

Validity of Retrospective Patient Reported Assessment of Pre-Surgical Hip Pain and Disability Following Hip Arthroscopy

Matthew Joel Thomas Wain

A thesis

submitted in partial fulfilment of the requirements for the degree

of Master of Osteopathy Unitec Institute of Technology, 2021

Abstract

Background:

Retrospective patient reported outcome measures (PROMs), which evaluate a time-point prior to collection, can be beneficial both clinically and in research. The purpose of this study was to assess the validity of pre-surgery PROMS completed retrospectively 6 and 24 months after hip arthroscopy compared with prospective completion. Validity was judged in relation to preoperative test-retest reliability.

Methods:

Three separate groups of hip arthroscopy patients were recruited. In two retrospective groups, patients who had undergone a hip arthroscopy and had completed preoperative PROM questionnaires were contacted 6 or 24 months after surgery and asked to recall their preoperative state. In a third test-retest group, patients completed prospective PROMs twice prior to surgery, 2–8 weeks apart. PROMs included Non-Arthritic Hip Score (NAH), Hip Disability and Osteoarthritis Outcome score (HOOS), International Hip Outcome Tool (IHOT-12), and Visual Analogue Scale (VAS) for Medical Care, Feeling and Pain.

Results:

In the 6-month, 24-month and test-retest groups, patient sample sizes were 52, 40 and 61, respectively. Retrospective 6- and 24-month and test-retest PROMs demonstrated 'moderate' to 'very large' correlations (ICC=0.42–0.79), except for 6-month VAS-Medical-Care and VAS-Feeling ('small') and 24-month VAS-Feeling ('trivial'). PROMs for the 24-month group reduced, from prospective to retrospective, for NAH-Total (MD -4.10 [-7.93--0.27; (95%CI)]; p=0.03), HOOS Symptoms (-8.21 [-13.44--2.97]; p=0.003), HOOS-Daily-Living (-7.53 [-12.71--2.36]; p=0.005), HOOS-Sports (-6.49 [-12.97--0.00]; p=0.05), and HOOS-Short-Form (-5.16 [-9.83--0.49]; p=0.03). Similar non-significant reductions were observed for 6-month retrospective PROMs, except for VAS-Feeling which increased retrospectively (7.47 [0.13--14.81]; p=0.05). Three test-retest PROMs also showed statistically significant reduction, with worse outcomes approaching surgery: HOOS-Daily-Living (-3.87 [-7.73--0.00]; p=0.05), HOOS-Short-Form (-4.49 [-8.41--0.53]; p=0.03), and IHOT-12-Total (-4.91 [-8.7--1.23]; p=0.01).

Conclusion:

Correlations for retrospective versus prospective PROM scores spanned a similar range to test-retest correlations. Therefore, PROM scores relating to pre-surgery condition, but completed 6 or 24 months after surgery, showed agreement with PROM scores completed before surgery which were consistent with the day-to-day variability expected before surgery. This consistency supports the use of retrospective PROMs completed up to 2 years after hip arthroscopy.

Acknowledgements

I would like to thank my family for all of the support I have received throughout my studies. A special thanks to Emma, for your support and patience, and to my father for helping me through difficult times.

I feel blessed to have worked with my exceptional supervisor, Dr Catherine Bacon of the University of Auckland, whose work in collating participants' data was extraordinary. Her close interest in this study was evident in the large amount of time she invested in providing advice and encouragement, which was greatly appreciated. Thank you Catherine for sticking with me over the course of this lengthy project.

Many thanks to my associate supervisor, Dr James Furness, Assistant Professor of Physiotherapy at Bond University Queensland, who was generous with his time and provided valued guidance and insightful advice.

I particularly want to thank Abbey Lissaman and acknowledge her assistance with the later data collection for this study. Special mention should be made of Dr Matthew Brick and Orthosports North Harbour Ltd., for their constant support. Without the use of data from their patients this project would not have been possible.

To all the participants in this study, thank you for your time, contribution, and cooperation.

Abbreviations:

ACL QOL:	Anterior Cruciate Ligament-Quality of Life
CTT:	Classical Test Theory
COSMIN:	Consensus-based Standards for the selection of health status Measurement Instruments
DASH:	Disabilities of the Arm, Shoulder, and Hand questionnaire
EQ-5D-3L:	EuroQol Group 5 Dimension 3 Levels
FAI:	Femoroacetabular Impingement
FFI:	Foot Function Index
HOOS:	Hip Osteoarthritis Outcome Score
KOOS:	Knee Osteoarthritis Outcome Score
ICC:	Interclass Correlation Coefficient
IHOT-12:	International Hip Outcome Tool 12-Item
IKDC:	International Knee Documentation Committee
MD:	Mean Difference
MTT:	Modern Test Theory
NAH:	Non-Arthritic Hip Score
OHS:	Oxford Hip Score
OKS:	Oxford Knee Score
PROMs:	Patient Reported Outcome Measures
SD:	Standard Deviation
SEM:	Standard Error of the Mean
SF:	Short-Form
VAS:	Visual Analogue Scale
SF-12:	Short-Form 12-Item
SF-33:	Short-Form 33-Item
WOMAC:	Western Ontario and McMaster Universities Osteoarthritis Index
WOMET:	Western Ontario Meniscal Evaluation Tool

Table of Contents

Declaration.....	2
Abstract	3
Acknowledgements.....	4
Abbreviations:	5
Introduction to Thesis.....	9
Section I: Literature Review	11
1.1 Literature Review Overview	11
1.2 Patient Report Outcome Measures (PROMs).....	12
1.2.1 History of PROMs.....	12
1.2.2 Types of PROMs.....	13
1.2.3 Validity and Reliability of PROMs	13
1.2.4 Tools for PROMs.....	15
1.3 Memory and Missing Data in PROMs.....	17
1.3.1 The Role of Memory in PROMs.....	17
1.3.2 Missing Data and PROM Databases	18
1.4 Research into the Validity of Retrospective PROMs.....	20
1.4.1 Research Studies Investigating the Validity of Retrospective PROMs	20
1.4.2 General Research Studies Investigating Retrospective PROMs.....	21
1.4.3 Studies Investigating Retrospective PROMs for Pain Prior to Lumbar Fusion in Patients with Chronic Low Back Pain.....	23
1.4.4 Studies Investigating Retrospective PROMs after Shoulder and Upper Arm Surgeries.....	25
1.4.5 Studies Investigating Retrospective PROMs after Hip and Knee Arthroplasty	27
1.5 PROMs and the Hip	31
1.5.1 The Use of PROMs in Hip Arthroscopy	31
1.5.2 Nature of the Hip Joint.....	31
1.5.3 Anatomy of the Hip.....	32
1.5.4 Hip Pathologies	33

1.5.5	Hip Surgery	34
1.5.6	Retrospective PROMs in Arthroscopy	35
	References	37
	Section II - Manuscript	51
	Validity of Retrospective Patient Reported Assessment of Pre-Surgical Hip Pain and Disability Following Hip Arthroscopy.....	52
	Abstract	53
	Introduction	54
	Methods	55
	Overall Design.....	55
	Inclusion and Exclusion Criteria:.....	56
	Data Collection	56
	Outcome Measures	57
	Data Analysis	57
	Results	58
	Participant Characteristics	58
	PROM Scatterplots.....	60
	Intraclass Correlations (Table 2).....	61
	Bland-Altman Plots.....	63
	Changes in PROM Scores (Table 3)	64
	Discussion	66
	Comparison to Other Studies of Retrospective PROMs.....	67
	Variability between Test-retest and Retrospective PROMs.....	68
	Retrospective PROM Findings.....	70
	Test-retest Group	71
	Limitations	72
	Conclusion	73
	References	73
	SECTION 3: Appendices	77

Appendix A: Scatterplots for each group.....	77
Appendix B: Bland-Altman Plots.....	83
Appendix C: Hip Osteoarthritis Outcome Score (HOOS)	89
Appendix D: Non-Arthritic Hip (NAH) Score	94
Appendix E: International Hip Outcome Tool 12-Item (IHOT-12)	97
Appendix F: Preoperative Visual Analogue Scale (VAS) for Medical Care, Feeling and Pain	100
Appendix G: Arthroscopy: The Journal of Arthroscopic and Related Surgery Guide for Authors	101

Introduction to Thesis

Patient-reported outcome measures (PROMs) have become increasingly important in clinical practice and trials. This is because they provide a reliable way to understand a patient's pain, satisfaction, quality of life, function and disability (1). Collected before an intervention, these self-reported outcome measures establish a patient's pre-intervention status (2), which can then be used as a benchmark to assess change resulting from the intervention (3). In clinical research trials, PROMs establish a baseline and post-intervention status for each group involved. This can allow researchers to better detect differences between groups when analysing the outcome data. The applicability and ease of using PROMs have led to them becoming a mainstay in assessing outcomes from interventions (4). The use of PROMs, and whether they can be applied to a specific population, can be evaluated with respect to validity and reliability (5). Validity is the degree to which a measurement tool successfully quantifies the construct that it is designed to measure (6). Reliability concerns whether the measure obtained from a tool is consistent following repeated measurement in the same population (5).

Occasionally, when PROMs are acquired during clinical practice or research, key time-points can be missed due to patient non-compliance, forgetfulness, or clerical error. In both clinical trials and practice, this may compromise the integrity of patient data and the ability to accurately assess the patient's perception of change, whether improvement or deterioration (7,8).

In surgical assessment, pre-operative questionnaires in particular may be missed because there is sometimes quite a short window of opportunity between a definite surgery decision and the date of the operation itself. Such omission results in a lack of a recorded baseline level of pain or disability for each individual from which follow-up measurements can be compared (9). In surgical research trials, average follow-up time for all participants is commonly reported (10–12). Since a longer follow-up is generally regarded as more clinically important, a missing post-operative PROM can be replaced by measurement at a later time-point. However, missed pre-operative PROMs cannot be substituted with later information in the same way, and later measurement relies solely on the patient's memory of a past condition.

Clinically, patients are often asked about improvement in outcomes compared with a retrospective assessment of their condition prior to treatment (13). In surgical research trials, pre-operative PROMs may also be replaced with retrospective outcome measures in order to avoid low response rates from participants, provide researchers with access to a larger

cohort and to allow the incorporation of longer observation periods (14). By way of example, a large study, which was based on a previous World Hip Trauma Evaluation (WHiTE) multicentre study in the U.K., assessed patients' pre-fracture quality of life relating to patient characteristics and treatment choices. This was undertaken by having patients, who participated in the WHiTE study, engage in a retrospective assessment on their quality of life before their fracture at the time of admittance (15).

It is common for PROMs to enquire about symptoms and a patient's condition covering a recent duration of time prior to the questionnaire, for example in the last week, or in the last month. However, whilst there is often an element of retrospective assessment, PROMs include the patient's current condition and are not designed to gather data retrospectively, that is, not from a period not immediately before the current timepoint. To the extent that PROMs involve retrospective assessment, their validity may depend on the age of a patient, impact of the injury, time of recovery and the intensity of pain (16). If a condition from a past period of time is asked about, the duration of elapsed time after the recall period, and other comorbidities that have occurred in the intervening period, may affect validity (17).

The aim of this thesis is to investigate the validity of the retrospective assessment of PROMs in hip arthroscopy patients. The thesis is presented as two sections: the first section of the thesis contains a review of the surrounding literature, introduces the use of PROMs including the effect of memory and missing data on outcomes. This literature review also analyses past studies which have researched the validity of retrospective PROMs and draws conclusions about what is known and what is unclear in this area. Further, the literature review focuses on hip arthroscopy and explores how the targeted demographic could affect the validity of retrospective assessment of PROMs.

The second section of this thesis, presented as a draft paper for publication, reports on the findings of a study designed to investigate the aim of the thesis. This study assesses the validity of retrospectively collected PROMs administered after hip arthroscopic surgery by comparing these measures with prospectively collected PROMs (taken before surgery). Patients, to the best of their ability, have completed retrospective PROMs recalling their prospective status, which is their pre-surgery condition. The retrospective PROMs have been taken at two different time-points, 6 and 24 months, in two different groups of patients. Validity is determined partly by comparing reliability statistics with those from a test-retest group who were assessed twice before surgery, using the same measures. Appendices contain graphs and supplementary material to the thesis.

Section I: Literature Review

1.1 Literature Review Overview

Patient Reported Outcome Measures, commonly referred to as PROMs, are based on questionnaires that comprise of series of questions that ask the patient to recall their health status, quality of life, or symptoms over a recent period of time, either prior to or following intervention.

The primary importance of PROMs is that they can assist physicians in both patient care and in clinical administration. They achieve this in four main ways. First, due to the nature of the self-reporting questionnaires, PROMs provide a means to better assess both a patient's perceived changes in pain and disability pre- to post-intervention, and also patient satisfaction. This enables physicians to better evaluate treatment outcomes from the perspective of the patients (18). Secondly, PROMs completed by patients enable physicians to compare results with data already obtained from other patients in a database, taking into account individual patient or intervention factors. This can assist physicians in tailoring their recommendations and advice to patients on such matters as effectiveness of treatment, rehabilitation, and expected time of recovery. Thirdly, once again due to the nature of the questionnaire, PROMs can provide physicians with a ready means of measuring some parameters such as patients' levels of pain which would otherwise be unquantifiable. This can assist in such things as accessing patient outcomes and effectiveness of treatment. Finally, in providing valuable data, PROMs are cost-effective in their implementation and have minimal administration requirements (19,20).

A secondary benefit of PROMs is that they can provide a framework for patients which enables them to better understand or appreciate their present condition and the improvements achieved as a result of the procedure. A study by Teela et al., noted that discussing PROMs in the consultation room empowered patients, enhanced practitioner-patient communication, and promoted shared decision making (21).

Occasionally, in the case of PROMs, certain time-points, either in pre-surgical prospective consultations or in follow-ups, can be missed. The result of missing data is that the degree of improvement achieved by the outcome can be uncertain. Missing data in turn may impair a physician's ability to accurately gauge the patient's condition or benefit of the treatment. In clinical research studies, there are two main ways of dealing with missing data, namely the use of certain statistical methods or, alternatively, the use of retrospective PROMs.

As far as the use of statistical methods are concerned, these are designed to reduce and identify the extent that missing data has on creating bias in the results (7). Selection of

statistical analytical methods to best reduce the impact of missing data can depend on whether missing data are treated as random, completely at random or not at random (22).

The other common approach to handling missing data is to use retrospective PROMs. Retrospective PROMs are outcome measures taken after a planned time point, where patients are asked to recall their status as they were before their consultation (17,23). The use of retrospective PROMs can enable clinical studies and practices to easily gather potentially reliable data on a patient's pre-intervention condition, without concern of time restraints immediately before or after treatment.

