of 0.2065 and an ME of -0.0003.[39] These forecast error measures are larger than those we have reported. The MAE for our Model 4 (WOMAC® and HUI-3) was 0.1067, the RMSE was 0.1542 and the ME was 5.2775E-16. While Grootendorst et al. concluded that the model was appropriate for mapping at the group level, they do not recommended that

researchers use the model to estimate the health utilities of individual patients.[39] Because 32.5% of the forecast errors were greater than 0.1, we also do not recommend that HUI-3 utility values be estimated from the WOMAC® at the individual level. With respect to the HUI-3, the KOOS mapping is more accurate than the other disease specific measures. Nevertheless, the forecast errors of all of the HUI-3 models were the largest among the three preference-based measures. Therefore, we do not recommend using the HUI-3 health utility with either TKA or THA cohorts.

For the TKA cohort, Model 6, which mapped the KSS to the EQ-5D, provided the most accurate health utility estimates. For this model, the MAE was 0.051, the RMSE was .064 and the ME was 6.323E-16. In comparison, the forecast errors with the Grootendorst et al. hip OA mapping model, were substantially larger. Thus, we feel confident that researchers can apply this regression model to their respective TKA group cohorts to estimate health utilities derived from mapping the HHS.

Limitations

The cross sectional survey used to obtain data for the TTU mapping is subject to a number of internal validity threats. Because the questions measure general constructs of pain, function, mobility and general health, the questions across tools are similar. The presentation of one question may affect the answer on another question presented later. Therefore, testing and instrumentation threats to internal validity are concerns. To minimize these threats, random assignment to blocks defining different orders of questionnaires was done. Selection bias is also a concern. A sample of OrthoCarolina, P.A. patients with no emails in the patient registry was compared with the OrthoCarolina, P.A. study sample of those with email addresses. There was a significantly greater

proportion of females with no emails than men. On average, patients with no emails were three years older than patients with email addresses. Therefore, the convenience sample of those who have email addresses may represent a healthier sample with higher levels of independence, cognitive ability, and functional ability. It is also possible that those who volunteer participation may elect to do so either because they are more or less satisfied with their health status compared to non-volunteers and may not be representative of the larger sample. There was a significantly higher proportion of females that did not respond to the survey as compared to those that did respond. However, there were no differences between these two groups with respect to age or time since surgery.

External validity is also a limitation of this study. The sample is from two private orthopedic practices of high volume, specialty surgeons. We compared demographic variables between the study sample and the OrthoCarolina TJA population to evaluate the extent to which the study sample is representative of a larger population. We found that significant differences did exist. In comparing the demographic characteristics of the two knee samples, there was a significantly lower proportion of females in the study sample and the study sample was significantly younger than the OC population of TKA patients. These demographic differences were not as profound in the hip patients and the reason for this finding is unclear. Nevertheless, the results may not be generalizable to the populations of TKA and THA patients. Additionally, as noted in the literature review, due to the methodological challenges with TTU, predictive validity of the mapping models may be limited with the relatively narrow coverage of the disease specific measures relative to the broader, preference-based measures. The models developed may

require further research to larger arthroplasty populations to determine reliability and validity.

Future Research

The results of this study highlight potential areas for future research. The accuracy of these linear regression equations may be improved by investigating other nonlinear regression models. A comparison of the forecast errors associated with the linear and nonlinear models would inform future researchers in this area to the most accurate models to use to estimate health utilities. Additionally, it may be useful to further investigate the precision of models in estimating health utilities for various subpopulations, i.e males vs. females and different age-groups. One next logical step is to evaluate how the calculation of the incremental cost-effectiveness ratio varies using the estimated health utility compared to the actual health utility. It is also important for other researchers to use these regression equations on their datasets to test the robustness of the models in accurately predicting health utilities. Finally, this line of research can be extended into other subspecialties of orthopedic medicine.

Summary

In spite of the limitations noted, a total of five osteoarthritis disease specific measures commonly used to evaluate TJA were mapped to three commonly used preference-based health utility scores derived from multi-attribute health assessment instruments. While TTU has been used in other health-related studies, the few studies relevant to TJA included relatively small sample sizes and have only mapped the WOMAC® to the HUI-3. These models were developed and tested using data collected

from two large orthopedic practices, which provided the largest sample that has been used to date to develop TTU regression models. For cost utility analysis evaluating THA intervention options, the HHS subscores most precisely estimated an EQ-5D health utility. Given one has HHS pain and function scores, we recommend that Model 6 be used to map to the EQ-5D and derive a health utility value either at the individual patient level or at the group level. If one only has access to WOMAC® scores, we recommend that Model 4 be used to estimate an SF-6D health utility. EQ-5D health utilities are also precisely predicted using Model 10 and HOOS subscores. In the event that one needs to estimate health utilities for TKA interventions, the models predicting EQ-5D health utilities, regardless of the disease specific measure, were the most precise. If one has access to all three of these disease specific measures, we recommend using Model 6 to map the KSS to the EQ-5D. The models developed in this dissertation will allow clinician-researchers to translate disease specific outcome scores to utilities, thus improving the ability of osteoarthritis researchers and policymakers to consider the cost-effectiveness of osteoarthritis-related interventions relative to interventions for very different diseases and conditions. TTU offers a useful method to estimate utilities from disease specific measures and facilitate economic evaluations of current and new TJA interventions to better understand the true societal benefit of the interventions.

FIGURES

Total Joint Implant A:
Five years X 0.75 health utility = 3.75 QALYs

Total Joint Implant B:
Five years X 0.50 health utility = 2.5 QALYs

1.25 QALYs gained with intervention A

$$\text{Incremental Cost Effective Ratio} = \frac{\text{Cost of Implant A} - \text{Cost of Implant B}}{\text{QALYs Implant A - QALYs Implant B}}$$

Figure 1. Cost-Effective Analysis Example

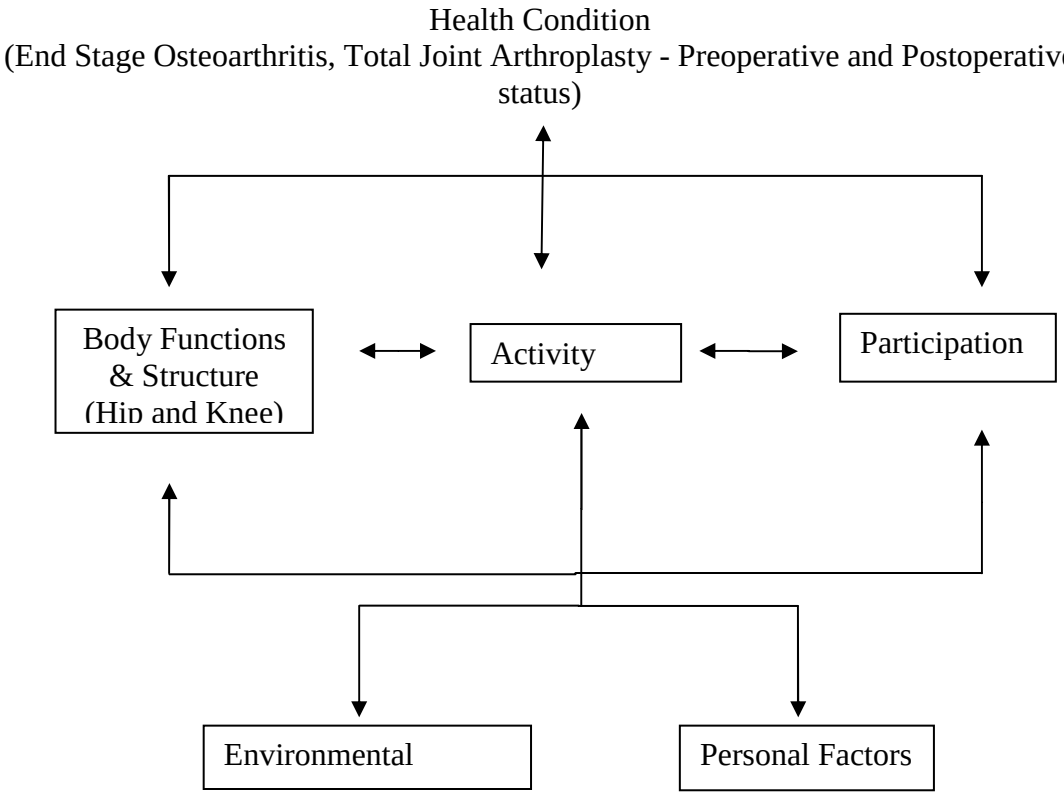

Figure 2. International Classification of Functioning, Disability and Health (ICF): Conceptual Framework of Disability