Whilst many PROM questionnaires involve a degree of recollection on the part of the patient, they are not inherently designed to be filled out after an extended period following the evaluation period (24). Several studies have looked into the validity of retrospective PROMs and have found varied results (2,25–27). Results have varied due to differences in the retrospective timeframe, age of the cohort, nature of medical conditions and their associated pain or disability, specific PROM questionnaires used and sample size. Nevertheless, despite a number of studies examining issues of reliability, the applicability of retrospective PROMs has been limited to a fairly narrow range of medical procedures, and their transferability to other clinical scenarios remains inconclusive.

The following review outlines the history and types of PROMs available. It also examines relevant literature in order to consider the role of memory in PROMs, their use in assessing surgical outcomes, how missing data may be handled in clinical practice and in research studies, the validity of retrospective PROMs, and whether peculiarities of hip arthroscopy may influence this validity.

1.2 Patient Report Outcome Measures (PROMs)

1.2.1 History of PROMs

In 1988, the use of PROMs was proposed as a means to provide outcome-focused patient information whilst not adversely impacting upon practitioner-patient rapport, and also provide a reliable source of information on the effectiveness of treatment (28). Over the last 30 years PROMs have since been used to a considerable extent in clinical practice, clinical trials and other research studies (29). This trend is likely to continue due to an increasing focus on quality assessment in health care which emphasises patient-centred care and healthcare value (30). PROMs can assist in gathering details of a patient's pain, quality of life and condition. They can also assist in understanding the patient's perspective on their treatment outcome (31). PROMs are increasingly common now in healthcare due to their ability to gain

a large amount of information and provide a reliable way of gauging change over time for a patient (32). PROMs have now become the gold standard in quantifying treatment outcomes (33).

1.2.2 Types of PROMs

PROMs can be categorised into several different forms, depending on their method of gaining data and their purpose. Christensen et al. (34) identified several forms of PROMs. The first, are mixed outcome measures, and are derived from data gained from patients during physical examinations and combined with clinical assessment measures, such as range of motion (35). The second form are general health-related and overall quality of life outcome measures based solely upon the patient's questionnaire with consideration of their ability to perform daily activities with reference to physical, mental and social factors (32). The third type are PROMs focused solely on system-specific and disease-specific outcome measures evaluating symptoms related to a specific body region or disease (36). PROMs reported within the musculoskeletal medicine and orthopaedics literature are predominately the second and third types of PROMs as identified by Christensen et al. (34). They deal with general health-related and quality of life outcomes and in the assessment of disease- and system-specific outcomes (34). They can be generally classified into measurements of pain, disability, and quality of life, encountered by the patient both before and after surgery (20).

In addition to these types of PROMs, it should be mentioned, but only by way of association, that the patient reported outcome measure information system (PROMIS) is a commonly used global tool designed to cover symptoms and outcomes over a wide range of chronic conditions (37). The topic of PROMIS has, however, no direct applicability to the use of PROMs in hip arthroscopy and is therefore outside the scope of this thesis.

All types of PROMs have become increasingly important for clinical purposes and research. This is because there has been a growing demand for patient-centred assessment from both health care professionals and patients (38) and due to the fact that PROMs provide a means to measure treatment outcomes. However, the ability of PROMs to measure the success of a procedure depends on both their validity and their reliability.

1.2.3 Validity and Reliability of PROMs

The validity of a PROM depends on whether it successfully quantifies the construct that it is designed to measure. For this to occur, the construct needs to be based upon either a model, theoretical framework or be empirically justified, so that it can be clearly defined (39). Questions, or items, must be constructed in such a manner that statistical analysis can

successfully identify relevant outcomes and draw accurate conclusions from observed patient responses with regards to the construct. In order to do this, a questionnaire must be reliable. However, a measure might be completely reliable and repeatable, but not at all valid for measuring the construct which it is supposed to measure. By way of example, Vitale et al. (2001) found that Quality of Life PROMs, which were valid for adults, were not appropriate for children, due to different expectations and health status (40). In another study, Suryavanshi et al. (2019), noted that outcome measures evaluating adults could not be applied universally to a younger population (41).

With respect to questionnaires, there are in fact different forms of validity, of which criterion validity, internal consistency, and construct validity are the most relevant. Briefly, criterion validity involves the practice of using a test-retest score as a 'gold standard' in order to determine the validity of a questionnaire (42). Internal consistency is an inductive estimate on the degree interrelatedness among the items, based on measurements used and the research design (43). Construct validity is the degree of correlation between a specific measuring device, or a procedure, and a theoretical concept (43).

The reliability of a PROM is concerned primarily with the degree of measurement error and whether it can measure a particular construct consistently. Therefore, reliability depends to a considerable extent on the repeatability and consistency of results from a PROM (44). If there is little correlation in results from patients, then the PROM can be regarded as being less reliable. Reliability of a PROM is an important consideration in determining its usefulness, that is whether the PROM is appropriate for a specific group of patients (39).

Validity and reliability of a PROM determines its quality. Responsiveness to change over time is another important parameter (45), but this factor is not always achieved with certain PROMs applied to particular cohorts (46). In order for studies to ensure that PROMs are valid and reliable, a number of checklists and guidelines exist, most notably the COSMIN checklist (47). All of these guidelines refer to a number of statistical methods applicable to a providing a range of quantitative metrics, encompassed within a body of knowledge known as the psychometric test theory. These methods can be used to investigate whether a PROM provides a valid and reliable measurement (39).

In terms of measurement, there are a number of ways to assess the validity and reliability of a PROM. One method applies statistical methods within psychometric test theories, one of which is known as 'classical test theory' (CTT). The aim of CTT, which is also referred to as the 'true score theory,' is to analyse the degree to which a measure provides a precise, valid, and reliable final score, in respect of the construct it represents. In order to do this, CTT

takes into account certain deficiencies with the data and participants' recollection, and so regards the final score of a test as the sum of a 'true score' and what is described as an 'error score' (39). Consequently, CTT can be applied to determine the reliability and validity of a PROM by considering such things as the ability of participants and the difficulty of questions, which can be determined by both the importance of the construct being assessed and items that few participants answer correctly (48).

Historically, there was a change of focus from test-level validity, that is validity of the PROM in its entirety (49), to consideration of how responses to individual questions depended on unobserved background characteristics. This shift of emphasis saw CTT develop into another psychometric test theory, known specifically as 'modern test theory' (MTT). MTT does not in fact compete with CTT, but is instead a way of assessing a PROM on the basis that the probability of a correct answer to a question is a mathematical function of person and question parameters, where the person parameter is allocated a single number (39). However, despite their differences, MTT and CTT are concerned with the same issues of validity of PROMs and they differ only in their method of statistical analysis. MTT focuses on the relationship between a latent variable, as in unobserved values, and a question, rather than on the sum of the 'error score' and 'true score' which, as noted, is the focus of CTT (39). What this does mean, however, is that MTT provides a method that analyses all of the relevant properties when determining the validity and reliability of PROMs, which is not the case for CTT. As a result, claims of validity for PROMs that are based solely on the use of CTT, may be considered to be less accurate, as CTT does not effectively consider the utility of a measure in clinical situations (50).

1.2.4 Tools for PROMs

A range of specific PROMs are in common use in musculoskeletal medicine and orthopaedics. These include general measures for health-status and quality of life such as SF-36 Health Survey, and Musculoskeletal Health Questionnaire and Visual Analogue Scale (VAS). VAS items contain an analogue scale which allows patients to subjectively rate levels of sensation, typically pain or normality of feeling (51,52).

Some PROMs have validated subscales oriented around particular joints, such as Roland Morris Disability Questionnaire and Oswestry Disability Index (ODI) for low back pain, Shoulder Pain and Disability Index and Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH) for shoulders, and Knee Injury and Osteoarthritis Outcome Score (KOOS) and Oxford Hip Score (OHS) for knees. A number of PROMs relating to the hip joint are in common use. These are the Harris Hip Score, Western Ontario and McMaster

Universities Osteoarthritis Index (WOMAC), Oxford Hip Score (OHS), Hip Outcome Score (HOS) and Hip Disability and Osteoarthritis Outcome Score (HOOS) (53). With the exception of HOOS, these PROMs better deal with treatment of late-stage osteoarthritis and are typically administered to older patients, without specifically addressing younger or more active demographics (41). In younger or more active cohorts, these PROMs exhibit signs of a 'ceiling effect,' and tend to lose their sensitivity to improvements that occur at the higher end of the activity spectrum. This 'ceiling effect' means that sometimes elderly patients could display similar results as a young active patient, even though they greatly differ in terms of lost capability and impact of their condition (54,55).

In the case of hip arthroscopy, a procedure typically performed on younger more active patients, there are several prominent and reliable PROMs that are used. These particular PROMs are the Non-Arthritic Hip score (NAH); HOOS Version LK 2.0; and the International Hip Outcome Tool (IHOT-12).

The PROM tool NAH was designed to assess functional pain in young and active patients (56). HOOS was designed to assess people with hip disability with or without hip osteoarthritis by adding dimensions onto the WOMAC (57). The NAH and HOOS PROMs both assess the degree of experienced hip discomfort or difficulty associated with listed specific activities. Both questionnaires define subscale outcomes based on subsets of items. In the case of NAH, the subscales are Pain, Symptoms, Function, and Activities and for HOOS they are Symptoms, Pain, Daily Living, Sports, and Quality of Life. One distinguishing feature of NAH is that it includes a total score. It also seeks information about the previous 48 hours, except for activities for which assessment is made of the degree of difficulty experienced over the last month. By comparison, the HOOS enquires about difficulties experienced in the last week.

The International Hip Outcome Tool 12-item (IHOT-12) is an updated and shortened version of the International Hip Outcome Tool 33-item (IHOT-33) (58). The IHOT-33 was designed to focus on measuring the impact of hip disease in young and active patients, and treatment-related change. It provides a sensitive measure on a wide range of symptoms and problems. The shortened version IHOT-12 loses very little information, despite being one third of the original length of its predecessor. Unfortunately, IHOT-12 items do not specify a time frame and ask a combination of questions relating to level of pain, difficulty/trouble, and concern.

1.3 Memory and Missing Data in PROMs

1.3.1 The Role of Memory in PROMs

As noted, PROMs require a patient to evaluate their state in order to better portray their current condition. PROMs, particularly in regard to pain and disability, generally require a certain amount of recall from a patient about their condition from the recent past, or over a long period, and therefore are to some extent reliant on a patient's memory (59). This means that the patient's ability to accurately recollect is always relevant when considering the reliability of these measures. In addition, in the case of retrospective PROMs, a patient's ability to recollect their pre-treatment condition can be used to assess the effectiveness of intervention after the event (38).

The accuracy of a patient's recollection can be influenced by several factors. The first factor is a patient's recollection accuracy, which may be affected by their level of education, age, mental health scores or comorbidities (60). The length of time that transpires between an event or period being evaluated and the recollection of that event or period is often called the 'recall period,' and its length can also potentially affect the accuracy of memory (8).

The second factor is recall bias, a well-documented and common source of systematic error in clinical research (27,61–63). Recall bias may be thought of as a form of differential misclassification bias which, simply put, is the result of an overestimation or underestimation by the patient in recalling a prior state. In clinical situations, recall bias could result in incorrect data concerning the effectiveness of treatment (60,64). With a patient's change in perception as to their health and other influences at the time of treatment, recall bias may occur and lead a patient to either overestimate or underestimate their experience.

There are two dominant theories about health-related memory that address the patient's ability to recall their prior state, being 'implicit theory of memory' and 'response shift' (2,65).

The implicit theory of memory hypothesises that individuals have an awareness about the status of their health and about any condition that may elicit a change in that status, such as an injury or illness, or surgery or another intervention. Proponents of implicit memory contend that it is difficult for a patient to recall their previous state, without relating to a specific distinctive reference point, or 'anchor point'. Without such an anchor point, they argue, patients extrapolate their previous state by posing the questions such as "How do I currently feel?", or "How do I think things may have changed?" Thus, when follow-up measurements are taken after a treatment, patients might rely on their pre-intervention state as a reference in order to complete their PROMs. This process can potentially alter the

memory of the pre-intervention condition from how it would have been rated at the time. This is due to the patient inferring their condition based on how they have perceived change from different anchor points.

The second theory, concerning health-related memory, is referred to as 'response-shift', which hypothesises that retrospective memory of health status may be subject to alteration in the mind of the patient. Through numerous internal and external influences, a patient's perception and judgment of a memory can change over time. The change in memory of an event has been termed by some as a 'response shift' (66), or 'scale recalibration' (67). The theory of response-shift infers that changes to a patient's health may cause cognitive and behavioural changes, which can alter their normal rating of their health status. This shift in a patient's perception in their health, and their understanding and interpretation of their previous health status, may subsequently influence their recollection concerning their pain, function or quality of life (61,68). In one study, Gotlin et al. (2020), reported a consistent change in their participants in underestimating pre-operative score retrospectively, with participants reporting worse outcomes than their actual prospective PROMs. Gotlin theorised that successful outcomes could lead to patients overestimating their preoperative state and consequently overestimating the effectiveness of treatment, which in turn may lead to worse, less reliable retrospective PROMs (64).

1.3.2 Missing Data and PROM Databases

Apart from matters relating to patients' memory, another factor which impacts upon validity and reliability of PROMs, obtained from clinical databases, is missing data.

Databases are electronic filing systems that record patient data in a variety of administrative and clinical settings (69). In the case of databases within the medical profession, particularly in surgery, it is important for physicians to accurately record the nature of treatment and patient outcomes in order that the likely benefits of the treatment can be better understood and appreciated (70). They may also wish to contribute to research and knowledge in the field by sharing these patient outcomes. To assist these matters, databases invariably provide an array of useful information on patients' clinical data, surgical outcomes and the success or failure of treatment (71).

However, information found in patient databases used in clinical practice and research may have been impacted by the problem of missing data. It is not uncommon for data to be incomplete for the reason that, at the time of collection, important time-points can be missed. Missing data can also be the result of a range of other factors including such things as miscommunication with staff delivering the PROMs, patient resistance to completing

questionnaires, or due to time sensitive scenarios, such as medical emergencies (2,25). The reason why missing data becomes an important consideration is that it can result in a database being less comprehensive. This may limit a physician's available resource for improving patient care and better understanding the efficacy of treatment based on the experiences of other patients (20).

Lacking data concerning a patient's experience and condition prior to or after treatment, can hinder a clinician's or a researcher's ability to understand the full impact of treatment and its outcome over time. Hence, when data are missed, clinicians and researchers can be motivated to either collect retrospective data or use statistical methods to reduce the impact of missing data (7,9).

Retrospectively collected PROMs may sometimes be the only information available in certain instances, such as gaining personal data in time sensitive situations (24) or situations that involve certain vulnerable populations which lead to difficulties in accessing the patients involved (72). Occasionally, retrospective PROMs are adopted and preferred over prospective data due to their potential to reduce project costs and participant drop-out rates, and perhaps also to decrease projected times for the completion of research projects (2,73). However, despite the ease of collecting retrospective data and its cost-effectiveness, there are currently drawbacks using it in place of prospective data, as validity has not been conclusively established.

Missing data within databases may also be addressed statistically. It has been noted that statistical reduction of bias is based on a premise that patients who fail to respond to a questionnaire, referred to as non-respondents, tend to be systematically different from those who provide complete data (9,74). There are a number of ways to statistically analyse the bias caused by missing data and analysis can depend upon whether missing data are due to observed or unobserved data. Patient characteristics, such as age and gender, can be classified as observed data and any unmeasured factors can be categorised as unobserved data.

When data are 'missing completely at random,' which means that response status is independent from observed and unobserved values, discarding non-respondents from analysis will not result in bias (9). This implicitly assumes that those who provided complete data are representative of the whole population. On the other hand, when data are only 'missing at random', as distinct from 'missing completely at random', differences in observed data, such as patient characteristics, may explain differences between respondents and non-respondents (9).

Omissions in data 'missing at random' can be addressed by multiple imputation, which is a process where missing data is replaced by a plausible set of substituted values (9,75). The natural variability of missing data is maintained by creating values which are based on variables correlated with causes of "missingness" and the missing data (76). Multiple imputation is available on standard statistical software, meaning it is readily available to researchers (77). However there are a few issues with its implementation in studies, one of which is that missing data may sometimes be incorrectly assumed to be 'missing at random' (78). This was apparent from a study into missing data in the UK primary care database which found that implementation of multiple imputation for data relating to smokers led to a lower number of smokers in their database compared to external sources. This caused the study to conclude that the data within the database may not have been 'missing at random,' or that there was some degree of error in the process (79).

Both statistical reduction of bias and retrospective PROMs each have their strengths and weaknesses in addressing missing data in surgeries and other clinical practices. Statistical methods, such as multiple imputation, are a fairly reliable way of reducing the impact of missing data especially in the context of research. However, despite the fact that their validity in some situations has not been confirmed, the use of retrospective PROMs involving patient input is nevertheless both a practical and efficient method for medical practitioners to obtain missing data.

As a result, the importance of missing data, and the significant impact it can have on the completeness of patient databases, reinforces the need for retrospective PROMs in some situations, in order to address the adverse effect of missing data and to assist in clinical patient evaluation and in research. Consequently, the question of how reliable retrospective PROMs are assumes some importance.

1.4 Research into the Validity of Retrospective PROMs

1.4.1 Research Studies Investigating the Validity of Retrospective PROMs

Research concerning the validity of retrospective PROMs has been carried out in a wide range of different fields within the medical profession, including general fields such as physical exercise, obstetrics and dentistry and surgical fields. Surgical studies, which typically analyse databases found in clinical practices and surgeries, have focused on retrospective PROM validity following lumbar fusion in patients with chronic low back pain, shoulder and elbow surgeries, and hip arthroplasties. However, studies to date have produced conflicting results on the validity of retrospective PROMs which has caused a lack of consensus over their use (2,25–27,38,64,67,80–90).

1.4.2 General Research Studies Investigating Retrospective PROMs

Many of the studies that have looked into the validity of retrospective PROMs have concerned general health-related areas. Although such studies have often reached mixed or uncertain conclusions, they have invariably contained a number of limitations, particularly in regard to either having a small cohort (80,88,89), or using an unreliable, invalid or insensitive questionnaire to gain the retrospective PROM data (91).

Retrospective PROMs, if confirmed to be reliable, could potentially change the way data is gained in multiple areas of healthcare. Throughout the medical profession, the application of retrospective PROMs could ease administrative burden and reduce the importance of a need to prospectively gain patient data. Any PROM can potentially be collected retrospectively. However, the factors that can influence the reliability of retrospective PROMs may vary, depending on the area of healthcare involved, and can depend on the outcome measure in question (92,93).

In order to better understand and determine whether retrospective PROMs are reliable in clinical studies and practice, future studies ought to consider the state of the patient's memory, and the possible effect of 'response-shift' and 'theory of memory' on their recollection. This means that it is important for studies to use validated PROMs and a large patient cohort not influenced by other factors that might affect judgement of PROMs items. The study by Waldenström et al., whilst having a large patient pool of 2482 participants, did not implement a validated PROM questionnaire (91). The questions used a series of items, some with 7-point and one with only a 3-point scale. Furthermore, the reliability of data from the study could have been reduced by patients being asked to recall their labour pain, at a time when they had been under the influence of a range of analgesics. The influence of analgesics may have not only altered their experience, but may have also affected their ability to recall the painful event. This factor could have been one of the reasons why no definitive conclusion could be drawn from this large study. The study by Beese and Morley exhibited a similar limitation for the reason that patients, who were recalling pain associated with having their wisdom teeth removed, had also been under the influence of general anaesthetics (80).

One way to eliminate limitations associated with compromised reliability would be to focus only on conditions when patients are in fully coherent states. For example, patients should not be asked to recollect a time when they were under the influence of analgesics. Furthermore, reliability can be compromised in studies that don't focus on situations where participant responses are likely to have a consistent meaning across the cohort.

The study by Babel et al., in focusing on acute physical exercise pain after a marathon, recruited a broad range of participants, who likely also experienced diverse sources of their pain (89). The lack of a homogenous group potentially limited the applicability of the findings to other clinical or sporting situations. Clinical situations in which patients from a homogenous cohort rate their status are often found in analyses of databases from clinical practices and surgeries. These patients generally have specific conditions that manifest with a consistent pattern of symptoms, signs, or pathological stages.

General studies on retrospective outcomes have looked into a variety of different fields ranging from pain after having wisdom teeth removed, childbirth and physical exercise (80,88–90). A summary of the general studies associated with pain memory is set out in Table 1.

Table 1. General Research Studies Investigating Retrospective PROMs

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Widnall et al., 2014 (88)	n = 36 Mean age 54.6 years 3 months	SF-12, FFI	Retrospective scoring lacked accuracy when compared to prospective methods, 4.5 months after foot and ankle surgery. SF12 was recalled more accurately than the FFI ($r = 0.30; 0.05$).
Babel, 2016 (89)	n = 62 Mean age 37.1 years 3 or 6 months	Numeric Rating Scale, Positive and Negative Affect Schedule (PANAS),	Concluded that pain induced by physical exercise was not remembered accurately. Memory of pain and its impact had significant recall bias, with patients recalling less pain after exercise ($F(1,60) = 25.14, p < .001, \eta^2 = 0.30, \text{power} = 0.99$)
Waldenström et al., 2006 (90)	n = 2482 N/A 2 months and 1 year	Scale of 1 to 7	Mixed findings, inconclusive. One finding was that epidural analgesia and nitrous oxide were associated with changes in women's long-term memory of labour pain, but in opposite directions (OR 1.39; 95% CI 1.11–1.73; $p = 0.003$). Scale of 1 to 7 was too insensitive to give a comprehensive result.
Beese & Morley, 1993 (80)	n = 49 18 – 50 years 2 weeks	McGill Pain Questionnaire (MPQ) and the UWIST Mood Adjective Checklist (UMACL).	Accuracy of memory was only 'fair' for both mood and pain ($\kappa = 0.5$), two weeks after having wisdom teeth removed under general anaesthetic.

1.4.3 Studies Investigating Retrospective PROMs for Pain Prior to Lumbar Fusion in Patients with Chronic Low Back Pain.

Focusing the application of retrospective PROMs into certain areas where conditions are reproducible, could potentially increase the reliability of a study's findings. Different surgical fields, such as those dealing with specific orthopaedic injuries, usually directly deal with comparable patient cohorts and can provide a reliable platform for studies to find large patient cohorts with similar conditions before and after surgery.

A majority of studies into the reliability of retrospective PROMS in musculoskeletal medicine are in relation to surgical outcomes involving pain and disability. In particular, retrospective PROMs assessed following lumbar fusion in relation to before surgery can be a particularly useful field for reliability research, due to the implementation of this procedure for chronic lower back pain, the most common musculoskeletal ailment (94). The few studies on the reliability of retrospective PROMs following a lumbar fusion have had conflicting results, with some supporting or rejecting their use.

The long period of chronic pain, which may fluctuate over the time prior to any surgery, could potentially impact upon the reliability of retrospective PROMs. Chronic low back pain additionally brings an interesting dynamic into assessing reliability of retrospective PROMs due to the potential effect of psychosocial factors, such as depression, anxiety, or pain catastrophising (95), on pain or disability recall. Fluctuations in experience of chronic pain or psychosocial factors could explain poor reliability noted in two studies of patients having lumbar fusion following chronic lower back pain. These patients had their ability to recall their preoperative status evaluated over a very long time period of two months to five years (81,96). One of these studies was by Pellise et al., who retrospectively evaluated pre-surgery back pain 37.5 months, on average, after lumbar fusion (81). In this study, there was only a small to moderate agreement between retrospectively and prospectively collected pain and function PROMS scores, with intraclass correlation coefficients (ICCs) ranging from 0.34 to 0.52. Pellise et al. noted a fair amount of recall bias in their patient cohort, as participants overestimated their preoperative pain retrospectively (Pain-Prolo: retrospective, 1.60 ± 1.03 ; prospective, 2.07 ± 0.77) (81). Similar findings were reported by Aleem et al., with patients' recollection of their preoperative back pain being significantly worse than indicated in the prospective PROM assessment, and unreliable (Mean Difference (MD) +9.6%, 95%CI 5.6 – 14.0; ICC = 0.44;) (96). The relatively long duration of follow-up in these studies, 43.3 months for Aleem et al. (96) and 37.5 months for Pellise et al. (81), could potentially be the reason behind the significant difference between the retrospective and prospective PROMs, compared to the study done by Kuittinen et al (82). However, the use of

non-validated, self-reported patient questions is also likely to play a part in observations of only moderate recall accuracy noted by Pellise et al. (81). This state of affairs emphasises the importance of using validated PROMs with sufficiently established psychometric properties when dealing with the evaluation of retrospective data.

The study by Kuittinen et al. not only looked at a shorter time frame of three months, but also took into account the effect surgery outcome has on recall bias (82). Unlike the studies by Aleem et al. (96) and Pellise et al. (81), Kuittinen et al. (82) noted what happened to patients who had good or poor surgical outcomes. It was observed that patients who had poor or very good results had a better memory of their preoperative pain than those with intermediate surgical outcomes. In addition, the retrospective 3-point PROM scale, which was used to determine the success of treatment by a scale of 0 (poor) to 2 (good), showed good intra-rater and inter-rater reliability ($\kappa = 0.682$, $P < 0.001$ and $\kappa = 0.630$, $P < 0.001$, respectively). The findings by Kuittinen et al. lend some credence to the ‘theory of memory,’ where a patient’s surgery stands as an anchor point (82). It is possible that a highly successful treatment, or a failure, results in a strong anchor-point, which might lead to a more dramatic and poignant memory of the operation and of the time before surgery. Kuittinen et al.’s findings support this notion as patients in their study who had experienced a mild change were less likely to recall their preoperative status (82). A summary of the studies on retrospective assessment of pain prior to lumbar fusion in patients with chronic low back pain is set out in Table 2.

Table 2. Studies Investigating Retrospective PROMs for Pain Prior to Lumbar Fusion in Patients with Chronic Low Back Pain.

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Pellis� et al., 2005 (81)	n = 58 Mean age 48.3 years 2 to 58 months (mean 37.5)	Prolo, Pain-Prolo, Function-Prolo, Waddell Disability Index and VAS	Comparisons between prospective and recalled data showed poor agreement and significant differences (ICC = 0.34 – 0.52; MD -18.47 to 0.69).
Kuittinen et al., 2012 (82)	n = 100 Mean age 62 years 3 months	Retrospective 3-point scale, Oswestry Disability Index questionnaire, VAS, and a patient satisfaction questionnaire	Retrospective assessment of spinal surgery outcome was highly reproducible. Accuracy was highest in the patients with poor and good surgical result, ODI (r = 0.59; P < 0.001) and VAS (r = 0.37; P < 0.001).

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Aleem et al., 2017 (96)	n = 62 Mean age 66.1 years 43 months (±16.3)	Numeric Pain Scores, Oswestry Disability Index Score	No significant correlation between actual and recalled scores with regards to back (r=0.18, MD +2.3, 95% CI, 1.5 – 3.2; p < 0.05) or leg (r=0.24, MD +1.8, 95%CI, 0.9 – 2.7; p < 0.05) pain and only moderate correlation with disability (r=0.44, +9.6%, 95%CI, 5.6 – 14.0; p < 0.05) after lumbar surgery at long-term follow up.

1.4.4 Studies Investigating Retrospective PROMs after Shoulder and Upper Arm Surgeries

Orthopaedic outcome research typically involves studies specific to joints or body regions commonly operated on. Due to the close location of the hip and knee, these regions are often combined when considering PROMs, and the same applies in the case of surgery to the shoulder and upper extremities.

Some of the studies available to date, which look at surgeries of the shoulder and upper arm, have reached similar conclusions, being that retrospective PROMs are unreliable (ICC = 0.23 – 0.29) (64,83,85). However, a number of studies have reached a contrary conclusion and have found retrospective PROMs to be reliable (ICC = 0.79 – 0.86) and have consequently promoted their use in place of prospective outcome measures (38,84). The conflicting position of these various studies is perhaps puzzling for the reason that they mostly use either DASH (Disability of the Arm, Shoulder and Hand) or QuickDASH questionnaires. It is apparent however, that the key difference between the studies is the time frame between the retrospective and the prospective outcome measures, even though the questionnaires may be the same.

Two particular studies that had positive results looked at shorter retrospective time frames, from 3 months to 2 years following the initial consultation. The study by Stepan et al., found that patients at four different timepoints, being 3 months, 6 months, 12 months, and 24 months respectively, could accurately recall their preoperative status (ICC = 0.89, 0.86, 0.79, 0.85 respectively) (38). Reynolds et al., who found a good to excellent ability for patients to recall their preoperative status ($r^2 = 0.79 – 0.87$), also reported a relatively short average time frame of 32 weeks after surgery, although this period ranged from 6 to 121 weeks (84).

In contrast, the studies by Gotlin et al. (64), Tashjian et al. (83) and Wormdall et al. (85) looked at significantly longer retrospective time frames. The study by Tashjian et al.,

surveyed patients on average 4.5 years after their surgery (83). Gotlin et al. (64) and Wormdall et al. (85) assessed their patients, on average, 3.26 and 3.75 years after surgery. In both these studies the retrospective PROMs were found to be unreliable (Gotlin: ICC = 0.29 95% CI, -0.07, 0.57; P = 0.068; Wormdall: QuickDASH MD 7.6, Standard Deviation (SD) 15.6; Standard Error of the Mean (SEM) 1.2) (64,83,85).