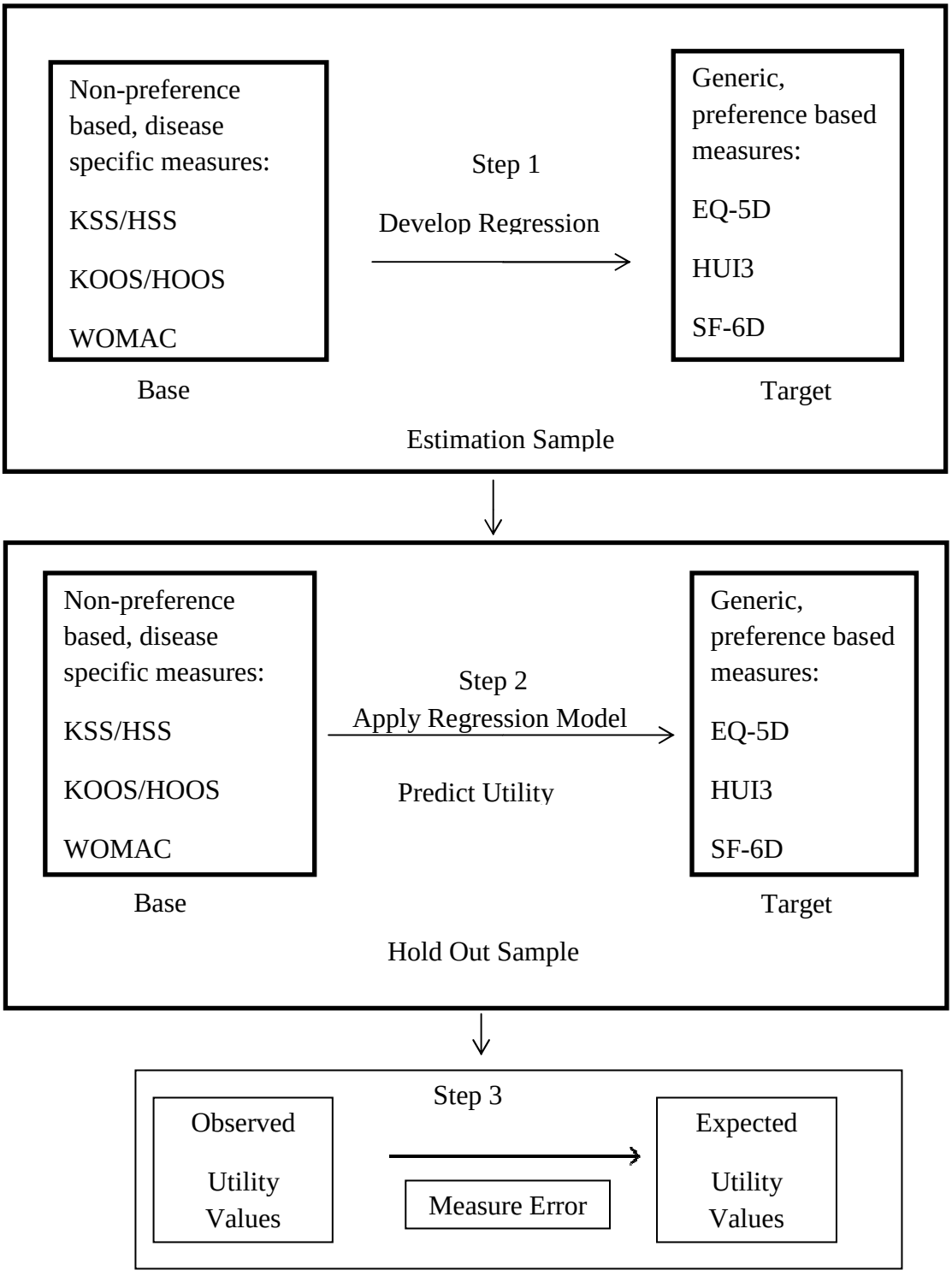

Figure 3. Schematic of the Analytic Steps in the Mapping Process (Chuang & Whitehead, 2012)[46]

Y = f(X)
X = set of numeric inputs (age, gender, time since surgery, disease specific score, relevant subscales for disease specific score)
w = weights
Y = set of numeric outputs
F() = unknown functional relationship between the inputs and the outputs

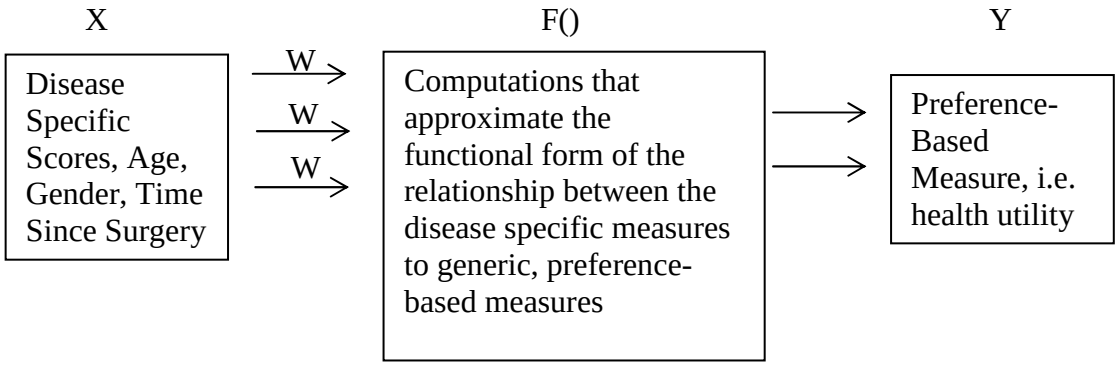

Figure 4. General Neural Network Diagram

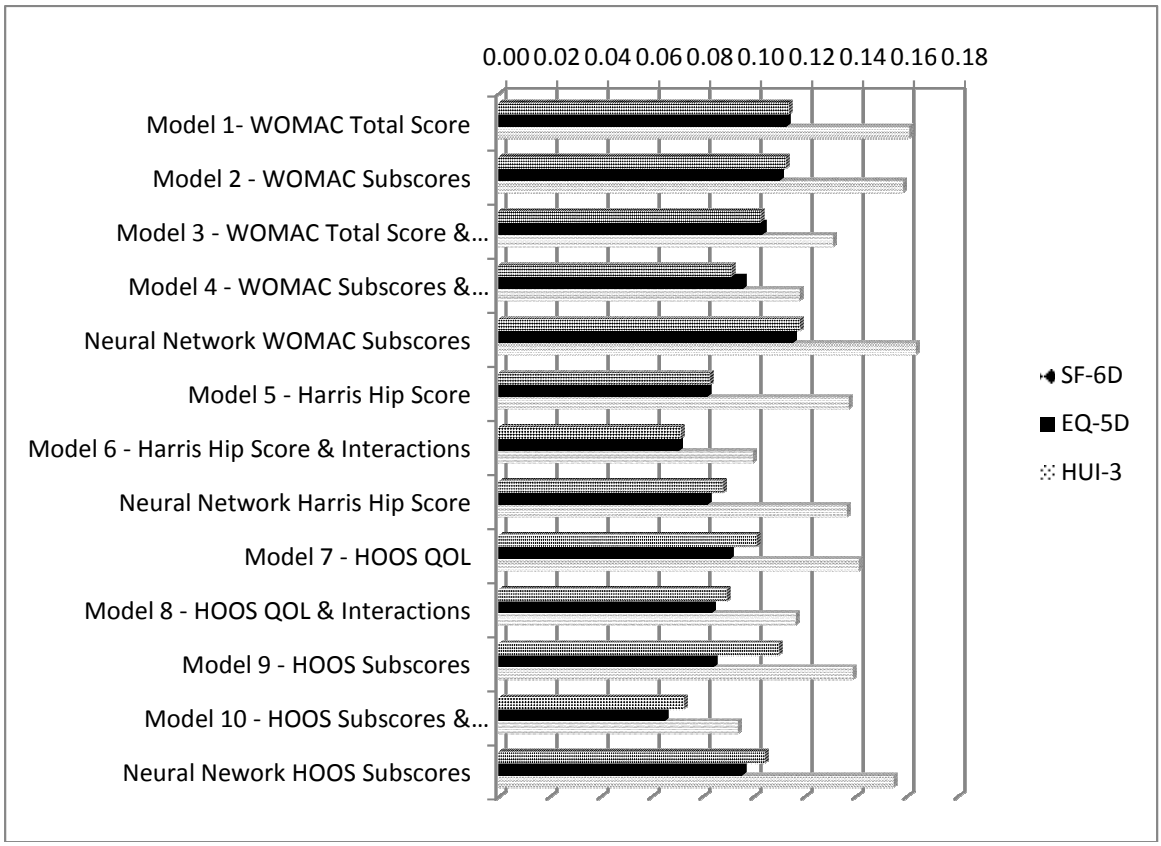

Figure 5. RMSE across all THA mapping models

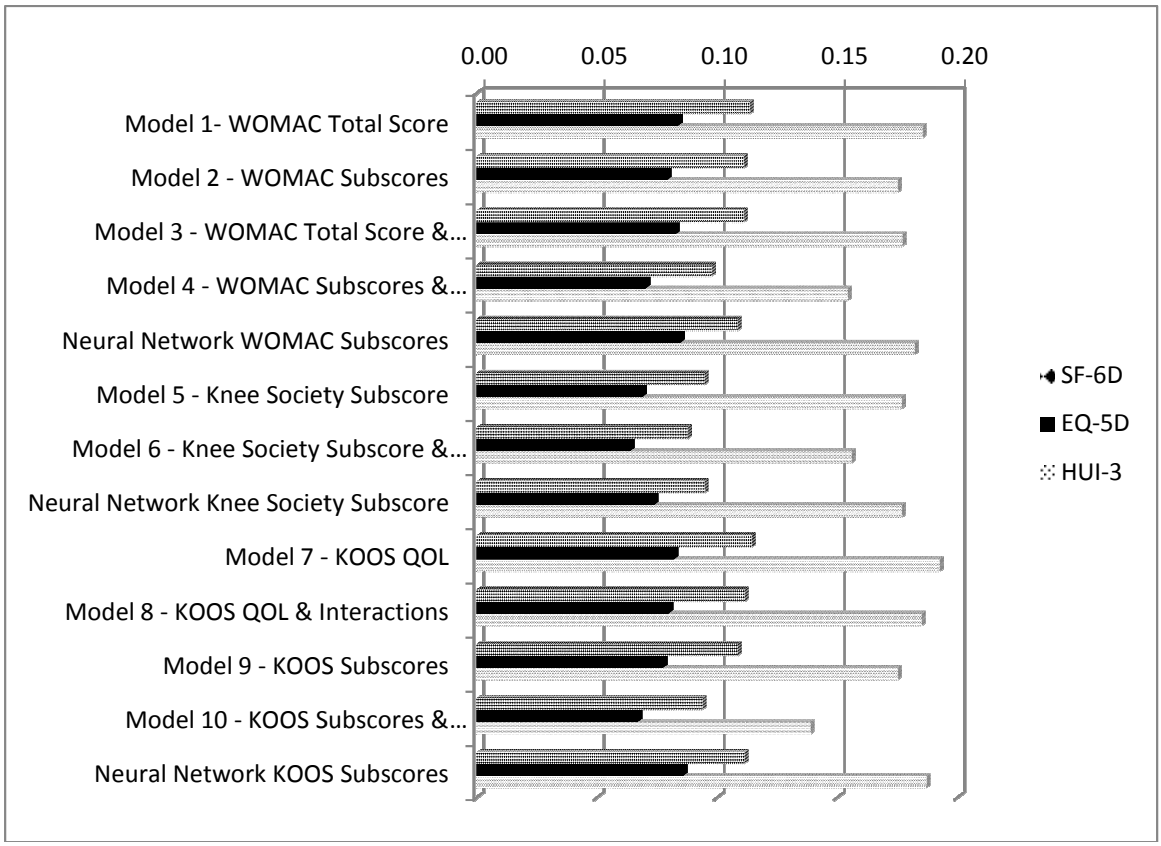

Figure 6. RMSE across all TKA mapping models

TABLES

Table 1. Subscales of the Disease Specific Measures and Preference-based Utility Measures