The effect of recall bias, that is the tendency of patients to either overestimate or underestimate their previous state, was observed in the studies by Gotlin et al. (64), Stepan et al. (38) and Wormdall et al. (85). Gotlin et al. (64), reported a consistent underestimation of pain and disability in the recalled American Shoulder and Elbow Surgeons (ASES) Standard Shoulder Assessment scores, with the retrospective score being significantly lower than the prospective PROM (30.69 ± 16.93 retrospective vs 51.42 ± 19.14 prospective; $P < 0.001$). Both Stepan et al. (38) and Wormdall et al. (85), in contrast, noted that patients consistently overestimated their retrospective pre-surgery QuickDASH score compared to prospective. However, unlike Wormdall et al.'s (85) study, difference in scores for patients in Stepan et al.'s (38) study was not greater than the 'minimal clinically important difference' for scores at 3-months, 6-months, 12-months, or 24-months after the operation (MD 3-months = -7.1; 6-months = 0.8; 12-months = -2.3; 24-months = -2.8). This means that even though Stepan noted an average increase in retrospective scores, it was not consistent in terms of direction or large enough to make them unreliable. It seems likely that the differences in results from Wormdall et al. (85) and Stepan et al. (38) were due to the fact that, in the Stepan et al. (38) study, the time frame was shorter. Therefore, the effect of recall bias was not as substantial in this study involving a shorter time frame (38,85). A summary of the studies on retrospective assessment of shoulder and upper arm surgeries is set out in Table 3.

Table 3. Studies Investigating Retrospective PROMs after Shoulder and Upper Arm Surgeries

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Gotlin et al., 2020 (64)	n = 84 Mean age 57.4 years 39.12 ± 17.37 months	American Shoulder and Elbow Surgeons (ASES) Standard Shoulder Assessment	Retrospective PROMs significantly affected by recall bias (30.69 ± 16.93 vs 51.42 ± 19.14 ; $P=0.001$). Poor agreement between recall and original PROM (ICC = 0.29; 95% CI, -0.07–0.57; $P = 0.07$). Recalled PROMs likely accurate in younger patients with severe symptoms over a longer period.

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Tashjian et al., 2008 (83)	n = 112 Mean age 59 years 34 to 85 months (mean 54)	DASH, Musculoskeletal Outcomes Data Evaluation and Management System, Simple Shoulder Test, VAS	Pain, function, and quality of life had fair correlations with the prospectively determined improvement (R = 0.23–0.25; P <.01). Patient satisfaction more highly correlated (R = 0.41– 0.58; P = 0.001).
Stepan et al., 2013 (38)	n = 140 Mean age 54.7 years 3, 6, 12 and 24 months	QuickDASH	QuickDASH can accurately assess a patient's preoperative function, 3 months (MD –7.1; P < 0.01), 6 months (MD 0.8; P = 0.79), 12 months (MD –2.3; P = 0.43), and 24 months (MD –2.8; P = 0.26) after shoulder surgery.
Reynolds & Thirkannad, 2013 (84)	n = 38 Mean age 50 years 32 weeks	DASH	Recall DASH was an accurate measure for research into hand surgery (y = 1.03, R ² = 0.87; MD 1.6, 95%CI, 10.7 – 12.7).
Wormdal et al., 2017 (85)	n = 160 Mean age 63 years 39 to 97 months (mean 45)	QuickDASH	Retrospective QuickDASH PROMs cannot be used in individual patients because of high inaccuracy, three or more years after surgery (mean 7.6, SD 15.6; SEM 1.2).

1.4.5 Studies Investigating Retrospective PROMs after Hip and Knee Arthroplasty

There are several studies that have researched the validity of retrospective outcomes after hip arthroplasty (total hip replacement) and/or knee arthroplasty (total knee replacement). These surgeries typically have older patients, with chronic conditions such as severe osteoarthritis and, in this regard, differ from hip arthroscopy (97). However, these studies can potentially provide some insight into what could be expected in retrospective PROMs after a hip arthroscopy.

Though many different PROMs have been used in this field, Oxford Hip Score (OHS) and Oxford Knee Score (OKS) are common. This is due to their short length, good response rate, internal consistency and reliability (98), with OHS Cronbach's alpha measures between 0.83 to 0.93 (98), and OKS between 0.87 to 0.93 (99). These strong reliability metrics mean

that OHS and OKS have been used regularly in retrospective validity studies (2,25,26,86,87).

Short-term studies that have looked into the validity of retrospective PROMs, immediately or up to 2 weeks after a hip and/or knee arthroplasty, have generally found reliable and reproducible results (67,86,100). These studies have recommended the use of retrospective PROMs in place of contemporaneous PROMs, in order to obtain baseline data.

A limited number of studies have looked at the validity of retrospective PROMs at longer time-points following hip arthroplasty. Medium term studies, which tend to look at time frames longer than 3 days and shorter than a year after surgery, have also generally found positive results regarding the use of retrospective PROMs (2,25). Marsh et al. examined 173 patients and concluded that patients could accurately recall their preoperative status, 6 weeks after their hip arthroplasty (2). Marsh et al. reported that there was a strong correlation between ratings for the region-specific measures (Pearson's $r = 0.89$ for WOMAC, 0.87 for the Oxford Hip Score, and $r = 0.86$ for the Lower Extremity Functional Scale) (2). However, there was only a moderate correlation between recalled and preoperative ratings of the general health measures ($r = 0.63$ for the feeling thermometer, $r = 0.62$ for the SF-12 physical component score, and $r = 0.48$ for the SF-12 mental component score). The high correlations, along with the lack of any systematic changes, caused Marsh et al. to recommend the use of retrospective PROMs regarding functional status, quality of life and general health, in place of prospectively collected baseline data (2).

The study by Howell et al. evaluated a patient's ability to recall their preoperative pain and function during the post-operation period 6 weeks and 3 months after a total hip arthroplasty (25). The study questioned 104 patients and looked primarily at the patients' ability to recall their quality of life before their surgery. The ICC between recalled and preoperative scores remained high throughout the study (6 weeks, Spearman $\rho: 0.714 - 0.804$, ICC: $0.769 - 0.878$; 3 months, Spearman $\rho: 0.743 - 0.923$, ICC: $0.847 - 0.958$). The only moderate correlation found was the recalled WOMAC stiffness scores obtained 6 weeks postoperatively. Therefore, this study also recommended the use of retrospective PROMs taken not only at six weeks postoperatively, but also up to three months following a hip arthroplasty.

Long-term studies, looking at retrospective PROMs taken a year after an arthroplasty, have generally found that the retrospective PROMs were unreliable and could not be used in place of prospective PROMs for a baseline measure (26,87). Murphy et al. (26) and Yeoman et al. (87) both used OHS and OKS. Murphy et al. examined 113 patients who completed

these retrospective measures approximately 12 months following surgery (26). The findings in Murphy et al.'s study showed no significant difference between prospective and retrospective pre-operative status in their patient cohort (26), for either OKS (MD -0.11, ± 7.34 (SD), $p = 0.912$) or OHS (MD 0.8, ± 6.21 , $p = 0.329$). However, the authors reported that individual differences were greater for both OKS ($r = 0.15 - 0.71$; MD 5.41, $p < 0.001$) and OHS ($r = 0.28 - 0.48$; MD 5.24, $p < 0.001$). From these results they concluded that individual retrospective PROMs were unreliable and had poor correlation with the corresponding prospective outcome measure.

The results of the study by Yeoman et al. (87), closely mirrored Murphy et al.'s (26) findings. Yeoman et al. found excellent reliability for retrospective PROMs a year after an arthroplasty, indicated by ICCs for both OHS ($r = 0.802$) and OKS ($r = 0.772$) (87). Yeoman et al. also observed negligible mean differences of 0.04 points for OHS ($p = 0.97$) and 1.59 points for OKS ($p = 0.10$), between the retrospective and prospective PROMs, indicating no apparent recall bias (87). However, as with Murphy et al. (26), Yeoman et al. (87) concluded that this reliability did not translate to an individual basis, for either OHS ($r = 0.670$) or OKS ($r = 0.629$). Murphy et al. (26) and Yeoman et al. (87) both concluded that an individual's responses to retrospective PROMs taken a year after arthroplasty could not be used in the place of prospective PROMs, as a baseline measure (26,87). To both reach this conclusion, it is unclear exactly what analyses Murphy et al. (26) and Yeoman et al. (87) applied. It appears that they may have correlated individual OHS and OKS item responses for each participant. Since the responses to individual items within these tools are not validated measures, their interpretations are questionable. The total OHS and OKS scores for both studies, which showed little systematic change and high agreement, indicate strong reliability, and so support the validity of retrospective PROMs.

All of these particular studies, except for Bryant et al. (100), dealt with elderly cohorts with participants ranging in average age between 61 to 72.5 years old. This aspect of these studies could be primarily due to low rates of hip and knee arthroplasties in younger people (97). It is possible that the elderly nature of the cohorts in these studies may have compromised the reliability of the retrospective outcomes, especially in the longer-term studies. A few of these studies tried to address this matter, but found differing results (25,26,86). Kwong et al. (86) and Murphy et al. (26) found that age, gender, and other pain had no influence on recollection accuracy. Murphy et al. found no change to their recollection score after adjusting for age, gender, and type of arthroplasty, despite reported age-ranges from 37 to 84 years for hip and from 52 to 81 years for knee arthroplasty (26). However, the study from Howell et al. found that age did have an effect on recollection

accuracy, with patients under 65 having a strong recollection (ICC 0.85 – 0.95), and patients over 65 having only a medium to strong recollection (ICC 0.60 – 0.89) (25). Howell et al. also found that gender had an insignificant effect on recall reliability (25). A summary of the studies concerning retrospective PROMs after hip and knee arthroplasty appears in Table 4.

Table 4. Studies Investigating Retrospective PROMs after Hip and Knee Arthroplasty

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Howell et al., 2008 (25)	n = 104 Mean age 61 years 3 months	WOMAC, OHS, SF-12	Patients were able to accurately recall the prospective status up to three months after a hip arthroplasty (ICC = 0.85 – 0.96). Supported obtaining baseline data up to three months after surgery.
Marsh et al., 2009 (2)	n = 174 Mean age 70.6 years 6 weeks	Lower Extremity Functional Scale, WOMAC, OHS, Short Form-12 health survey (SF-12), and the feeling thermometer	Patients could recall their prospective status six weeks after a hip arthroplasty, with high accuracy in disease-related PROMs (ICC = 0.86 – 0.88) and moderate accuracy for general PROMs (ICC = 0.48 – 0.60).
Murphy et al., 2015 (26)	n = 113 Mean age 63 years THA, 68.5 years TKA. Mean of 12.4 months	OHS or OKS	There was a large absolute difference and weak correlation between retrospective and prospective PROMs in both OKS (MD 5.41; r = 0.61) and OHS (MD 5.24; r = 0.7).
Kwong et al., 2018 (86)	No. 400 Mean age 69.1 years THA, 67.7 years TKA. Shortly after surgery	OHS or OKS, EQ-5D-3L	Patients could accurately recall the prospective status shortly after surgery, with disease-related PROMs (ICC = 0.87). Recommended use of retrospective PROMs when contemporary collection is not possible.
Lawson et al., 2020 (67)	No. 88 Mean age 68 years Shortly after surgery	EQ-5D-3L	High agreement at group level, moderate to substantial agreement at an individual level ($P_c = 0.76$, 95% CI = 0.66, 0.84; $P_c = 0.46$, 95% CI = 0.28, 0.61). Recommended use of retrospective PROMs following surgery.

Author/Study	Study Details	Scores/Outcome measures	Main Findings
Bryant et al., 2006 (100)	No. 398 Mean age 39 years 2 weeks	KOOS, ACL-QOL, WOMET, IKDC, SF-36	Agreement for disease-related PROMs was excellent (ICC = 0.81 – 0.93) and moderate for general mental health (ICC = 0.67). Recommended use of retrospective data in place of prospective data.
Yeoman et al., 2018 (87)	No. 335 Mean age 72.5 years 1 year	OHS or OKS	Recalled status a year after surgery was a reliable measure for assessing a cohort (OHS, $r = 0.8$; OKS, $r = 0.78$). However, the individual retrospective score was not reliable due to a diminished reliability (OHS, $r = 0.67$; OKS, $r = 0.63$). Did not recommend the use of retrospective PROMs.

1.5 PROMs and the Hip

1.5.1 The Use of PROMs in Hip Arthroscopy

To some extent, the validity and reliability of PROMs in hip arthroscopy are influenced by the nature of the hip joint and hip pathologies. The morphological features of the hip that contribute to soft tissue injury often manifest in younger and active patients, who are likely to have different expectations and awareness of hip function, and possibly stronger memory of symptoms, which may have been less long-standing. These factors may impact on issues such as recall bias and response shift associated with retrospective PROMs. For this reason, some discussion of hip anatomy and conditions of the hip can be useful in understanding the full extent of how they may impact upon retrospective PROMs.

1.5.2 Nature of the Hip Joint

Briefly, the hip is a true 'ball and socket joint,' located inferior to the iliac crest, anterior and lateral to the gluteal region. It displays remarkable stability whilst maintaining a wide range of motion in several physical planes (101). The hips are subjected to large amount torsional and axial forces, as they act as the structural link between the axial

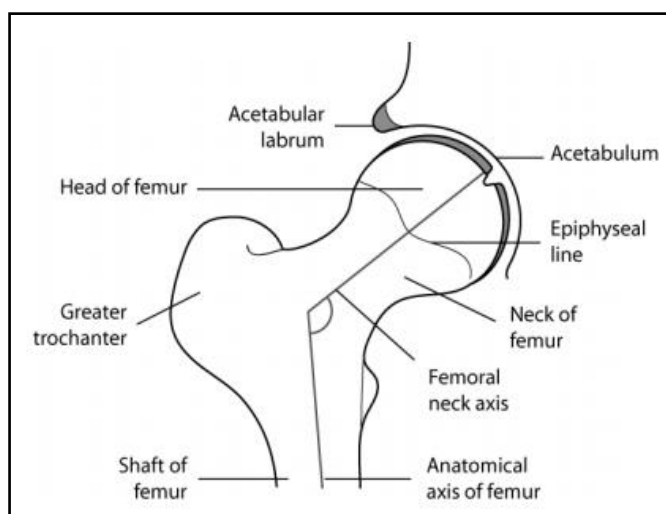


Figure 1, Basic anatomy of the Hip. Adapted from Byrne et al. (2010) (82)

skeleton and the lower extremities. Therefore, the hip is crucial in physical activities as it deals with forces from the ground up, and also carries forces from the trunk, upper extremities, head and neck (102). This means that any injury or issue with the hip usually has a very significant impact on a patient's mobility, quality of life and ability to move, whether in sports and leisure, or otherwise.