Questionnaire	Subscale
Disease Specific	
Knee Society Score (KSS)	Pain, Function
Harris Hip Score (HHS)	Pain, Function
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC®)	Pain, Mobility, Stiffness
Hip Disability and Osteoarthritis Outcomes Score (HOOS)	Pain, Symptoms, Daily Activity Limitations, Sport and Recreation Activity Limitations, Quality of Life
Knee Injury and Osteoarthritis Outcomes Score	
Generic, Prefence-Based	
EuroQOL 5D (EQ-5D)	Mobility, Self-care, Usual Activity, Pain and Discomfort, Anxiety and Depression
Health Utilities Index 3 (HUI3)	Vision, Hearing, Speech, Ambulation, Dexterity, Cognition, Pain, Emotion
Short Form 6D (SF-6D)	Role Physical, Bodily Pain,Vitality, Social Functioning, Role Emotional, Mental Health

Table 2. Questionnaire Ordered Blocks

Total Hip Arthroplasty	Block 1	SF-12, HOOS, EQ-5D, HHS, HUI3
Total Knee Arthroplasty	Block 1	SF-12, KOOS, EQ-5D, KSS, HUI3
Total Hip Arthroplasty	Block 2	EQ-5D, HOOS, HUI3, HHS, SF-12
Total Knee Arthroplasty	Block 2	EQ-5D, KOOS, HUI3, KSS, SF-12
Total Hip Arthroplasty	Block 3	HUI3, HOOS, SF-12, HHS, EQ-5D
Total Knee Arthroplasty	Block 3	HUI3, KOOS, SF-12, KSS, EQ-5D
Total Hip Arthroplasty	Block 4	SF-12, HHS, EQ-5D, HOOS, HUI3
Total Knee Arthroplasty	Block 4	SF-12, KSS, EQ-5D, KOOS, HUI3
Total Hip Arthroplasty	Block 5	EQ-5D, HHS, HUI3, HOOS, SF-12
Total Knee Arthroplasty	Block 5	EQ-5D, KSS, HUI3, KOOS, SF-12
Total Hip Arthroplasty	Block 6	HUI3, HHS, SF-12, HOOS, EQ-5D
Total Knee Arthroplasty	Block 6	HUI3, KSS, SF-12, KOOS, EQ-5D

Table 3. Completed Pairs of Disease-Specific and General, Preference-Based Measures

	Hip			Knee		
	HHS	HOOS	WOMAC®	KSS	KOOS	WOMAC®
EQ-5D	399	408	408	504	506	506
HUI 3	397	397	397	497	497	497
SF-6	404	410	410	505	505	505

Table 4. Transfer to Utility Regression Models

WOMAC® Regressed onto Utility

1) $\text{Utility} = \beta_0 + \beta_1\text{total} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \mu$
2) $\text{Utility} = \beta_0 + \beta_1\text{pain} + \beta_2\text{mobility} + \beta_3\text{stiffness} + \beta4\text{age} + \beta_5\text{gender} + \beta_6\text{years} + \mu$
3) $\text{Utility} = \beta_0 + \beta_1\text{total} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \beta_5\text{total*age} + \beta_6\text{total*gender} + \beta_7\text{total*years} + \beta_8\text{total}^2 + \beta_9\text{age}^2 + \beta_{10}\text{years}^2 + \mu$
4) $\text{Utility} = \beta_0 + \beta_1\text{pain} + \beta_2\text{mobility} + \beta_3\text{stiffness} + \beta4\text{age} + \beta_5\text{gender} + \beta_6\text{years} + \beta_7\text{pain*age} + \beta_8\text{pain*gender} + \beta_9\text{pain*years} + \beta_{10}\text{mobility*age} + \beta_{11}\text{mobility*gender} + \beta_{12}\text{mobility*years} + \beta_{13}\text{stiffness*age} + \beta_{14}\text{stiffness*gender} + \beta_{15}\text{stiffness*years} + \beta_{16}\text{pain}^2 + \beta_{17}\text{mobility}^2 + \beta_{18}\text{stiffness}^2 + \beta_{19}\text{age}^2 + \beta_{20}\text{years}^2 + \mu$

KSS/HHS Regressed onto Utility

1) $\text{Utility} = \beta_0 + \beta_1\text{total} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \mu$
2) $\text{Utility} = \beta_0 + \beta_1\text{pain} + \beta_2\text{function} + \beta_3\text{age} + \beta_4\text{gender} + \beta_5\text{years} + \mu$
3) $\text{Utility} = \beta_0 + \beta_1\text{total} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \beta_5\text{total*age} + \beta_6\text{total*gender} + \beta_7\text{total*years} + \beta_8\text{total}^2 + \beta_9\text{age}^2 + \beta_{10}\text{years}^2 + \mu$
4) $\text{Utility} = \beta_0 + \beta_1\text{pain} + \beta_2\text{function} + \beta_3\text{age} + \beta_4\text{gender} + \beta_5\text{years} + \beta_6\text{pain*age} + \beta_7\text{pain*gender} + \beta_8\text{pain*years} + \beta_9\text{function*age} + \beta_{10}\text{function*gender} + \beta_{11}\text{function*years} + \beta_{12}\text{pain}^2 + \beta_{13}\text{function}^2 + \beta_{14}\text{age}^2 + \beta_{15}\text{years}^2 + \mu$

KOOS/HOOS Regressed onto Utility

1) $\text{Utility} = \beta_0 + \beta_1\text{qualityoflife} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \mu$
2) $\text{Utility} = \beta_0 + \beta_1\text{pain} + \beta_2\text{ActivityDaily} + \beta_3\text{ActivitySport} + \beta_4\text{symptoms} + \beta_5\text{qualityoflife} + \beta_6\text{age} + \beta_7\text{gender} + \beta_8\text{years} + \mu$
3) $\text{Utility} = \beta_0 + \beta_1\text{total} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \beta_5\text{total*age} + \beta_6\text{total*gender} + \beta_7\text{total*years} + \beta_8\text{total}^2 + \beta_9\text{age}^2 + \beta_{10}\text{years}^2 + \mu$
4) $\text{Utility} = \beta_0 + \beta_1\text{qualityoflife} + \beta_2\text{age} + \beta_3\text{gender} + \beta_4\text{years} + \beta_5\text{qualityoflife*age} + \beta_6\text{qualityoflife*gender} + \beta_7\text{qualityoflife*years} + \beta_8\text{qualityoflife}^2 + \beta_9\text{age}^2 + \beta_{10}\text{years}^2 + \mu$
5) $\text{Utility} = \beta_0 + \beta_1\text{pain} + \beta_2\text{ActivityDaily} + \beta_3\text{ActivitySport} + \beta_4\text{symptoms} + \beta_5\text{qualityoflife} + \beta_6\text{age} + \beta_7\text{gender} + \beta_8\text{years} + \beta_9\text{pain*age} + \beta_{10}\text{pain*gender} + \beta_{11}\text{qualityoflife*age} + \beta_{12}\text{qualityoflife*gender} + \beta_{13}\text{qualityoflife*years} + \beta_{14}\text{pain*years} + \beta_{15}\text{ActivityDaily*age} + \beta_{16}\text{ActivityDaily*gender} + \beta_{17}\text{ActivityDaily*years} + \beta_{18}\text{ActivitySport*age} + \beta_{19}\text{ActivitySport*gender} + \beta_{20}\text{ActivitySport*years} + \beta_{21}\text{symptoms*age} + \beta_{22}\text{symptoms*gender} + \beta_{23}\text{symptoms*years} + \beta_{24}\text{pain}^2 + \beta_{25}\text{ActivityDaily}^2 + \beta_{26}\text{ActivitySport}^2 + \beta_{27}\text{symptoms}^2 + \beta_{28}\text{qualityoflife}^2 + \beta_{29}\text{age}^2 + \beta_{30}\text{years}^2 + \mu$

Table 5. Frequency and Proportions of Demographic Characteristics for THA Cohort

	Hip (n=437)	
	Freq	%
Gender		
Female	213	48.74
Male	224	51.26
Follow-up Period		
< 2 years	181	41.42
2 to 5 years	155	35.47
> 5 years	101	23.11
Age Group		
< 65 years	242	55.38
65 to 75 years	147	33.64
> 75 years	48	10.98

Table 6. Means and Standard Deviations for THA Cohort

	Hip		
	N	Mean	SD
Years Since Surgery at Time of Survey	435	3.66	3.63
Age at Time of Survey	435	62.97	10.75
SF-6D	419	0.6600*	0.26+
EQ-5D	408	0.7998*	0.20+
HUI-3	367	0.7536*	0.19+
HHS Pain	413	36.01	9.93
HHS Function	414	40.14	8.14
WOMAC Function	413	80.78	19.12
WOMAC Pain	416	82.99	19.22
WOMAC Stiffness	416	77.64	22.27
WOMAC TOTAL	411	81.12	19.11
HOOS FnADL	413	80.78	19.12
HOOS FnSRA	413	69.2	26.19
HOOS Pain	416	82.75	19.09
HOOS QOL	411	72.84	23.36
HOOS Symptoms	416	81.05	18.14