1.5.3 Anatomy of the Hip

In terms of the 'socket,' the acetabulum is formed by the innominate bone or pelvis, is composed of three bones, the pubis (approximately 20%), ilium (40%) and ischium (40%) (103). In skeletal development, these three are separated by a triradiate cartilage with fusion between the bones occurring typically between the ages of 15 to 17, a process usually complete by the age of 25 (104).

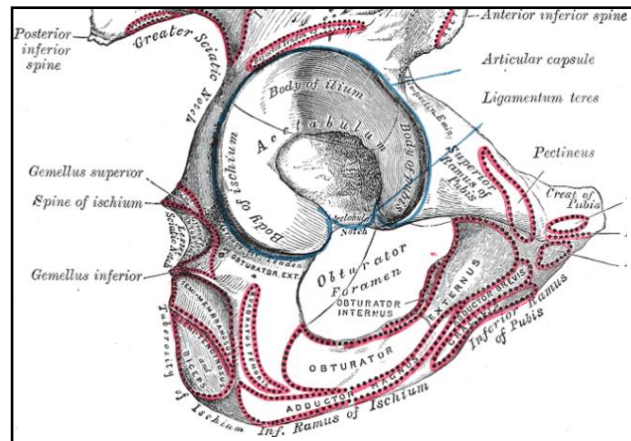


Figure 2, External surface of Pelvis, Joint capsule, n.d. (136)

The articular surface of the acetabulum appears lunate or moon shaped. The fossa, formed within the centre of the moon-shaped cartilage, is filled with synovial covered fat, and contains the attachment and origin of ligamentum teres. Inferior to the fossa, lies the inferior transverse ligament which completes the socket of the hip (101).

Attached to the rim of the acetabulum is a horseshoe-shaped fibrocartilaginous structure called the acetabular labrum. The inferior aspects of the labrum, the anterior and posterior horns, are also connected by the transverse ligament (105). However, the acetabulum labrum does not play a significant role in joint stability as it does in the glenoid labrum in the shoulder. The acetabular labrum helps distribute forces around the joint and plays a crucial role in joint development (101).

This junction of the articular cartilage and labrum has been observed as the most likely place for labral tears to occur, leading it to being dubbed the 'watershed region' (106). Acetabular labral tears have been described as a common cause of hip pain in active adults (107). They can be debilitating and severely impact sporting activities. Hip arthroscopies have identified labral tears as cause of moderate to severe groin pain, and long-term hip and groin pain in professional hockey players (108). Intraarticular hip pathologies frequently cause groin/hip pain (109), and can be related to a considerable amount of morbidity in active young adults (108,110).

As to the other main part of the hip joint, the head of the femur is the 'ball' that fits into the 'socket.' It is covered by an articular cartilage that reaches beyond the acetabular rim to accommodate the hip's full range of motion (101).

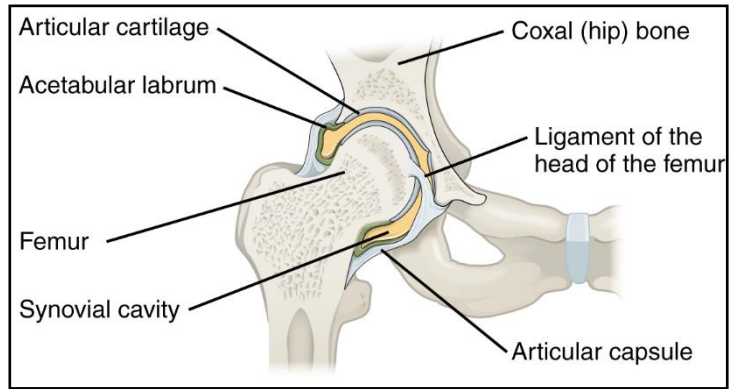


Figure 3, Front section through the right hip joint, n.d (137)

1.5.4 Hip Pathologies

One common intra-articular hip pathology is osteoarthritis which typically occurs in patients over the age of 50. Osteoarthritis can also occur in people under the age of 50 and can be caused by bony abnormalities, such as femoroacetabular impingement (FAI) (111).

FAI is a condition that occurs due to early contact of the proximal head of the femur and the acetabulum in extreme range of motion (112).

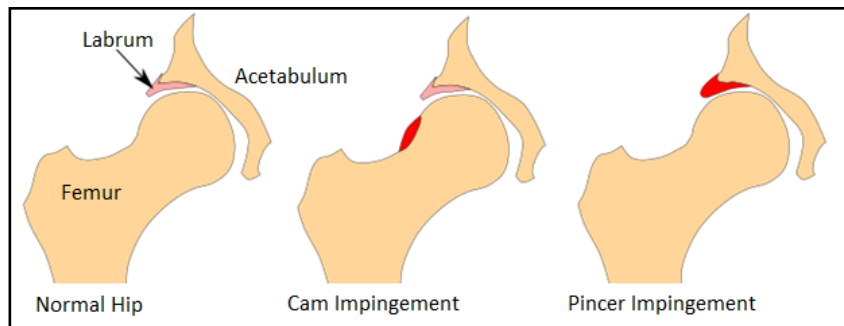


Figure 4, Different types of Femoroacetabular Impingement, n.d (138)

This can be due to an abnormal morphology of

the head of the femur (cam) and/or the acetabulum (pincer), or due to exposing the hip to excessive forces, beyond the hip's physiological range of motion (113). A typical FAI is due to a cumulation of these factors, as a patient with abnormal hip morphology can have a reduced range of motion, leading to greater susceptibility to damaging impingement (114). In addition, an abnormality in the anterosuperior femoral head, which can occur in a cam lesion, may lead to damage to the acetabulum, making it a cause of osteoarthritis development (115,116).

FAI is a major cause of early arthrosis of the hip (117). It typically presents in active young adults in their 20s – 40s as a slow onset of groin pain, beginning after a minor trauma. In the initial stages of the condition, the pain can be exacerbated by demanding activities and may be intermittent. Pain can also be present after sitting for extended periods (114), which further increases the adverse impact and disability of this condition. Through radiology, evidence of FAI has been found in up to 29% of young active asymptomatic athletes (118) and structural abnormalities predisposing to FAI in up to 48% of asymptomatic adults (119).

FAI is strongly associated with chondrolabral pathology, leading to micro-instability, hip migration and increased severity of labral pathologies (120,121).

1.5.5 Hip Surgery

Due to the current understanding of morphological aetiology of FAI, a range of surgical approaches have been developed to treat this condition, usually with femoral or acetabular osteoplasty, labral repair and mitigation of chondral lesions (106,122,123). These approaches include surgical dislocation of the hip (124), combined anterior hip exposure and arthroscopy (108,125–128). The aims of these surgeries are to improve movement and function, ease pain, and to protect the natural hip against degradation and degeneration caused by FAI (117). With appropriate indications, realistic patient expectations, sound patient selection, hip arthroscopy has been established as a reliable treatment for FAI, labral pathology, loose bodies, focal articular cartilage lesions and synovial abnormalities (111).

A systematic review by Kemp et al, 2012, examined the efficacy of hip arthroscopies and non-surgical interventions for intra-articular hip pathologies. The review analysed twenty-nine studies. The studies investigated pain relief and functional improvement, ranging from 1 year to 10 years after surgery. Evidence showed that hip arthroscopy alone can significantly improve function and reduce pain with patients who have intra-hip pathologies (128).

Although it was initially described in 1931, hip arthroscopy has been gaining popularity since the 1980s (111). Over recent decades, surgery for hip pain has undergone remarkable development. As understanding of hip pathology and surgical interventions has increased, there has been a greater emphasis placed on addressing abnormal bony morphology (FAI) (111).

One of the main surgical procedures for dealing with FAI, particularly with younger patients, is hip arthroscopy. This is a less invasive surgery than open reduction and may delay or prevent the need for hip resurfacing or replacement. The procedure involves a small incision into the hip and the insertion of an arthroscope to examine and treat intra-articular pathologies (129). Hip arthroscopy has a number of advantages. Not only is it less invasive than open procedures, but it also allows a clear view of the internal structures of the hip. It creates a means to view the internal anatomy of the hip, such as the acetabulum, femoral surfaces, the labrum, the posterior-superior capsule and other internal structures (130). This makes arthroscopy not only useful as a surgical procedure, but also as a means of scoping the internal state of a hip before further action is taken.

Early intra-articular and periarticular damage originating from FAI can progress to the development of osteoarthritis, if underlying impingement is not addressed (114). Hip arthroscopy can alleviate pain, restore function and improve quality of life, and extend the quality of life of the native hip joint by stopping the processes which precipitate osteoarthritis (112,116,120). Patients who have undergone hip arthroplasty usually return to sports and experience better short- and medium-term outcomes, compared to conservative treatment (128,131–133).

The Tönnis osteoarthritis grading system (134) is designed to measure radiographic changes in the hip due to degeneration, which can affect the success of hip arthroscopy. The Tönnis grading system categorises the extent of osteoarthritis from no changes (grade 0), mild changes (grade 1), moderate changes (grade 2), to severe changes (grade 3) (110). In the presence of pre-existing hip pathology, with the Tönnis grade being greater than 2 and the joint space width being less than 2 mm, the outcomes of hip arthroscopy are worse than those without pre-existing articular damage. This often leads to total hip arthroplasty or arthroscopic revision surgery after an arthroscopy (123).

1.5.6 Retrospective PROMs in Arthroscopy

The relevance of hip anatomy, pathologies, and treatment by hip arthroscopy on validity of retrospective PROMs may lie in three main areas. The first area is concerned with the validity of retrospective PROMs that arises because the underlying pathology is usually highly impactful for the patient. The hip is crucial in physical activities and is even more so in young people who are generally more active. As noted, the hip deals with forces from the ground up and also carries forces from the trunk and upper body. Therefore, any injury to the hip usually has a significant impact on quality of life and mobility. In particular, acetabular labral tears are painful and debilitating, and are a common cause of significant hip pain (135). Further, FAI is a major cause of early arthrosis of the hip and can be successfully treated by hip arthroscopy.

All these things mean that, in the event of a successful hip arthroscopy which either deals with an acetabular labral tear or addresses underlying bony structural morphology precipitating FAI, it is likely that, at the time a patient retrospectively considers their preoperative condition, these typically active patients are more likely to have a clear memory of the pain and impact that was present prior to the operation. These factors can provide an 'anchor-point' for a patient's memory and therefore lessen any tendency for the patient to assess their retrospective condition by simply extrapolating back from their current status. Furthermore, the severity of hip pathologies, and the active status of this cohort, is likely to

mean that patients could have a clearer appreciation of the deficits caused by the injury and improvement gained by having the arthroscopy. All these matters can potentially increase the reliability of retrospective PROMs in arthroscopy.

The second area where the nature of hip arthroscopy may affect PROMs lies in the fact that the procedure itself is less invasive and has quicker recovery time. It has a lesser impact on the patient than a hip arthroplasty, not only in terms of the operation but also with regards rehabilitation. Consequently, the patient may be less likely to confuse their preoperative condition with any difficulties they experienced during or after the operation. This in turn may serve to diminish the effect of recall bias on retrospective PROMs for hip arthroscopy and therefore enhance their reliability.

The final main area which may affect the reliability of retrospective PROMs for hip arthroscopy is the fact that the procedure invariably involves a younger cohort of patients, whose injuries generally have shorter chronicity than for patients undergoing hip replacement. The result of this is that factors such as chronicity, age-related memory deficit and the effect of comorbidities are likely to be less relevant in hip arthroscopy compared to arthroplasty patients. In addition, hip arthroscopies have a high success rate and effectively deal with restoring function and improving quality of life. Particularly in the case of younger people who are active, these factors may also be relevant in diminishing the effect of recall bias. This is because the combination of restored hip function, quick recovery and restoration of physical capability means that the patient might be likely to have a relatively accurate recollection of their preoperative function.

Given the lack of relevant literature in hip arthroscopy or in younger age-groups, there would appear to be real benefit in a study which investigates retrospective PROMs in hip arthroscopy in order to confirm whether they are in fact valid and reliable both for clinical practice and research. The importance of such a study also arises due to fact that there is currently a predominance of research on joint-replacement procedures and a lack of information concerning the validity of retrospective PROMs relating to the lower body in younger and more active patients.