* Median values are reported
+ Interquartile range is reported

Table 7. Means and Standard Deviations for THA Cohort by Gender

	Hip		
	Female	Male	pvalue
Years Since Surgery at Time of Survey	3.48 (3.38)	3.83 (3.86)	0.316
Age at Time of Survey	61.96 (11.99)	63.92 (9.35)	0.059
SF-6D	.8590 (.26)*	.8170 (.26)*	0.874
EQ-5D	.8438 (.20)*	.8438 (.21)*	0.836
HUI-3	.9188 (.20)*	.8629 (.19)*	0.21
HHS Pain	36.54 (9.58)	35.50 (10.25)	0.285
HSS Function	40.04 (7.84)	40.23 (8.43)	0.821
WOMAC Function	82.10 (19.44)	79.54 (18.77)	0.174
WOMAC Pain	83.84 (19.86)	82.18 (18.60)	0.379
WOMAC Stiffness	78.02 (22.20)	77.29 (22.38)	0.739
WOMAC TOTAL	82.11 (19.38)	80.18 (18.84)	0.306
HOOS FnADL	82.10 (19.44)	79.54 (18.77)	0.174
HOOS FnSRA	70.96 (26.18)	67.54 (64.00)	0.186
HOOS Pain	84.16 (19.30)	81.41 (18.83)	0.142
HOOS QOL	74.81 (23.95)	70.95 (22.67)	0.094
HOOS Symptoms	81.67 (18.13)	80.45 (18.18)	0.491

* Median value and interquartile range reported

Table 8. Means and Standard Deviations for THA Cohort by Age Group

	Hip			
	< 65 years	65 to 75 years	> 75 years	
	Mean (SD)	Mean (SD)	Mean (SD)	Pvalue
SF-6D	.8170 (.26)*	.8590 (.22)*	.8085 (.22)*	0.17
EQ-5D	0.8438 (0.22)*	0.8603 (0.19)*	0.8540 (0.19)*	0.17
HUI-3	0.9054 (0.22)*	0.9047 (0.17)*	0.8543(0.36)*	0.26
HHS Pain	34.29 (10.83)	37.94 (7.92)	38.62 (9.35)	<0.0001
HHS Function	39.82 (8.73)	41.03 (6.34)	38.86 (9.86)	0.21
HHS Modified Total	81.40 (19.78)	86.99 (13.09)	85.45 (17.64)	0.01
WOMAC Function	79.52 (21.30)	81.62 (17.54)	84.71 (17.34)	0.21
WOMAC Pain	80.96 (20.92)	84.76 (17.02)	87.84 (15.23)	0.04
WOMAC Stiffness	75.38 (23.48)	79.55 (20.58)	83.24 (19.80)	0.05
WOMAC TOTAL	79.34 (20.50)	82.48 (17.25)	86.07 (16.21)	0.06
HOOS FnADL	79.52 (20.30)	81.62 (17.54)	84.71 (17.34)	0.21
HOOS FnSRA	67.54 (26.98)	70.38 (24.22)	74.13 (27.94)	0.26
HOOS Pain	80.50 (20.98)	84.81 (16.19)	87.78 (15.84)	0.02
HOOS QOL	68.50 (25.37)	76.95 (18.54)	82.27 (21.64)	<0.0001
HOOS Symptoms	78.78 (19.96)	83.29 (15.29)	85.57 (15.03)	0.01

* Median value and interquartile range reported

Table 9. Means and Standard Deviations for THA Cohort by Time Since Surgery

	Hip			
	< 2 years	2 to 5 years	> 5 years	
	Mean (SD)	Mean (SD)	Mean (SD)	Pvalue
SF-6D	0.8170 (0.26)*	0.8590 (0.24)*	0.8085 (0.26)*	0.19
EQ-5D	0.8435 (0.21)*	1.000 (0.20)*	0.8438 (0.20)*	0.32
HUI-3	0.8543 (0.22)*	0.9188 (0.17)*	0.8543 (0.22)*	0.06
HHS Pain	34.45 (10.29)	37.08 (9.66)	37.18 (9.37)	0.03
HHS Function	38.99 (8.73)	41.06 (7.93)	40.78 (7.10)	0.05
HHS Modified Total	80.85 (18.60)	85.88 (17.34)	85.75 (15.85)	0.02
WOMAC Function	78.55 (19.74)	82.79 (18.58)	81.65 (18.62)	0.13
WOMAC Pain	81.10 (19.71)	84.25 (18.88)	84.44 (18.75)	0.24
WOMAC Stiffness	74.13 (22.88)	80.48 (21.92)	79.60 (21.04)	0.02
WOMAC TOTAL	78.70 (19.58)	83.11 (18.78)	82.39 (18.50)	0.09
HOOS FnADL	78.55 (19.74)	82.79 (18.58)	81.65 (18.62)	0.13
HOOS FnSRA	64.68 (27.18)	73.80 (24.27)	70.15 (26.16)	0.01
HOOS Pain	80.76 (19.58)	84.25 (18.74)	84.03 (18.60)	0.2
HOOS QOL	69.71 (23.42)	75.78 (24.46)	73.90 (21.02)	0.06
HOOS Symptoms	79.39 (18.43)	82.64 (18.48)	81.58 (17.01)	0.27

* Median value and interquartile range reported

Table 10. Frequency and Proportions of Demographic Characteristics for TKA Cohort

	Knee (n=550)	
	Freq	%
Gender		
Female	291	52.91
Male	259	47.09
Follow-up Period		
< 2 years	264	48
2 to 5 years	192	34.91
> 5 years	94	17.09
Age Group		
< 65 years	270	49.09
65 to 75 years	211	38.36
> 75 years	69	12.55

Table 11. Means and Standard Deviations for TKA Cohort

	Knee		
	N	Mean	SD
Years Since Surgery at Time of Survey	548	2.89	2.5
Age at Time of Survey	548	65.3	8.28
SF-6D	521	0.6810*	0.24+
EQ-5D	506	0.7998*	0.20+
HUI-3	460	0.7056*	0.23+
KSS Pain	515	40.83	13.81
KSS Function	513	83.13	20.61
WOMAC Function	509	87.04	16.58
WOMAC Pain	517	87.73	16.2
WOMAC Stiffness	521	78.89	20.46
WOMAC TOTAL	509	85.76	16.07
KOOS FnADL	509	87.04	16.58
KOOS FnSRA	509	64.91	27.5
KOOS Pain	517	85.91	17.11
KOOS QOL	509	70.65	24.3
KOOS Symptoms	521	80.72	16.87

* Median value is reported
+ Interquartile range is reported

Table 12. Means and Standard Deviations for TKA Cohort by Gender

	Knee		
	Female	Male	pvalue
Years Since Surgery at Time of Survey	3.05 (2.70)	2.71 (2.25)	0.1044
Age at Time of Survey	64.89 (8.61)	65.76 (7.89)	0.2205
SF-6D	.79660 (0.18)*	.8380 (0.20)*	0.009
EQ-5D	.8271 (.21)*	.8438 (.18)*	0.025
HUI-3	.8543 (.22)*	.8794 (.22)*	0.270
KSS Pain	39.20 (14.50)	42.68 (12.76)	0.0039
KSS Function	79.17 (21.38)	87.59 (18.75)	<.0001
WOMAC Function	85.80 (17.26)	88.44 (15.71)	0.0729
WOMAC Pain	85.95 (17.37)	89.71 (14.56)	0.0077
WOMAC Stiffness	76.77 (21.80)	81.25 (18.62)	0.0118
WOMAC TOTAL	84.20 (16.86)	87.51 (14.98)	0.0205
KOOS FnADL	85.80 (17.26)	88.44 (15.71)	0.0729
KOOS FnSRA	61.69 (28.44)	68.52 (25.99)	0.0051
KOOS Pain	84.47 (18.09)	87.51 (15.83)	0.0422
KOOS QOL	69.49 (24.57)	71.95 (23.99)	0.2548
KOOS Symptoms	79.83 (17.56)	81.72 (16.05)	0.2019

* Median value and interquartile range reported

Table 13. Means and Standard Deviations for TKA Cohort by Age Group

	Knee			
	< 65 years	65 to 75 years	> 75 years	
	Mean (SD)	Mean (SD)	Mean (SD)	pvalue
SF-6D	.8000 (.25)*	.8000 (.20)*	.7680 (.1189)*	0.37
EQ-5D	0.8271 (0.20)*	0.8438 (0.18)*	0.8271 (0.22)*	0.19
HUI-3	0.9054 (0.22)*	0.8543 (0.19)*	0.8004 (.29)*	0.10
KSS Pain	40.12 (14.34)	41.12 (13.45)	42.74 (12.66)	0.38
KSS Function	84.26 (21.14)	83.43 (19.25)	77.58 (22.06)	0.07
WOMAC Function	87.21 (16.95)	87.19 (15.69)	85.90 (18.04)	0.84
WOMAC Pain	87.56 (17.08)	88.20 (14.76)	86.92 (17.13)	0.84
WOMAC Stiffness	78.62 (20.72)	79.86 (19.67)	76.89 (21.95)	0.57
WOMAC TOTAL	85.72 (16.61)	86.16 (15.12)	84.66 (17.04)	0.81
KOOS FnADL	87.21 (16.95)	87.19 (15.69)	85.90 (18.04)	0.84
KOOS FnSRA	64.21 (27.43)	65.68 (26.42)	65.24 (31.26)	0.85
KOOS Pain	85.23 (18.06)	86.67 (15.81)	86.20 (17.35)	0.66
KOOS QOL	68.75 (24.76)	72.46 (23.54)	72.42 (24.71)	0.23
KOOS Symptoms	78.85 (17.51)	82.04 (16.15)	83.82 (15.95)	0.04

* Median value and interquartile range is reported

Table 14. Means and Standard Deviations for TKA Cohort by Time Since Surgery

	Knee			
	< 2 years	2 to 5 years	> 5 years	
Variable	Mean (SD)	Mean (SD)	Mean (SD)	Pvalue
SF-6D	0.8000 (0.24)*	0.8000 (0.23)*	0.7910 (0.21)*	0.35
EQ-5D	0.8271 (0.22)*	0.8438 (0.18)*	0.8271 (0.20)*	0.34
HUI-3	0.8458 (0.25)*	0.8794 (0.21)*	0.8668 (0.17)*	0.44
KSS Pain	38.61 (15.06)	42.87 (11.83)	42.92 (12.99)	0.002
KSS Function	80.65 (22.11)	85.82 (19.21)	84.72 (18.12)	0.03
WOMAC Function	84.07 (17.44)	89.48 (15.75)	90.30 (14.42)	0.0005
WOMAC Pain	85.43 (17.03)	89.72 (15.04)	90.06 (15.39)	0.0081
WOMAC Stiffness	73.65 (20.97)	83.10 (18.24)	84.86 (19.87)	<0.0001
WOMAC TOTAL	82.61 (16.76)	88.48 (14.68)	88.97 (15.32)	0.0001
KOOS FnADL	84.07 (17.44)	89.48 (15.75)	90.30 (14.42)	0.0005
KOOS FnSRA	61.46 (28.22)	65.93 (26.93)	72.30 (25.19)	0.005
KOOS Pain	82.64 (18.49)	88.80 (14.98)	89.07 (15.64)	0.0002
KOOS QOL	65.79 (25.28)	74.22 (23.05)	76.83 (21.47)	<0.0001
KOOS Symptoms	76.65 (17.60)	84.52 (14.40)	84.33 (16.99)	<0.0001

* Median value and interquartile range is reported

Table 15. Comparison of Predictive Performance Criteria for the SF-6D and the THA Cohort

	n	MAE	Error >.1	RMSE
Estimation Sample				
Model 1- WOMAC Total Score	346	0.1040	43.4%	0.1248
Model 2 - WOMAC Subscores	346	0.1011	38.8%	0.1229
Model 3 - WOMAC Total Score & Interactions	346	0.0999	43.1%	0.1210
Model 4 - WOMAC Subscores & Interactions	346	0.0957	40.2%	0.1173
Neural Network WOMAC Subscores	346	0.1045		0.1245
Model 5 - Harris Hip Score	341	0.0838	28.3%	0.1052
Model 6 - Harris Hip Score & Interactions	341	0.0822	27.5%	0.1040
Neural Network Harris Hip Score	341	0.0838		0.1052
Model 7 - HOOS QOL	347	0.0991	38.8%	0.1205
Model 8 - HOOS QOL & Interactions	347	0.0952	37.2%	0.1163
Model 9 - HOOS Subscores	345	0.0958	36.9%	0.1173
Model 10 - HOOS Subscores & Interactions	345	0.0899	34.5%	0.1114
Neural Nework HOOS Subscores	345	0.0944		0.1177
Holdout Sample				
Model 1- WOMAC Total Score	65	0.0923	37.9%	0.1136
Model 2 - WOMAC Subscores	65	0.0934	36.4%	0.1123
Model 3 - WOMAC Total Score & Interactions	65	0.0840	30.3%	0.1025
Model 4 - WOMAC Subscores & Interactions	65	0.0742	25.8%	0.0912
Neural Network WOMAC Subscores	65	0.0971		0.1179
Model 5 - Harris Hip Score	62	0.0667	22.7%	0.0826
Model 6 - Harris Hip Score & Interactions	62	0.0574	15.2%	0.0711
Neural Network Harris Hip Score	62	0.0720		0.0878
Model 7 - HOOS QOL	64	0.0849	27.3%	0.1009
Model 8 - HOOS QOL & Interactions	64	0.0733	30.3%	0.0893
Model 9 - HOOS Subscores	64	0.0949	45.5%	0.1096

Table 15 (continued)

Model 10 - HOOS Subscores & Interactions	64	3.1398E-16	0.0561	19.7%	0.0724
Neural Network HOOS Subscores	64		0.0851		0.1040

Table 16. Comparison of Predictive Performance Criteria for the EQ-5D and the THA Cohort

	N	MAE	Error > .1	RMSE
Estimation Sample				
Model 1- WOMAC Total Score	342	0.10622	34.0%	0.13738
Model 2 - WOMAC Subscores	342	0.1049	34.2%	0.1361
Model 3 - WOMAC Total Score & Interactions	342	0.1058	35.8%	0.1306
Model 4 - WOMAC Subscores & Interactions	342	0.1030	38.8%	0.1267
Neural Network WOMAC Subscores	342	0.1071		0.1374
Model 5 - Harris Hip Score	337	0.0797	25.9%	0.1025
Model 6 - Harris Hip Score & Interactions	337	0.0797	24.5%	0.0993
Neural Network Harris Hip Score	337	0.0883		0.1199
Model 7 - HOOS QOL	344	0.0971	34.5%	0.1301
Model 8 - HOOS QOL & Interactions	342	0.0974	32.6%	0.1241
Model 9 - HOOS Subscores	342	0.0948	33.7%	0.1249
Model 10 - HOOS Subscores & Interactions	342	0.0903	27.5%	0.1137
Neural Network HOOS Subscores	342	0.1017		0.1350
Holdout Sample				
Model 1- WOMAC Total Score	64	0.0907	30.3%	0.11279
Model 2 - WOMAC Subscores	64	0.08671	33.3%	0.11021
Model 3 - WOMAC Total Score & Interactions	64	0.0828	25.8%	0.10337
Model 4 - WOMAC Subscores & Interactions	64	0.0750	27.3%	0.09558
Neural Network WOMAC Subscores	64	0.0950		0.11537
Model 5 - Harris Hip Score	62	0.0659	16.7%	0.0817
Model 6 - Harris Hip Score & Interactions	62	0.0551	12.1%	0.0705
Neural Network Harris Hip Score	62	0.0653		0.0818

Table 16 (continued)

Model 7 - HOOS QOL	64	0.0668	24.2%	0.0906
Model 8 - HOOS QOL & Interactions	64	0.0654	18.2%	0.0833
Model 9 - HOOS Subscores	64	0.0622	16.7%	0.0842
Model 10 - HOOS Subscores & Interactions	64	0.0522	12.1%	0.0649
Neural Network HOOS Subscores	64	0.0756		0.0954

Table 17. Comparison of Predictive Performance Criteria for the HUI-3 and the THA Cohort

	n	MAE	Error > .1	RMSE
Estimation Sample				
Model 1- WOMAC Total Score	308	0.13146	36.9%	0.18554
Model 2 - WOMAC Subscores	308	0.1290	38.5%	0.1826
Model 3 - WOMAC Total Score & Interactions	308	0.1249	40.2%	0.1747
Model 4 - WOMAC Subscores & Interactions	308	0.1216	39.4%	0.1676
Neural Network WOMAC Subscores	308	0.1333		0.1851
Model 5 - Harris Hip Score	304	0.0929	25.9%	0.1375
Model 6 - Harris Hip Score & Interactions	304	0.0896	24.3%	0.1333
Neural Network Harris Hip Score	304	0.1114		0.1606
Model 7 - HOOS QOL	310	0.12672	40.2%	0.17933
Model 8 - HOOS QOL & Interactions	310	0.1211	36.1%	0.1692
Model 9 - HOOS Subscores	308	0.1244	38.5%	0.1751
Model 10 - HOOS Subscores & Interactions	308	0.1133	35.3%	0.1577
Neural Network HOOS Subscores	308	0.1312		0.1818
Holdout Sample				
Model 1- WOMAC Total Score	57	0.1244	43.9%	0.1608
Model 2 - WOMAC Subscores	57	0.1217	45.5%	0.1583
Model 3 - WOMAC Total Score & Interactions	57	0.1001	33.3%	0.1309
Model 4 - WOMAC Subscores & Interactions	57	0.0901	30.3%	0.1177
Neural Network WOMAC Subscores	57	0.1260		0.1634
Model 5 - Harris Hip Score	57	0.0990	27.3%	0.1370
Model 6 - Harris Hip Score & Interactions	57	0.0747	18.2%	0.0994
Neural Network Harris Hip Score	55	0.0990		0.1364
Model 7 - HOOS QOL	57	0.10728	34.8%	0.14088
Model 8 - HOOS QOL & Interactions	57	0.0833	30.3%	0.1162

Table 17 (continued)

Model 9 - HOOS Subscores	57	0.1023	33.3%	0.1388
Model 10 - HOOS Subscores & Interactions	57	0.0657	16.7%	0.0938
Neural Network HOOS Subscores	57	0.1137		0.1547

Table 18. Best Performing THA Models and Recommendations for Mapping Pairs

	SF-6D	EQ-5D	HUI-3	Preferred Utility Measure
WOMAC	Model 4 Group only	Model 4 Group only	Model 4 Group only	SF-6D
Harris Hip	Model 6 Individual & Group	Model 6 Individual & Group	Model 6 Group only	EQ-5D
HOOS	Model 10 Group only	Model 10 Individual & Group	Model 10 Group only	EQ-5D
Preferred Disease Specific	Harris Hip	Harris Hip / HOOS	Harris Hip / HOOS	

Table 19. Coefficient Estimates for Best Performing Models Estimating SF-6D

WOMAC Model 4 Variable	WOMAC Model 4 Coefficient Estimate	Harris Hip Model 6 Variable	Harris Hip Model 6 Coefficient Estimate	HOOS Model 10 Variable	HOOS Model 10 Coefficient Estimate
R^2	0.3053	R^2	0.4545	R^2	0.3729
Intercept	-0.34484	Intercept	0.06275	Intercept	-0.07232
Pain	0.01604^	HHS Pain	0.01039	FnDL	-0.00587
Function	-0.0095	HHS Function	0.00454	FnSRA	0.00354
Stiffness	0.00694^	Age	0.00664	Pain	0.00943
Age	0.01633^	Gender	-0.04104	QOL	0.00589^
Gender	-0.10882	Followup	0.0005755	Symptoms	-0.0001377
Followup	0.00061898	Pain * Age	-0.0001564	Age	0.01053
Pain * Age	-0.00016158	Pain * Followup	0.0001366	Gender	-0.09771
Pain * Gender	0.00129	Pain * Gender	0.00146	Followup	0.00537
Pain * Followup	0.00006254	Function*Age	6.96E-06	Pain * Age	-0.0000738
Function * Age	0.0001744	Function * Followup	-0.000165	Pain * Gender	0.0016
Function * Gender	-0.00015343	Function * Gender	-0.0005784	Pain * Followup	0.0003259
Function * Followup	-0.00007699	Pain * Pain	1.684E-05	Pain * Pain	-0.0000452
Stiffness * Age	-0.00012813^	Function * Function	6.477E-05	FnDL * Age	0.0000677
Stiffness * Gender	0.00009293	Age * Age	-6.23E-06	FnDL * Gender	0.0009557
Stiffness * Followup	0.00001421	Followup * Followup	-7.31E-06	FnDL * Followup	-0.0005052
Pain * Pain	-0.00004944			FnDL * FnDL	0.0000272
Function * Function	0.00002386			FnSRA * Age	-0.0000128
Stiffness * Stiffness	7.86E-07			FnSRA * Gender	-0.0008074
Age * Age	-0.00005283			FnSRA * Followup	0.0001330