References

1. McClimans LM, Browne J. Choosing a patient-reported outcome measure. *Theor Med Bioeth.* 2011;31(1):48–59.
2. Marsh J, Bryant D, MacDonald SJ. Older Patients Can Accurately Recall Their Preoperative Health Status Six Weeks Following Total Hip Arthroplasty. *J Bone Jt Surgery-American Vol* [Internet]. 2009;91(12):2827–37. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004623-200912000-00007>
3. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. Routine use of patient reported outcome measures in healthcare settings. *BMJ.* 2010;340:464–7.
4. Ruseckaite R, Maharaj AD, Krysinska K, Dean J, Ahern S. Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries. *Patient Relat Outcome Meas.* 2019;10(10):355–72.
5. Holmes C, Briffa N. Patient-Reported Outcome Measures (PROMS) in patients undergoing heart valve surgery: why should we measure them and which instruments should we use? *Open Hear.* 2016;3(1):e000315.
6. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. *J Psychosom Res.* 2010;68(4):319–23.
7. Zhang Y, Alyass A, Vanniyasingam T, Sadeghirad B, Flórez ID, Pichika SC, et al. A systematic survey of the methods literature on the reporting quality and optimal methods of handling participants with missing outcome data for continuous outcomes in randomized controlled trials. *J Clin Epidemiol.* 2017;88:67–80.
8. Van den Bergh O, Walentynowicz M. Accuracy and bias in retrospective symptom reporting. *Curr Opin Psychiatry.* 2016;29(5):302–8.
9. Gomes M, Gutacker N, Bojke C, Street A. Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance. *Heal Econ (United Kingdom).* 2016;25:515–28.
10. Palmer AJR, Rombach I, Cooper CD, Pollard TC, Hollinghurst D, Taylor A, et al. Protocol for the Femoroacetabular Impingement Trial (FAIT) A MULTI-CENTRE

- RANDOMISED CONTROLLED TRIAL COMPARING SURGICAL AND NON-SURGICAL MANAGEMENT OF FEMOROACETABULAR IMPINGEMENT. *Bone Jt Res.* 2014;3(11):321–7.
11. Sheila Sprague, Frede Frihagen, Ernesto Guerra, Mohit Bhandari, Thomas A Einhorn, Ydo Vincent Kleinlugtenbelt, Rudolf W. Poolman, Sofia Bzovsky RSP. Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. *N Engl J Med.* 2019;381(21).
 12. Nauth A, Creek AT, Zellar A, Lawendy AR, Dowrick A, Gupta A, et al. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *Lancet.* 2017;389(10078):1519–1527.
 13. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain. *Arthritis Care Res (Hoboken).* 2011;63(11):S210–52.
 14. Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, et al. A checklist for retrospective database studies - Report of the ISPOR task force on retrospective databases. *Value Heal.* 2003;6(2):90–7.
 15. Parsons NR, Costa ML, Achten J, Griffin XL. Baseline quality of life in people with hip fracture: Results from the multicentre WHiTE cohort study. *Bone Jt Res.* 2020;9(8):468–76.
 16. Uttil B, White CA, Cnudde K, Grant LM. Prospective memory, retrospective memory, and individual differences in cognitive abilities, personality, and psychopathology. *PLoS One.* 2018;13(3):1–23.
 17. Helfand BT, Fought A, Manvar AM, McVary KT. Determining the utility of recalled lower urinary tract symptoms. *Urology.* 2010;76(2):442–7.
 18. Thorborg K, Roos E, Bartels E, Petersen J, Holmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *Br J Sports Med [Internet].* 2010;44(16):1186–96. Available from: <http://bjsm.bmj.com/cgi/doi/10.1136/bjsm.2009.060889>
 19. Griggs CL, Schneider JC, Kazis LE, Ryan CM. Patient-reported outcome measures: A stethoscope for the patient history. *Ann Surg.* 2017;256(6):1066–7.
 20. Kitaoka HB, Meeker JE, Phisitkul P, Adams SB, Kaplan JR, Wagner E. AOFAS Position Statement Regarding Patient-Reported Outcome Measures. *Foot Ankle Int.*

- 2018;39(12):1389–93.
21. Teela L, van Muilekom MM, Kooij LH, Gathier AW, van Goudoever JB, Grootenhuis MA, et al. Clinicians' perspective on the implemented KLIK PROM portal in clinical practice. *Qual Life Res.* 2020;1–11.
 22. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. In: 2nd editio. Hoboken: Wiley; 2002.
 23. Bahl V, Hu HM, Henke PK, Wakefield TW, Campbell DA, Caprini JA. A validation study of a retrospective venous thromboembolism risk scoring method. *Ann Surg.* 2010;251(2):344–50.
 24. Snelgrove R, Havitz ME. Looking back in time: The pitfalls and potential of retrospective methods in leisure studies. *Leis Sci.* 2010;32(4):337–51.
 25. Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS. A Comparison Between Patient Recall and Concurrent Measurement of Preoperative Quality of Life Outcome in Total Hip Arthroplasty. *J Arthroplasty [Internet]*. 2008 [cited 2017 May 4];23(6):843–9. Available from: <http://www.sciencedirect.com.libproxy.unitec.ac.nz/science/article/pii/S0883540307004627>
 26. Murphy MT, Vardi R, Journeaux SF, Whitehouse SL. A patient's recollection of pre-operative status is not accurate one year after arthroplasty of the hip or knee. *Bone Jt J.* 2015;97-B(8):1070–5.
 27. Aleem IS, Duncan J, Ahmed AM, Zarrabian M, Eck JC, Rhee J, et al. Do lumbar decompression and fusion patients recall their preoperative status? A cohort study of recall bias in patient-reported outcomes. *Spine (Phila Pa 1976) [Internet]*. 2016;(In press):1–28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/?term=Do+Lumbar+Decompression+and+Fusion+Patients+Recall+Their+Preoperative+Status%3F+A+Cohort+Study+of+Recall+Bias+in+Patient-Reported+Outcomes>
 28. Ellwood PM. Shattuck Lecture - outcomes management. A technology of patient experience. *N Engl J Med.* 1988;318(23):1549–56.
 29. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S. Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res.*

- 2013;22(3):475–83.
30. Porter ME. What Is Value in Health Care? *N Engl J Med*. 2010;363(26):2477–81.
 31. Fleurence R, Whicher D, Dunham K, Gerson J, Newhouse R, Luce B. The Patient-centered outcomes research Institute's Role in advancing methods for patient-centered outcomes research. *Med Care*. 2015;53(1):2–8.
 32. Jackowski D, Guyatt G. A Guide to Health Measurement. *Clin Orthop Relat Res*. 2003;413:80–9.
 33. Gagnier JJ. Patient reported outcomes in orthopaedics. *J Orthop Res*. 2017;35(10):2098–108.
 34. Christensen DL, Dickens JF, Freedman B, Mauntel T, Owens BD, Potter BK, et al. Patient-reported outcomes in orthopaedics. *Journal of Bone and Joint Surgery - American Volume*. 2018. p. 436–42.
 35. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, De Vet HCW. Outcome instruments: Rationale for their use. *J Bone Jt Surg*. 2009;91(3):41–9.
 36. Elsenbeck M, Tucker CJ, Dickens JF. Patient Reported Outcomes: The Currency of Value-Based Healthcare. *Sports medicine update*. 2018;1(1):2–6.
 37. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Med Care*. 2007;45(5):S3–11.
 38. Stepan JG, London DA, Boyer MI, Calfee RP. Accuracy of patient recall of hand and elbow disability on the quickdash questionnaire over a two-year period. *J Bone Jt Surg - Ser A*. 2013;95:1–8.
 39. Christensen KB, Comins J, Krogsgaard MR, Brodersen J, Jensen J, Hansen CF, et al. Psychometric validation of PROM instruments. Article four in a series of ten. *Scand J Med Sci Sports*. 2020;1:1–14.
 40. Vitale MG, Levy DE, Johnson MG, Gelijns AC, Moskowitz AJ, Roye BP, et al. Assessment of quality of life in adolescent patients with orthopaedic problems: Are adult measures appropriate? *J Pediatr Orthop*. 2001;21:622–628.

41. Suryavanshi JR, Goto R, Jivanelli B, Aberdeen J, Duer T, Lam KC, et al. Age-Appropriate Pediatric Sports Patient-Reported Outcome Measures and Their Psychometric Properties: A Systematic Review. *Am J Sports Med.* 2019;47(13):3270–6.
42. Thoomes-de Graaf M, Scholten-Peeters GGM, Schellingerhout JM, Bourne AM, Buchbinder R, Koehorst M, et al. Evaluation of measurement properties of self-administered PROMs aimed at patients with non-specific shoulder pain and “activity limitations”: a systematic review. *Qual Life Res.* 2016;25(9):2141–2160.
43. Fitzner K. Reliability and validity: A quick review. *Diabetes Educ.* 2007;33(5):775–80.
44. Davidson M, Keating J. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *Br J Sports Med.* 2014;48:792–796.
45. Hawkins RJ. Recommendations for Evaluating and Selecting Appropriately Valued Outcome Measures. *Instr Course Lect.* 2016;65:587–91.
46. Fidai MS, Saltzman BM, Meta F, Lizzio VA, Stephens JP, Bozic KJ, et al. Patient-Reported Outcomes Measurement Information System and Legacy Patient-Reported Outcome Measures in the Field of Orthopaedics: A Systematic Review. *Arthrosc - J Arthrosc Relat Surg.* 2018;34(2):605–14.
47. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res.* 2018;27(5):1171–9.
48. DeVellis RF. Classical test theory. *Med Care.* 2006;44(11):S50–9.
49. Wang W, Song L, Chen P, Ding S. An Item-Level Expected Classification Accuracy and Its Applications in Cognitive Diagnostic Assessment. *J Educ Meas.* 2019;56(1):51–75.
50. Schauber SK, Hecht M, Nouns ZM. Why assessment in medical education needs a solid foundation in modern test theory. *Adv Heal Sci Educ.* 2018;23(1):217–32.
51. Huskisson EC. Measurement fo pain. *Lancet.* 1974;304(7889):1127–31.
52. Aitken RCB. A Growing Edge of Measurement of Feelings: Measurement of Feelings Using Visual Analogue Scales. *J R Soc Med.* 1969;62:989–93.

53. Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: A Short-form Hip Replacement Survey. *Clin Orthop Relat Res*. 2016;474(6):1472–82.
54. Hamilton DF, Giesinger JM, MacDonald DJ, Simpson AHRW, Howie CR, Giesinger K. Responsiveness and ceiling effects of the Forgotten Joint Score-12 following total hip arthroplasty. *Bone Jt Res*. 2016;
55. Puliero B, Blakeney WG, Beaulieu Y, Vendittoli PA. Joint Perception After Total Hip Arthroplasty and the Forgotten Joint. *J Arthroplasty*. 2018;34(1):65–70.
56. Christensen CP, Althausen PL, Mittleman MA, Lee J, McCarthy JC. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res [Internet]*. 2003;406(406):75–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12579003>
57. Klässbo 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.
58. Griffin DR, Parsons N, Mohtadi NGH, Safran MR. A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. *Arthrosc - J Arthrosc Relat Surg*. 2012;28(5):611–8.
59. Chang EM, Gillespie EF, Shaverdian N. Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality. *Patient Relat Outcome Meas*. 2019;10:171–86.
60. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol*. 1990;43(1):87–91.
61. Blome C, Augustin M. Measuring change in quality of life: Bias in prospective and retrospective evaluation. *Value Heal*. 2015;18(1):110–5.
62. Raphael K. Recall bias: A proposal for assessment and control. *Int J Epidemiol*. 1987;16(2):167–70.
63. Schmier J, Halpern MT. Patient recall and recall bias of health state and health status. *Expert Rev Pharmacoeconomics Outcomes Res*. 2004;4(2):159–63.
64. Gotlin MJ, Kingery MT, Baron SL, McCafferty J, Jazrawi LM, Meislin RJ. Recall Bias in Retrospective Assessment of Preoperative Patient-Reported American Shoulder

- and Elbow Surgeons Scores in Arthroscopic Rotator Cuff Repair Surgery. *Am J Sports Med* [Internet]. 2020;48(6):1471–5. Available from: <https://doi.org/10.1177/0363546520913491>
65. Baddeley A. *Human memory. Theory and practice*. Book. Hove, East Sussex: Psychology Press; 2005. 420 p.
 66. Howard GS, Dailey PR. Response-shift bias: A source of contamination of self-report measures. *J Appl Psychol*. 1979;64(2):144–50.
 67. Lawson A, Tan AC, Naylor J, Harris IA. Is retrospective assessment of health-related quality of life valid? *BMC Musculoskelet Disord*. 2020;21(1):1–10.
 68. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. *Health Qual Life Outcomes*. 2010;8(1):1–9.
 69. Loke YK. Use of databases for clinical research. *Archives of Disease in Childhood*. 2014. p. 587–589.
 70. Briffa N. The employment of Patient-Reported Outcome Measures to communicate the likely benefits of surgery. *Patient Relat Outcome Meas*. 2018;9:263–6.
 71. Nachira D, Bertolaccini L, Ismail M, Chiappetta M, Meacci E, Margaritora S. How to create a surgical database? *J Thorac Dis*. 2018;10(11):6352–5.
 72. Ronai CR. Multiple reflections of child sex abuse: An Argument for a Layered Account. *J Contemp Ethnogr*. 1995;23(4):395–426.
 73. Hardt J, Rutter M. Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *J Child Psychol Psychiatry Allied Discip*. 2004;45(2):260–73.
 74. Schenker N, Raghunathan TE, Chiu PL, Makuc DM, Zhang G, Cohen AJ. Multiple imputation of missing income data in the national health interview survey. *J Am Stat Assoc*. 2006;101(475):924–33.
 75. Royston P. Multiple imputation of missing values: Update. *Stata J*. 2004;4(3):227–41.
 76. Wayman JC. *Multiple Imputation For Missing Data: What Is It and How Can I Use It?* *Res Methods Psychol*. 2003;2:1–16.

77. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*. 2009;339(7713):157–60.
78. Allison PD. Multiple imputation for missing data: A cautionary tale. *Sociol Methods Res*. 2000;28(3):301–9.
79. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf*. 2010;19(6):618–26.
80. Beese A, Morley S. Memory for acute pain experience is specifically inaccurate but generally reliable. *Pain*. 1993;53(2):183–9.
81. Pellisé F, Vidal X, Hernández A, Cedraschi C, Bagó J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine (Phila Pa 1976)*. 2005;30(3):365–8.
82. Kuittinen P, Aalto T, Heikkilä T, Leinonen V, Savolainen S, Sipola P, et al. Accuracy and reproducibility of a retrospective outcome assessment for lumbar spinal stenosis surgery. *BMC Musculoskelet Disord*. 2012;13(1):83.
83. Tashjian RZ, Bradley MP, Tocci S, Henn RF, Rey J, Green A. A comparison of prospective and retrospective assessment of functional outcome after rotator cuff repair. *J Shoulder Elbow Surg [Internet]*. 2008;17(6):853–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18693118>
84. Reynolds N, Thirkannad S. The recall DASH score - a novel research tool. *Hand Surg*. 2013;18(1):11–4.
85. Wormdal AH, Tallaksen SH, Hagen Ø, Foss OA, Finsen V. Remembered preoperative Quick disabilities of the arm, shoulder and hand (QuickDASH) scores. *J Hand Surg (European Vol [Internet])*. 2017;42(1):78–83. Available from: <http://journals.sagepub.com/doi/10.1177/1753193416665299>
86. Kwong E, Neuburger J, Black N. Agreement between retrospectively and contemporaneously collected patient-reported outcome measures (PROMs) in hip and knee replacement patients. *Qual Life Res*. 2018;27(7):1845-1854.
87. Yeoman TFM, Clement ND, Macdonald D, Moran M. Recall of preoperative Oxford