Table 19 (continued)

Followup * Followup	-0.00006199			FnSRA * FnSRA	-0.0000168
				QOL * Age	-0.0000918
				QOL * Gender	0.0008835
				QOL * Followup	-0.0000258
				QOL * QOL	0.00000898
				Symptoms * Age	0.0000156
				Symptoms * Gender	-0.00157
				Symptoms * Followup	0.0000391
				Symptoms * Symptoms	-0.0000018
				Age * Age	-0.0000329
				Followup * Followup	-0.0001578

^ Statistical significance $p < .05$

Table 20. Coefficient Estimates for Best Performing Models Estimating EQ-5D

WOMAC Model 4 Variable	WOMAC Model 4 Coefficient Estimate	Harris Hip Model 6 Variable	Harris Hip Model 6 Coefficient Estimate	HOOS Model 10 Variable	HOOS Model 10 Coefficient Estimate
R^2	.3789	R^2	.6181	R^2	.4999
Intercept	-0.51514	Intercept	0.15452	Intercept	-0.09453
Pain	0.01374	HHS Pain	0.01692^	FnDL	0.00382
Function	0.00227	HHS Function	0.00927	FnSRA	0.00149
Stiffness	0.00248	Age	-0.000251	Pain	0.01166
Age	0.01778^	Gender	-0.046	QOL	0.00662
Gender	-0.15935^	Followup	-0.00673	Symptoms	-0.007
Followup	0.00644	Pain * Age	-0.000116	Age	0.00924
Pain * Age	-0.00007262	Pain * Followup	-0.000221	Gender	-0.0975
Pain * Gender	0.00491^	Pain * Gender	0.00404^	Followup	-0.00038357
Pain * Followup	0.00034801	Function*Age	8.199E-05	Pain * Age	-0.0000803
Function * Age	0.00000524	Function * Followup	0.0003064	Pain * Gender	0.004
Function * Gender	-0.00434	Function * Gender	-0.00271	Pain * Followup	0.00064539
Function * Followup	-0.00064623	Pain * Pain	-8.99E-05	Pain * Pain	-0.00005836
Stiffness * Age	-0.00005096	Function * Function	-6.54E-05	FnDL * Age	-0.00007325
Stiffness * Gender	0.00121	Age * Age	0.000011	FnDL * Gender	-0.00411
Stiffness * Followup	0.00024734	Followup * Followup	0.0001277	FnDL * Followup	-0.00102
Pain * Pain	-0.00007335			FnDL * FnDL	0.0000386
Function * Function	0.00003213			FnSRA * Age	-0.00001707
Stiffness * Stiffness	-0.00001225			FnSRA * Gender	0.00094022
Age * Age	-0.00006463			FnSRA * Followup	0.00009245
Followup * Followup	-0.00013706			FnSRA * FnSRA	-0.00000731

Table 20 (continued)

				QOL * Age	-0.00002745
				QOL * Gender	-0.00058718
				QOL * Followup	-0.00010118
				QOL * QOL	-0.00001583
				Symptoms * Age	0.00016249
				Symptoms * Gender	0.00093174
				Symptoms * Followup	0.00040434
				Symptoms * Symptoms	-0.00003168
				Age * Age	-0.00006077
				Followup * Followup	-0.00021263

^ Statistical significance $p < .05$

Table 21. Coefficient Estimates for Best Performing Models Estimating HUI-3

WOMAC Model 4 Variable	WOMAC Model 4 Coefficient Estimate	Harris Hip Model 6 Variable	Harris Hip Model 6 Coefficient Estimate	HOOS Model 10 Variable	HOOS Model 10 Coefficient Estimate
R^2	.3596	R^2	.5855	R^2	.4327
Intercept	-0.69489	Intercept	-0.16269	Intercept	-0.43747
Pain	0.03542^	HHS Pain	0.02558^	FnDL	-0.02289^
Function	-0.01266	HHS Function	0.01874	FnSRA	0.00742
Stiffness	0.00591	Age	-0.0009344	Pain	0.02693^
Age	0.01285	Gender	-0.01153	QOL	0.00731
Gender	-0.04806	Followup	-0.01324	Symptoms	0.01095
Followup	-0.01865	Pain * Age	-0.0002025	Age	0.0079
Pain * Age	-0.00023851	Pain * Followup	0.0005105	Gender	-0.0706
Pain * Gender	-0.00297	Pain * Gender	-0.00282	Followup	-0.01951
Pain * Followup	-0.00022107	Function*Age	0.00011522	Pain * Age	-0.000298
Function * Age	0.00025384	Function * Followup	-0.0002155	Pain * Gender	-0.00159
Function * Gender	-0.00009856	Function * Gender	0.00167	Pain * Followup	-4.36E-05
Function * Followup	0.00025739	Pain * Pain	-0.0001594	Pain * Pain	-5.07E-05
Stiffness * Age	-0.00016855	Function * Function	-0.0001527	FnDL * Age	0.0003783^
Stiffness * Gender	0.0033	Age * Age	0.00000996	FnDL * Gender	0.00128
Stiffness * Followup	0.00020134	Followup * Followup	0.00005754	FnDL * Followup	0.0001057
Pain * Pain	-0.00012567			FnDL * FnDL	9.99E-07
Function * Function	0.00001196			FnSRA * Age	-5.23E-05
Stiffness * Stiffness	0.00000758			FnSRA * Gender	-0.000582
Age * Age	-0.00002089			FnSRA * Followup	-0.000193

Table 21 (continued)

Followup * Followup	-0.00003164			FnSRA * FnSRA	-2.63E-05
				QOL * Age	-2.82E-05
				QOL * Gender	-0.00234
				QOL * Followup	0.0002797
				QOL * QOL	-2.16E-05
				Symptoms * Age	-0.000117
				Symptoms * Gender	0.00322
				Symptoms * Followup	0.0001116
				Symptoms * Symptoms	-3.53E-05
				Age * Age	-1.71E-05
				Followup * Followup	-0.000234

^ Statistical significance $p < .05$

Table 22. Comparison of Predictive Performance Criteria for the SF-6D and the TKA Cohort

	n	MAE	Error > .1	RMSE
Estimation Sample				
Model 1- WOMAC Total Score	430	0.0915	37.7%	0.1075
Model 2 - WOMAC Subscores	430	0.0890	34.5%	0.1053
Model 3 - WOMAC Total Score & Interactions	430	0.0902	37.9%	0.1065
Model 4 - WOMAC Subscores & Interactions	430	0.0865	33.6%	0.1032
Neural Network WOMAC Subscores	430	0.0895		0.1060
Model 5 - Knee Society Subscore	429	0.0873	33.2%	0.1046
Model 6 - Knee Society Subscore & Interactions	429	0.0863	33.0%	0.1038
Neural Network Knee Society Subscore	429	0.0877		0.1049
Model 7 - KOOS QOL	430	0.0927	38.5%	0.1105
Model 8 - KOOS QOL & Interactions	430	0.0912	35.5%	0.1088
Model 9 - KOOS Subscores	430	0.0877	33.2%	0.1043
Model 10 - KOOS Subscores & Interactions	430	0.0836	32.8%	0.0997
Neural Network KOOS Subscores	430	0.0857		0.1027
Holdout Sample				
Model 1- WOMAC Total Score	76	0.0893	28.9%	0.1135
Model 2 - WOMAC Subscores	76	0.0883	31.3%	0.1109
Model 3 - WOMAC Total Score & Interactions	76	0.0851	28.9%	0.1110
Model 4 - WOMAC Subscores & Interactions	76	0.0750	21.7%	0.0974
Neural Network WOMAC Subscores	76	0.0857		0.1085
Model 5 - Knee Society Subscore	77	0.0781	28.9%	0.0945
Model 6 - Knee Society Subscore & Interactions	77	0.0709	19.3%	0.0876
Neural Network Knee Society Subscore	77	0.0801		0.0946

Table 22 (continued)

Model 7 - KOOS QOL	76	0.0865	30.1%	0.1142
Model 8 - KOOS QOL & Interactions	76	0.0866	26.5%	0.1113
Model 9 - KOOS Subscores	76	0.0863	31.3%	0.1084
Model 10 - KOOS Subscores & Interactions	76	0.0735	25.3%	0.0936
Neural Network KOOS Subscores	76	0.0859		0.1110

Table 23. Comparison of Predictive Performance Criteria for the EQ-5D and the TKA Cohort

	n	MAE	Error > .1	RMSE
Estimation Sample				
Model 1- WOMAC Total Score	430	0.0847	28.9%	0.1087
Model 2 - WOMAC Subscores	430	0.0847	25.5%	0.1064
Model 3 - WOMAC Total Score & Interactions	430	0.0829	25.5%	0.1060
Model 4 - WOMAC Subscores & Interactions	430	0.0810	23.1%	0.1025
Neural Network WOMAC Subscores	430	0.0852		0.1070
Model 5 - Knee Society Subscore	428	0.0816	27.2%	0.1030
Model 6 - Knee Society Subscore & Interactions	428	0.0812	25.3%	0.1020
Neural Network Knee Society Subscore	428	0.0806		0.1105
Model 7 - KOOS QOL	430	0.0879	31.5%	0.1169
Model 8 - KOOS QOL & Interactions	430	0.0884	29.3%	0.1130
Model 9 - KOOS Subscores	430	0.0834	25.5%	0.1060
Model 10 - KOOS Subscores & Interactions	430	0.0790	25.5%	0.1008
Neural Network KOOS Subscores	430	0.0834		0.1053
Holdout Sample				
Model 1- WOMAC Total Score	76	0.0712	19.3%	0.0837
Model 2 - WOMAC Subscores	76	0.0677	13.3%	0.0792
Model 3 - WOMAC Total Score & Interactions	76	0.0702	15.7%	0.0827
Model 4 - WOMAC Subscores & Interactions	76	0.0554	9.6%	0.0703
Neural Network WOMAC Subscores	76	0.0754		0.0847
Model 5 - Knee Society Subscore	76	0.0576	15.7%	0.0690
Model 6 - Knee Society Subscore & Interactions	76	0.0505	12.0%	0.0639
Neural Network Knee Society Subscore	76	0.0591		0.0737
Model 7 - KOOS QOL	76	0.0640	16.9%	0.0820
Model 8 - KOOS QOL & Interactions	76	0.0644	18.1%	0.0800

Table 23 (continued)

Model 9 - KOOS Subscores	76	0.0643	18.1%	0.0777
Model 10 - KOOS Subscores & Interactions	76	0.0511	13.3%	0.0671
Neural Network KOOS Subscores	76	0.0762		0.0858

Table 24. Comparison of Predictive Performance Criteria for the HUI-3 and the TKA Cohort

Estimation Sample	n	MAE	Error > .1	RMSE
Model 1- WOMAC Total Score	389	0.1157	38.1%	0.1520
Model 2 - WOMAC Subscores	389	0.1134	37.5%	0.1511
Model 3 - WOMAC Total Score & Interactions	389	0.1142	36.2%	0.1505
Model 4 - WOMAC Subscores & Interactions	389	0.1086	33.8%	0.1465
Neural Network WOMAC Subscores	389	0.1074		0.1449
Model 5 - Knee Society Subscore	389	0.1050	30.2%	0.1420
Model 6 - Knee Society Subscore & Interactions	389	0.1028	30.6%	0.1386
Neural Network Knee Society Subscore	389	0.1047		0.1408
Model 7 - KOOS QOL	389	0.1185	37.5%	0.1560
Model 8 - KOOS QOL & Interactions	389	0.1158	36.2%	0.1531
Model 9 - KOOS Subscores	389	0.1108	34.0%	0.1496
Model 10 - KOOS Subscores & Interactions	389	0.1061	31.3%	0.1439
Neural Network KOOS Subscores	389	0.1041		0.1397
Holdout Sample				
Model 1- WOMAC Total Score	71	0.1351	42.2%	0.1853
Model 2 - WOMAC Subscores	71	0.1256	42.2%	0.1750
Model 3 - WOMAC Total Score & Interactions	71	0.1349	48.2%	0.1769
Model 4 - WOMAC Subscores & Interactions	71	0.1067	32.5%	0.1542
Neural Network WOMAC Subscores	71	0.1280		0.1819
Model 5 - Knee Society Subscore	71	0.1236	37.3%	0.1765
Model 6 - Knee Society Subscore & Interactions	71	0.1119	34.9%	0.1558
Neural Network Knee Society Subscore	71	0.1191		0.1767
Model 7 - KOOS QOL	71	0.1375	41.0%	0.1926

Table 24 (continued)

Model 8 - KOOS QOL & Interactions	71	0.1359	43.4%	0.1851
Model 9 - KOOS Subscores	71	0.1253	34.9%	0.1750
Model 10 - KOOS Subscores & Interactions	71	0.0989	34.9%	0.1387
Neural Network KOOS Subscores	71	0.1283		0.1868

Table 25. Best Performing TKA Models and Recommendations for Mapping Pairs

	SF-6D	EQ-5D	HUI-3	Preferred Utility Measure
WOMAC	Model 4 Group only	Model 4 Individual & Group	Model 4 Group only	EQ-5D
Harris Hip	Model 6 Group Only	Model 6 Individual & Group	Model 6 Group only	EQ-5D
HOOS	Model 10 Group only	Model 10 Individual & Group	Model 10 Group only	EQ-5D
Preferred Disease Specific	Knee Society	All Very Close	All Very Close	

Table 26. Coefficient Estimates for Best Performing Models Estimating SF-6D

WOMAC Model 4 Variable	WOMAC Model 4 Coefficient Estimate	Knee Society Model 6 Variable	Knee Society Model 6 Coefficient Estimate	KOOS Model 10 Variable	KOOS Model 10 Coefficient Estimate
R^2	.3502	R^2	.3458	R^2	.3937
Intercept	0.24165	Intercept	0.75456	Intercept	0.72149
Pain	-0.00985	KS Pain	0.00156	FnDL	-0.0019
Function	0.0096	KS Function	0.00139	FnSRA	0.0046
Stiffness	0.00352	Age	-0.00783	Pain	-0.01397
Age	0.00448	Gender	-0.00765	QOL	0.00652
Gender	0.02829	Followup	0.0093	Symptoms	0.01069^
Followup	0.01829	Pain * Age	0.0000021	Age	-0.00905
Pain * Age	0.00009899	Pain * Followup	-0.000225	Gender	-0.00685
Pain * Gender	0.00053116	Pain * Gender	0.00122	Followup	0.02229
Pain * Followup	0.00024302	Function * Age	0.00000599	Pain * Age	8.498E-05
Function * Age	-0.00005461	Function * Followup	-3.735E-05	Pain * Followup	0.00153^
Function * Gender	-0.00058954	Function * Gender	-0.0005275	Pain * Gender	0.0007748
Function * Followup	-0.00053807	Pain * Pain	0.00000395	Pain * Pain	3.366E-05
Stiffness * Age	-0.00008981	Function * Function	0.00001039	FnDL * Age	0.0001522
Stiffness * Gender	-0.00005338	Age * Age	0.00005873	FnDL * Gender	-0.00197
Stiffness * Followup	-0.00000716	Followup * Followup	0.00013034	FnDL * Followup	-0.000812^
Pain * Pain	0.00001275			FnDL * FnDL	-6.99E-06
Function * Function	0.00000628			FnSRA * Age	-6.1E-05
Stiffness * Stiffness	0.00001569			FnSRA * Gender	-0.000632
Age * Age	-0.0000076			FnSRA * Followup	0.0002105

Table 26 (continued)

Followup * Followup	0.00043385			FnSRA * FnSRA	-2.14E-06
				QOL * Age	-0.000107^
				QOL * Gender	0.0007698
				QOL * Followup	-0.000389^
				QOL * QOL	9.98E-06
				Symptoms * Age	-0.000116
				Symptoms * Gender	0.0005989
				Symptoms * Followup	-0.000158
				Symptoms * Symptoms	-1.78E-05
				Age * Age	7.226E-05
				Followup * Followup	0.0004987

^ Statistical significance $p < .05$

Table 27. Coefficient Estimates for Best Performing Models Estimating EQ-5D

WOMAC Model 4 Variable	WOMAC Model 4 Coefficient Estimate	Knee Society Model 6 Variable	Knee Society Model 6 Coefficient Estimate	KOOS Model 10 Variable	KOOS Model 10 Coefficient Estimate
R^2	.5050	R^2	.5021	R^2	.5238
Intercept	-0.69812	Intercept	1.55973	Intercept	-0.38483
Pain	-0.01315	KS Pain^	-0.00163	FnDL	0.01855^
Function	0.03007^	KS Function	-0.00412	FnSRA	0.00391
Stiffness	-0.00251	Age	-0.02275	Pain	-0.0165^
Age	0.01786^	Gender	-0.14146	QOL	0.00322
Gender	0.10322	Followup	0.01842	Symptoms	0.00784
Followup	0.00417	Pain * Age	-4.253E-05	Age	0.00963
Pain * Age	0.00015205	Pain * Followup	-9.226E-05	Gender	0.08534
Pain * Gender	-0.00103	Pain * Gender	0.00117	Followup	-0.01693
Pain * Followup	-0.00031612	Function * Age	0.00013491	Pain * Age	0.0001742
Function * Age	-0.00027657^	Function * Followup	-0.0002163	Pain * Followup	-0.00091
Function * Gender	-0.00008981	Function * Gender	0.00076288	Pain * Gender	4.796E-05
Function * Followup	-0.00008639	Pain * Pain	0.00012672	Pain * Pain	3.612E-05
Stiffness * Age	-0.00001243	Function * Function	-1.105E-05	FnDL * Age	-0.000129
Stiffness * Gender	0.00009724	Age * Age	0.0001089	FnDL * Gender	-0.00142
Stiffness * Followup	0.00019987	Followup * Followup	0.00051714	FnDL * Followup	6.292E-05
Pain * Pain	0.00003086			FnDL * FnDL	-3.06E-05
Function * Function	-0.00003795^			FnSRA * Age	-6.91E-05

Table 27 (continued)

Stiffness * Stiffness	0.00001942			FnSRA * Gender	-0.000489
Age * Age	-0.0000373			FnSRA * Followup	0.0002924
Followup * Followup	0.00106			FnSRA * FnSRA	-1.91E-06
				QOL * Age	-0.000257
				QOL * Gender	-0.00128
				QOL * Followup	0.0003846
				QOL * QOL	1.659E-05
				Symptoms * Age	0.0002285
				Symptoms * Gender	0.00327
				Symptoms * Followup	0.0004765
				Symptoms * Symptoms	-6.59E-05
				Age * Age	5.329E-05
				Followup * Followup	-0.00326^