- Hip and Knee Scores one year after arthroplasty is an alternative and reliable technique when used for a cohort of patients. *Bone Jt Res.* 2018;7(5):351–6.
88. Widnall J, Ralte P, Selvan D, Molloy A. An evaluation of retrospective outcome scores in elective foot and ankle surgery. *Foot* [Internet]. 2014 [cited 2017 Apr 18];24(1):28–30. Available from:
<http://www.sciencedirect.com.libproxy.unitec.ac.nz/science/article/pii/S0958259214000078?np=y&npKey=5556adeb3d2d9d4bfae8ab1c57508794a85884bde2fd1b394f7299404872b9a0>
 89. Babel P. Memory of pain induced by physical exercise. *Memory* [Internet]. 2016;24(4):548–59. Available from:
<http://www.tandfonline.com/doi/abs/10.1080/09658211.2015.1023809>
<http://www.tandfonline.com/doi/full/10.1080/09658211.2015.1023809>
 90. Waldenström U, Irestedt L. Obstetric pain relief and its association with remembrance of labor pain at two months and one year after birth. *J Psychosom Obstet Gynecol.* 2006;
 91. Waldenström U, Irestedt L, Waldenstrom U, Irestedt L, Waldenström U, Irestedt L, et al. Obstetric pain relief and its association with remembrance of labor pain at two months and one year after birth. *J Psychosom Obstet Gynecol* [Internet]. 2006;27(3):147–56. Available from:
[http://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=cmedm&AN=17214449&site=ehost-live&scope=site%5Cnhttp://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=a9h&AN=23022186&site=ehost-live&scope=site%5Cn%3CGo to ISI%3E:/](http://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=cmedm&AN=17214449&site=ehost-live&scope=site%5Cnhttp://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=a9h&AN=23022186&site=ehost-live&scope=site%5Cn%3CGo to ISI%3E/)
 92. Devlin NJ, Appleby J, Buxton M, Vallance-Owen A. Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision making. *Health Econ.* 2010;1:8–12.
 93. Clement N. Patient factors that influence the outcome of total knee replacement: A critical review of the literature. *OA Orthop.* 2013;1(2):1–4.
 94. Willems P. Decision making in surgical treatment of chronic low back pain: The performance of prognostic tests to select patients for lumbar spinal fusion. *Acta Orthop.* 2013;84(349):1–37.
 95. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs

- Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain*. 1993;52(2):157–68.
96. Aleem IS, Duncan J, Ahmed AM, Zarrabian M, Eck J, Rhee J, et al. Do Lumbar Decompression and Fusion Patients Recall Their Preoperative Status? *Spine (Phila Pa 1976)* [Internet]. 2017;42(2):128–34. Available from: <http://insights.ovid.com/crossref?an=00007632-201701150-00018>
 97. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Determining the Need for Hip and Knee Arthroplasty: The Role of Clinical Severity and Patients' Preferences. *Med Care*. 2001;39(3):206–16.
 98. Nilsson A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (A. Arthritis Care Res. 2011;63(S11):S200–7.
 99. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Ou. *Arthritis Care Res*. 2011;63(S11):S208–28.
 100. Bryant D, Norman G, Stratford P, Marx RG, Walter SD, Guyatt G. Patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery. *J Clin Epidemiol*. 2006;59(9):984–93.
 101. Byrne DP, Mulhall KJ, Baker JF. Anatomy & Biomechanics of the Hip. *Open Sport Med J*. 2010;4:51–7.
 102. Alderink GJ. *Joint Structure and Function: A Comprehensive Analysis*, ed 4. *Phys Ther*. 2006;356–69.
 103. Bareither D. *Thieme Atlas of Anatomy: General Anatomy and Musculoskeletal System*. *J Am Podiatr Med Assoc*. 2006;96(2):186.
 104. Oliver J. Moore - *Clinically Oriented Anatomy 7th Edition*. *Journal of Chemical Information and Modeling*. 2013.

105. Nguyen ML, Safran MR. Anatomy: Labrum. In: Hip Joint Restoration. Springer, New York, NY; 2017. p. 23–6.
106. McCarthy JC, Noble PC, Schuck MR, Wright J, Lee J. The role of labral lesions to development of early degenerative hip disease. *Clin Orthop Relat Res*. 2001 Dec 1(393):25–37.
107. Seldes RM, Tan V, Hunt J, Katz M, Winiarsky R, Fitzgerald RH. Anatomy, histologic features, and vascularity of the adult acetabular labrum. In: *Clinical Orthopaedics and Related Research*. 2001. p. 232–40.
108. Philippon MJ, Weiss DR, Kuppersmith DA, Briggs KK, Hay CJ. Arthroscopic Labral Repair and Treatment of Femoroacetabular Impingement in Professional Hockey Players. *Am J Sports Med* [Internet]. 2010;38(1):99–104. Available from: <http://ajs.sagepub.com/lookup/doi/10.1177/0363546509346393>
109. Parvizi J, Bican O, Bender B, Mortazavi SMJ, Purtill JJ, Erickson J, et al. Arthroscopy for Labral Tears in Patients with Developmental Dysplasia of the Hip: A Cautionary Note. *J Arthroplasty*. 2009;24(6):110–3.
110. Burnett RSJ, Della Rocca GJ, Prather H, Curry M, Maloney WJ, Clohisy JC. Clinical presentation of patients with tears of the acetabular labrum. *J Bone Jt Surg - Ser A*. 2006;88(7):1448–57.
111. Ross JR, Larson CM, Bedi A. Indications for Hip Arthroscopy. *Sports Health*. 2017;9(5):402–13.
112. Johnston TL, Schenker ML, Briggs KK, Philippon MJ. Relationship Between Offset Angle Alpha and Hip Chondral Injury in Femoroacetabular Impingement. *Arthrosc - J Arthrosc Relat Surg*. 2008;24(6):669–75.
113. Clohisy JC, Knaus ER, Hunt DM, Leshner JM, Harris-Hayes M, Prather H. Clinical presentation of patients with symptomatic anterior hip impingement. In: *Clinical Orthopaedics and Related Research*. 2009. p. 638–44.
114. Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock K a. Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res*. 2003;(417):112–20.
115. Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock KA. Femoroacetabular

- impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res.* 2003;(417):112–20.
116. Ganz R, Leunig M, Leunig-Ganz K, Harris WH. The etiology of osteoarthritis of the hip: An integrated mechanical concept. *Clin Orthop Relat Res.* 2008;466:264–72.
 117. Tannast M, Siebenrock KA, Anderson SE. Femoroacetabular impingement: Radiographic diagnosis - What the radiologist should know. Vol. 188, *American Journal of Roentgenology.* 2007. p. 1540–52.
 118. Abellàn J, Esparza F, Blanco A, Martínez M, Merino GR, Lisón A. Radiological evidence of femoroacetabular impingement in asymptomatic athletes. *Br J Sports Med [Internet].* 2011 Apr 1;45(4):333 LP – 333. Available from: <http://bjsm.bmj.com/content/45/4/333.1.abstract>
 119. Kang ACL, Gooding AJ, Coates MH, Goh TD, Armour P, Rietveld J. Computed tomography assessment of hip joints in asymptomatic individuals in relation to femoroacetabular impingement. *Am J Sports Med.* 2010;38(6):1160–5.
 120. Eijer H, Hogervorst T. Femoroacetabular impingement causes osteoarthritis of the hip by migration and micro-instability of the femoral head. *Med Hypotheses.* 104(2017):93–6.
 121. Lievense AM, Bierma-Zeinstra SMA, Verhagen AP, Verhaar JAN, Koes BW. Influence of hip dysplasia on the development of osteoarthritis of the hip. *Ann Rheum Dis.* 2004;63(6):621–6.
 122. Hunt D, Prather H, Hayes MH, Clohisy JC. Clinical Outcomes Analysis of Conservative and Surgical Treatment of Patients With Clinical Indications of Prearthritic, Intra-articular Hip Disorders. *Pm&R [Internet].* 2012;4(7):479–87. Available from: <http://dx.doi.org/10.1016/j.pmrj.2012.03.012>
 123. Kemp JL, MacDonald D, Collins NJ, Hatton AL, Crossley KM. Hip Arthroscopy in the Setting of Hip Osteoarthritis: Systematic Review of Outcomes and Progression to Hip Arthroplasty. *Clin Orthop Relat Res.* 2015;473(3):1055–1073.
 124. Yun HH, Shon WY, Yun JY. Treatment of femoroacetabular impingement with surgical dislocation. *Clin Orthop Surg.* 2009;1(3):146–54.
 125. Laude F, Sariali E, Nogier A. Femoroacetabular impingement treatment using

- Arthroscopy and Anterior approach. In: Clinical Orthopaedics and Related Research. 2009. p. 747–52.
126. Byrd JWT, Jones KS. Arthroscopic Femoroplasty in the management of cam-type Femoroacetabular impingement. In: Clinical Orthopaedics and Related Research. 2009. p. 739–46.
127. Guanche CA, Bare AA. Arthroscopic treatment of femoroacetabular impingement. *Arthrosc - J Arthrosc Relat Surg* [Internet]. 2006;22(1):95–106. Available from: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L43053188%5Cnhttp://dx.doi.org/10.1016/j.arthro.2005.10.018>
128. Kemp JL, Collins NJ, Makdissi M, Schache a. G, Machotka Z, Crossley K. Hip arthroscopy for intra-articular pathology: a systematic review of outcomes with and without femoral osteoplasty. *Br J Sports Med*. 2012;46(9):632–43.
129. McCarthy JC, Day B, Busconi B. Hip Arthroscopy: Applications and Technique. *J Am Acad Orthop Surg*. 1995;3:115–22.
130. Kelly BT, Weiland DE, Schenker ML, Philippon MJ. Arthroscopic labral repair in the hip: Surgical technique and review of the literature. *Arthroscopy - Journal of Arthroscopic and Related Surgery*. 2005.
131. Minkara AA, Westermann RW, Rosneck J, Lynch TS. Systematic Review and Meta-analysis of Outcomes After Hip Arthroscopy in Femoroacetabular Impingement. *Am J Sports Med*. 2019;47(2):488–500.
132. Philippon MJ. Outcomes following hip arthroscopy for femoroacetabular impingement with associated chondrolabral dysfunction MINIMUM TWO-YEAR FOLLOW-UP. *J Bone Jt Surg [Br]*. 2009;91(1):16–23.
133. Bedi A, Kelly BT, Khanduja V. Arthroscopic hip preservation surgery: Current concepts and perspective. *J Bone Jt Surg - Ser B*. 2013;95(1):10–9.
134. Tönnis D, Heinecke A. Current concepts review-acetabular and femoral anteversion: relationship with osteoarthritis of the hip. *J Bone Jt Surg - Ser A*. 1999;81(12):1747–70.
135. Narvani AA, Tsiridis E, Kendall S, Chaudhuri R, Thomas P. A preliminary report on prevalence of acetabular labrum tears in sports patients with groin pain. *Knee*

Surgery, Sport Traumatol Arthrosc. 2003;11:403–8.

136. *Right hip bone. External surface*, n.d. photograph, retrieved from https://en.wikipedia.org/wiki/Joint_capsule#/media/File:Gray235.png
137. *Front Section Through Right Hip Bone*, n.d photograph, retrieved from https://commons.wikimedia.org/wiki/File:916_Hip_Joint.jpg
138. *Femoral Acetabular Impingement*, n.d photograph, retrieved from https://commons.wikimedia.org/wiki/File:Femoral_acetabular_impingement_FAI_de.svg

SECTION II - Manuscript

Validity of Retrospective Patient Reported Assessment of Pre-Surgical Hip Pain and Disability Following Hip Arthroscopy.

Note to reader:

Manuscript presented in this section is drafted according to guidelines for submission for the journal *Arthroscopy: The Journal of Arthroscopic and Related Surgery* ([https://www.arthroscopyjournal.org/article/S0749-8063\(06\)01331-4/fulltext](https://www.arthroscopyjournal.org/article/S0749-8063(06)01331-4/fulltext)).

The required word limit for Arthroscopy has been exceeded in order to allow a deeper discussion and evaluation of results for the purposes of this thesis. Instructions relating to authorship, blinding for review, copyright and conflicts of interest statements have been omitted.

Abstract

Purpose:

The purpose of this study was to assess the validity of pre-surgery PROMS completed retrospectively 6 and 24 months after hip arthroscopy compared with prospective completion. Validity was judged in relation to preoperative test-retest reliability.

Methods:

Three separate groups of hip arthroscopy patients were recruited. In two retrospective groups, patients who had undergone a hip arthroscopy and had completed preoperative PROM questionnaires were contacted 6 or 24 months after surgery and asked to recall their preoperative state. In a third test-retest group, patients completed prospective PROMs twice prior to surgery, 2–8 weeks apart. PROMs included Non-Arthritic Hip Score (NAH), Hip Disability and Osteoarthritis Outcome score (HOOS), International Hip Outcome Tool (IHOT-12), and Visual Analogue Scale (VAS) for Medical Care, Feeling and Pain.

Results:

Patient sample sizes were 52, 40 and 61 for the 6-month, 24-month and test-retest groups, respectively. Retrospective 6- and 24-month and test-retest PROMs demonstrated 'moderate' to 'very large' correlations (ICC=0.42–0.79), except for three retrospective VAS PROMs. PROMs for the 24-month group reduced, from prospective to retrospective, for NAH-Total (MD -4.10 [-7.93--0.27; (95%CI)]; p=0.03), HOOS-Symptoms (-8.21 [-13.44--2.97]; p=0.003), HOOS-Daily-Living (-7.53 [-12.71--2.36]; p=0.005), HOOS-Sports (-6.49 [-12.97--0.00]; p=0.05), and HOOS-Short-Form (-5.16 [-9.83--0.49]; p=0.03). Similar non-significant reductions were observed for 6-month retrospective PROMs, except for VAS-Feeling which increased retrospectively. Three test-retest PROMs also reduced significantly, with worse outcomes approaching surgery.

Conclusion:

Correlations for retrospective versus prospective PROM scores spanned a similar range to test-retest correlations. Therefore, PROM scores relating to pre-surgery condition, but completed 6 or 24 months after surgery, showed agreement with PROM scores completed before surgery which were consistent with the day-to-day variability expected before surgery. This consistency supports the use of retrospective PROMs completed up to 2 years after hip arthroscopy.

Level of Evidence: Level 3

Introduction

Patient reported outcome measures (PROMs) can be a very useful tool when evaluating a patient's perception of the outcome of a surgical procedure (1). Self-reported ratings of function, general health and quality of life can provide valuable insight to a patient's condition and help measure otherwise unquantifiable data (2).

To better determine the effectiveness of treatment, preoperative or baseline measures are compared to post-operative or follow-up measures. Collection of baseline data prior to surgery is sometimes costly, difficult, and time consuming (3). As it may involve coordination with preadmission staff or require additional pre-surgery visits from patients, the collection of PROMs prior to surgery may not be feasible. This can potentially lead to this time point being missed. Situations can also arise where pre-operative data collection is not possible. For example, following a traumatic injury the patient may not be able to complete questionnaires before surgery and there are often other clinical priorities, and as result a baseline PROM cannot be gathered.

If a patient can accurately recall their prior status after surgery, then it could be possible to supplement this missing data with retrospectively collected PROMs. The use of retrospective PROMs taken shortly after surgery, has been shown to be relatively valid compared to PROMs collected prospectively (before surgery) by several studies (4–7). In the most recent studies, Kwong et al. (4) and Lawson et al. (5) both recommend the use of retrospective PROMs in creating a baseline measure.