^ Statistical significance $p < .05$

Table 28. Coefficient Estimates for Best Performing Models Estimating HUI-3

WOMAC Model 4 Variable	WOMAC Model 4 Coefficient Estimate	Knee Society Model 6 Variable	Knee Society Model 6 Coefficient Estimate	KOOS Model 10 Variable	KOOS Model 10 Coefficient Estimate
R^2	.3225	R^2	.3932	R^2	.3459
Intercept	-1.08313	Intercept	-0.31597	Intercept	-0.80949
Pain	-0.00262	KS Pain	0.02329	FnDL	0.0141
Function	0.01787	KS Function	-0.0007454	FnSRA	0.00089262
Stiffness	-0.00464	Age	0.02247^	Pain	-0.00825
Age	0.03646^	Gender	-0.34406	QOL	0.00912
Gender	0.13007	Followup	-0.01135	Symptoms	-0.00268
Followup	-0.02687	Pain * Age	-0.00005973	Age	0.03087^
Pain * Age	0.00006775	Pain * Followup	0.00247^	Gender	0.0687
Pain * Gender	-0.00271	Pain * Gender	-0.00811	Followup	-0.05002^
Pain * Followup	-0.00038595	Function * Age	0.00015118	Pain * Age	0.00015231
Function * Age	-0.00023242	Function * Followup	-0.00141^	Pain * Followup	-0.0026
Function * Gender	0.00331	Function * Gender	0.00772	Pain * Gender	0.00003069
Function * Followup	0.00007207	Pain * Pain	-0.00029971	Pain * Pain	-0.00001626
Stiffness * Age	0.00006185	Function * Function	-0.00002504	FnDL * Age	-0.00020313
Stiffness * Gender	-0.00223	Age * Age	-0.0002583^	FnDL * Gender	0.00164
Stiffness * Followup	0.00051683	Followup * Followup	0.00266	FnDL * Followup	0.00058007
Pain * Pain	-0.00000312			FnDL * FnDL	0.00001528
Function * Function	0.00001668			FnSRA * Age	0.00002924
Stiffness * Stiffness	0.00000111			FnSRA * Gender	-0.00093486

Table 28 (continued)

Age * Age	-0.00021262			FnSRA * Followup	-0.00047149
Followup * Followup	0.00051208			FnSRA * FnSRA	-0.00000183
				QOL * Age	-0.00013392
				QOL * Gender	-0.00033242
				QOL * Followup	0.00036461
				QOL * QOL	8.53E-07
				Symptoms * Age	0.00005089
				Symptoms * Gender	0.00124
				Symptoms * Followup	-0.00011033
				Symptoms * Symptoms	-0.00000957
				Age * Age	-0.00017994^
				Followup * Followup	0.00018633

^ Statistical significance $p < .05$

REFERENCES

1. Centers for Disease Control. National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions-United States, 2003. MMWR. 2007;56(1):4-7.

2. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am. Apr 2007;89(4):780-785.

3. Bozic KJ, Rosenberg AG, Huckman RS, Herndon JH. Economic evaluation in orthopaedics. J Bone Joint Surg Am. Jan 2003;85-A(1):129-142.

4. Bozic KJ, Saleh KJ, Rosenberg AG, Rubash HE. Economic evaluation in total hip arthroplasty: analysis and review of the literature. J Arthroplasty. Feb 2004;19(2):180-189.

5. Brauer CA, Rosen AB, Olchanski NV, Neumann PJ. Cost-utility analyses in orthopaedic surgery. J Bone Joint Surg Am. Jun 2005;87(6):1253-1259.

6. Singh J, Sloan JA, Johanson NA. Challenges With Health-related Quality of Life Assessment in Arthroplasty Patients: Problems and Solutions. J Am Acad Orthop Sur. Feb 2010;18(2):72-82.

7. Riddle DL, Stratford PW, Singh JA, Strand CV. Variation in outcome measures in hip and knee arthroplasty clinical trials: a proposed approach to achieving consensus. J Rheumatol. Sep 2009;36(9):2050-2056.

8. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. J Bone Joint Surg Am. May 2004;86-A(5):963-974.

9. Bombardier C, Melfi CA, Paul J, et al. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Med Care. Apr 1995;33(4 Suppl):AS131-144.

10. Boardman DL, Dorey F, Thomas BJ, Lieberman JR. The accuracy of assessing total hip arthroplasty outcomes: a prospective correlation study of walking ability and 2 validated measurement devices. J Arthroplasty. Feb 2000;15(2):200-204.
11. Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol. Mar 1997;50(3):239-246.

12. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. Clin Orthop Relat Res. Nov 1989(248):13-14.

13. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. J Bone Joint Surg Am. Dec 2001;83-A(12):1856-1864.

14. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am. Jun 1969;51(4):737-755.

15. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scand J Rheumatol. 2003;32(1):46-51.

16. Nilsdotter AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. BMC Musculoskelet Disord. May 30 2003;4:10.

17. Davis AM, Perruccio AV, Canizares M, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis Cartilage. Jul 2009;17(7):843-847.

18. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports. Dec 1998;8(6):439-448.

19. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther. Aug 1998;28(2):88-96.

20. Group TE. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. Dec 1990;16(3):199-208.

21. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics. Jan 2000;17(1):13-35.

22. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol. Nov 1998;51(11):1115-1128.

23. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. Mar 2002;21(2):271-292.

24. Tolley K. What are health utilities? What is...? series. 2009. http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Health-util.pdf.

25. Phillips C. What is a QALY? What is...? series. 2009. http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/QALY.pdf.

26. Organization WH. Towards a Common Language for Functioning, Disability and Health. 2002:1-22. http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf. Accessed November 1, 2010.

27. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health Utilities Index. Pharmacoeconomics. Jun 1995;7(6):503-520.

28. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. Arthritis Rheum. Jun 15 2008;59(6):876-883.

29. McGrory BJ, Freiberg AA, Shinar AA, Harris WH. Correlation of measured range of hip motion following total hip arthroplasty and responses to a questionnaire. J Arthroplasty. Aug 1996;11(5):565-571.

30. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes. 2003;1:54.

31. QualityMetric. SF-12v2 Health Survey. http://www.qualitymetric.com/WhatWeDo/GenericHealthSurveys/SF12v2HealthSurvey/tabid/186/Default.aspx.

32. Barton GR, Sach TH, Avery AJ, Doherty M, Jenkinson C, Muir KR. Comparing the performance of the EQ-5D and SF-6D when measuring the benefits of alleviating knee pain. Cost Eff Resour Alloc. 2009;7:12.

33. Blanchard C, Feeny D, Mahon JL, et al. Is the Health Utilities Index valid in total hip arthroplasty patients? Qual Life Res. Mar 2004;13(2):339-348.

34. Mortimer D, Segal L, Sturm J. Can we derive an 'exchange rate' between descriptive and preference-based outcome measures for stroke? Results from the transfer to utility (TTU) technique. Health Qual Life Outcomes. 2009;7:33.

35. Brazier JE, Kolotkin RL, Crosby RD, Williams GR. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. Value Health. Jul-Aug 2004;7(4):490-498.

36. Brennan DS, Spencer AJ. Mapping oral health related quality of life to generic health state values. BMC health services research. 2006;6:96.

37. Kulkarni AV. Distribution-based and anchor-based approaches provided different interpretability estimates for the Hydrocephalus Outcome Questionnaire. J Clin Epidemiol. Feb 2006;59(2):176-184.

38. Longworth L, Buxton MJ, Sculpher M, Smith DH. Estimating utility data from clinical indicators for patients with stable angina. The European journal of health economics : HEPAC : health economics in prevention and care. Dec 2005;6(4):347-353.

39. Grootendorst P, Marshall D, Pericak D, Bellamy N, Feeny D, Torrance GW. A model to estimate health utilities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. J Rheumatol. Mar 2007;34(3):534-542.

40. Marshall D, Pericak D, Grootendorst P, et al. Validation of a prediction model to estimate health utilities index Mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the hip. Value Health. May-Jun 2008;11(3):470-477.

41. Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. Med Decis Making. Jan-Feb 2008;28(1):66-89.

42. Segal L, Day SE, Chapman AB, Osborne RH. Can we reduce disease burden from osteoarthritis? Med J Aust. Mar 1 2004;180(5 Suppl):S11-17.

43. Gatsonis C, Sampson AR. Multiple correlation: exact power and sample size calculations. Psychological bulletin. Nov 1989;106(3):516-524.

44. Gershenson C. Artificial Neural Networks for Beginners. 2003:1-8.

45. Nissen S. Neural Networks Made Simple. 2005:14-19.

46. Chuang LH, Whitehead SJ. Mapping for economic evaluation. British medical bulletin. 2012;101:1-15.

APPENDIX A: INTRODUCTORY SCREEN FOR ONLINE SURVEY

Dear [total joint replacement patient],
You are receiving this email because you have had a total [hip/knee] replacement surgery by [physician name] of the OrthoCarolina, P.A. Hip and Knee Surgeons. Dr [physician name] is collaborating with researchers at the University of North Carolina at Charlotte to determine accurate measures of health related quality of life for patients after surgery.

Several questionnaires have been used for decades to evaluate health related quality of life or health outcomes, i.e. pain and function, after total joint replacement surgery. While these questionnaires are very useful in evaluating health outcomes following total joint replacement surgery, there are other questionnaires that provide another type of health related quality of life measure called health utility. Health utility measures are necessary for the calculations used to determine the cost-effectiveness of various osteoarthritis treatments including total joint replacement surgery.

We are doing this study to develop calculations that can be used to determine the health utility scores based on the traditional pain and functional scores for patients who have had a joint replacement. Results from this study may help researchers in the future conduct cost-effectiveness analyses of total joint replacement procedures.

There are no foreseeable risks to participating in this study. The results of the study do not include any data that could be used to identify you. There are no direct benefits to you for participating in this study but the results may help others in the future. Your participation is voluntary and you have the choice to not participate. If you choose not to participate, there will be no loss of benefits to you and you may withdraw participation at any time without loss of benefits.

If you volunteer to participate, you will spend approximately 30 minutes completing these questionnaires. If you agree to participate, please click next. You will then be asked to log in to complete the questionnaires. If you are not able to complete all of the questions in one sitting, you may take a break and come back at a later time to complete the remaining questions. We do hope you can complete them in one day.

If you have any questions about this study, please contact Susan Odum at 704-323-2265.

Thank you,
Susan Odum, MEd
University of North Carolina at Charlotte; Doctoral Student, Health Services Research
OrthoCarolina, P.A. Research Institute

Jennifer Troyer, PhD
University of North Carolina at Charlotte; Associate Professor, Economics

Dr. [physician name]
OrthoCarolina, P.A. Hip and Knee Center