There is equivocal information about the validity of long-term retrospective PROMs taken after an extensive period following surgery. Two particular studies show that patients undergoing a total hip arthroplasty (hip replacement) were able to accurately recall their baseline health status 6 weeks and 3 months after surgery, when compared with responses provided by them before surgery (3,8). Another study by Murphy et al., reported contrasting findings and found retrospective PROMs, taken at a longer time point one year after a total hip arthroplasty, were inaccurate compared to prospective PROMs (9).

To date, studies investigating retrospective collection of PROMs prior to orthopaedic surgery have been mostly completed in populations aged 55 to 72 years (3,8–10). There is no information about the validity of retrospective PROMs collection following a hip arthroscopy, for which patients are typically significantly younger, more active, undergo a less invasive surgery and have shorter recovery times (11,12). The purpose of this study is to investigate the validity of retrospective PROMs 6 and 24 months after hip arthroscopy. Unique differences in the design of this study include the recruitment of a younger, more active

patient cohort than those involved in previous retrospective studies, and gathering recalled PROMs at two medium- and long-term time points. Consequently, this study will provide a unique insight into the validity of retrospective PROMs after arthroscopic hip surgery.

Methods

Overall Design

To investigate the validity of retrospective PROMs, pain and disability measures were collected from two separate groups patients 6 and 24 months following hip arthroscopy. Retrospective PROMs scores were compared with pre-surgery questionnaires completed prospectively at the time these questionnaires related to. Correlation coefficients of the retrospective and corresponding prospective PROMs scores from the two time periods were compared to determine whether the intervening duration affected recall validity.

In addition, the third group of patients completed the same PROMs questionnaires twice, between 2 to 8 weeks apart, before having surgery. Test-retest reliability coefficients provided a benchmark by which the coefficients obtained from retrospective versus prospective assessment could be compared. The overall design is summarised in Figure 1.

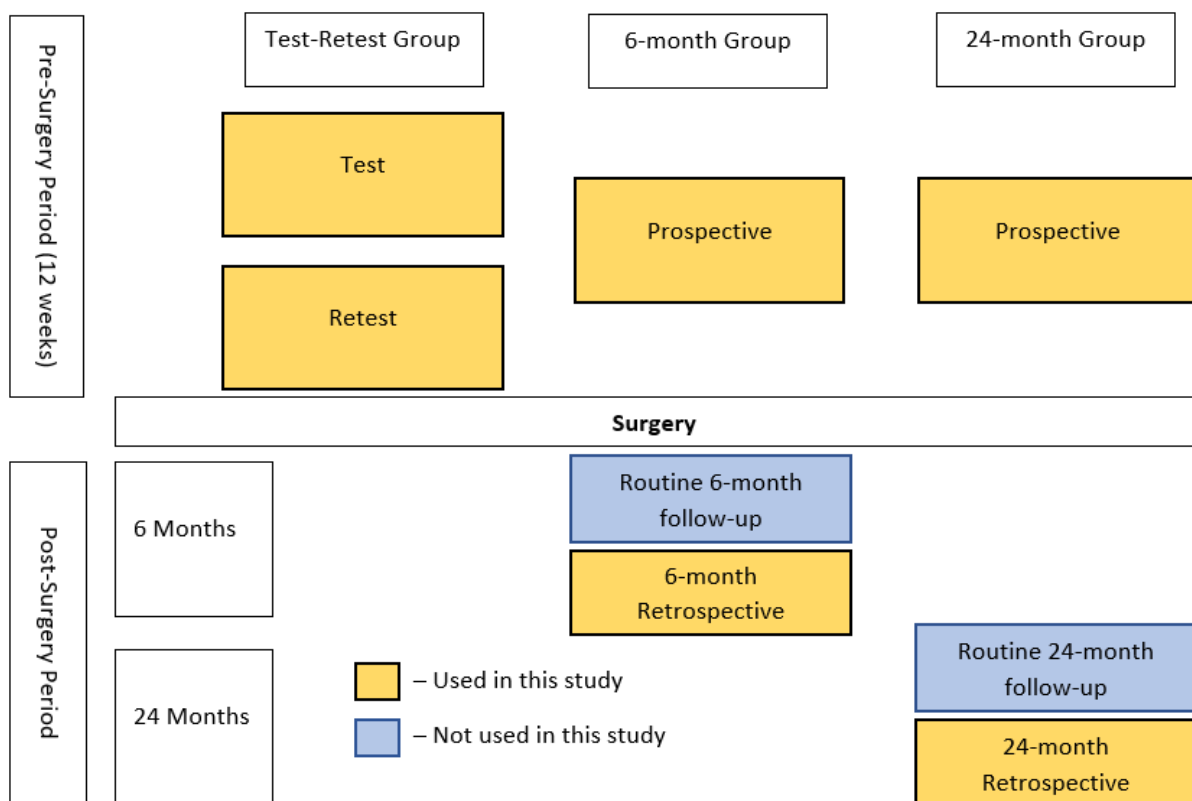


Figure 1: Overall design of the study showing time points of PROM questionnaire completions.

Inclusion and Exclusion Criteria:

Patients included in the study had received, or for test-retest were about to receive, primary arthroscopic hip surgery undertaken by one of two surgeons working for Orthosports North Harbour Ltd., Auckland, New Zealand. Participants were required to be at least 18 years of age and have completed pre-operative questionnaires within the 12 weeks prior to their operation. Patients were recruited either for the test-retest, 6-month retrospective or 24-month retrospective groups depending on the date of their surgery at recruitment.

As part of the surgeons' routine practice and according to procedures approved by the New Zealand Health and Disability Ethics Committee (17/NTA/269), Orthosports' patients complete 6-month and 24-month follow-up questionnaires. To avoid any confusion over the evaluation period to which the questionnaires related, patients were only recruited for the retrospective groups after their corresponding follow-up questionnaires were completed.

Patients in the retrospective groups were excluded if they had undergone previous hip, buttock or groin surgery on either side, or were undergoing simultaneous bilateral surgery or staged bilateral surgery within the time frame of the retrospective delay period (6 or 24 months). Patients who required revision or reoperative surgery within the time frame of the retrospective delay period (6 or 24 months) were also excluded from retrospective groups.

Data Collection

Between August 2018 and December 2020, eligible patients were contacted by email. They each were sent a link to information about the study and those willing to take part provided informed consent using an online platform, in a process approved by the New Zealand Health and Disability Ethics Committee (18/NTA/122). Participants then completed online questionnaires accessed via an emailed link. Information from patient records was kept in strict confidence by researchers and all data transferred out of the secure surgery database were deidentified.

Patients recruited for the retrospective assessment had surgery between 5 – 7 months earlier for the 6-month group, and between 22 – 26 months earlier for the 24-month group. If patients did not respond within 7 – 14 days, a follow-up email was sent.

For the prospective test-retest reliability study, eligible patients were contacted approximately 2 weeks after completing routine pre-operative questionnaires, which were completed between 2 and 12 weeks prior to their arthroscopy. They were asked if they were willing to complete questionnaires again and if willing provided online consent.

PROMs were emailed to patients via Socrates Orthopaedic Outcomes Software [Socrates Ortho, Pymble, NSW, Australia], and retained by this software for use in Orthosports surgical outcomes database.

Outcome Measures

Questionnaires included Non-Arthritic Hip Score (NAH), Hip Disability and Osteoarthritis Outcome score (HOOS), International Hip Outcome Tool 12-Item (IHOT-12), Visual Analogue Scale (VAS) for feeling and pain.

Non-Arthritic Hip Score (NAH)

NAH assesses functional pain, the degree of experienced hip discomfort or difficulty associated with listed specific activities. NAH generates a total score based on pain, symptoms, function experienced in the previous 48 hours and during activities in the previous month (13).

Hip Disability and Osteoarthritis Outcome score (HOOS)

HOOS generates a subscale scores based of the patient's symptoms, pain, daily living, sports, and quality of life experienced in the previous week (14).

International Hip Outcome Tool 12-Item (IHOT-12)

The iHOT-12 uses 12 questions taken from the IHOT-33 questionnaire. The questionnaire is very similar to the 33-item patient-reported outcome tool and is designed to measure the quality of life and the impact of disease in young and healthy patients, and loses little information from the complete tool. (15).

Visual Analogue Scale (VAS) for Feeling, Pain and Satisfaction with Medical Care

VAS is a simple outcome measure useful for gauging a subjective experience, (16) comprising a line with two verbal descriptors at either end, each denoting an extreme symptom (17). In this study three scales were scored of 100 using a cursor-controlled sliding scale. For satisfaction with Medical Care and normal Feeling, 100 indicated complete satisfaction and normal feeling, respectively. For Pain intensity, the scale was reversed with 'no pain' (score of 0) and 'worst pain imaginable' (score of 100) (18).

Data Analysis

Data were analysed using IBM SPSS Statistics, Version 26 [IBM Corporation, Armonk, NY, USA]. To assess the concordance between retrospective (remembered) and prospective

(actual) and between test-retest PROM scores, intraclass correlation, single-measure, two-way random effects model (2,1) was applied. Intraclass correlations coefficients (ICC) and mean differences with 95% confidence intervals (CI) were calculated. ICCs were interpreted using Hopkins' descriptors from Complete Scale of effect magnitudes (19). Systematic differences between retrospective minus prospective and retest minus test scores were examined using paired t-test. Concordance and difference statistics for 6- and 24-month groups were compared and assessed according to their level of correspondence with test-retest statistics.

Bland-Altman plots of differences between scores versus average score magnitudes were visually evaluated for heteroscedasticity. Pearson's correlation analyses were applied to determine whether any linear relationship existed between the observed differences between scores and average score magnitudes.

In order to determine whether follow-up time affected the validity of retrospective pre-surgical patient-reported measures, ICCs and differences between questionnaire administrations were compared in the two retrospective groups.

Results

Participant Characteristics

From a total of 437 potentially eligible patients 370 were approached. Of these, 157 did not respond to the invitation to take part, and the remaining 213 patients agreed to participate in the study. Upon inspection of the completed questionnaires for the retrospective groups it was realised that some participants had not followed instructions, completing the retrospective PROM questionnaires according to their current status. If retrospective PROM scores were very similar to the follow-up questionnaires that patients had recently completed with respect to their current status, then these were judged invalid. An exception to this was 3 patients for whom there was little change in scores between pre-surgery and follow-up time points, with retrospective scores being similar to both, and these were judged valid. In the 6-month retrospective group from 86 patients who consented 34 were invalid (52 valid) and for the 24-month group 26 from 66 patients completed invalid questionnaires (40 valid). Including 61 patients in the test-retest group, a final cohort of 153 participants were included in analysis (Figure 1).

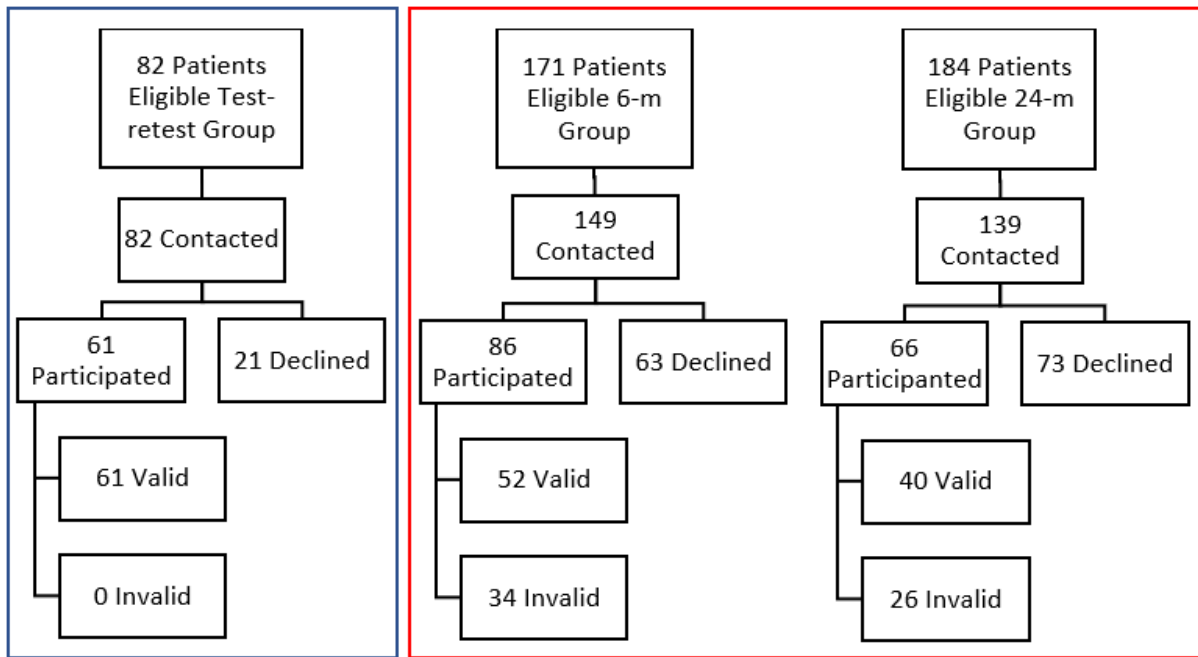


Figure 2. Participant flow chart for both the test re-test and validity components of the study

Durations between completions of the PROM questionnaires ranged from 12 to 56 days for the test-retest group, 25 to 47 weeks for the 6-month group, and 101 to 125 weeks for the 42-month group. Participants ranged in age at surgery from 15.8 to 63.4 years (mean \pm SD: 37.6 ± 11.1 years). The majority were female (109 of 153; 71%) and active, as 49.7% listed their sporting status as 'well-trained.' Participants' body mass index (BMI) ranged from 17.0 to 42.9 (25.2 ± 4.5). Patient characteristics by group are shown in Table 1.

Table 1. Patient characteristics for each of the three study groups.

	Test-Retest		6-month Retrospective		24-month Retrospective	
	mean or frequency	SD or percent	mean or frequency	SD or percent	mean or frequency	SD or percent
Number of Participants	61		52		40	
Duration between PROMs completion (weeks)	3.1	± 1.2	30.9	± 4.7	108.0	± 5.1
Age at surgery (years)	36.30	± 11	39.95	± 10.5	36.42	± 11.9
Gender (Male/Female) (% female)	21/40	66.7%	12/40	76.9%	11/29	72.5%
Height (cm)	171.7	± 9.2	170	± 8.5	171.8	± 9.7
Weight (kg)	74.45	± 16.8	73.6	± 16.6	74.2	± 13.6
BMI (kg/m²)	25.2	± 5.2	25.4	± 4.8	25	± 3.3

	Test-Retest		6-month Retrospective		24-month Retrospective	
	mean or frequency	SD or percent	mean or frequency	SD or percent	mean or frequency	SD or percent
Side (Left%)	22	36.1%	19	36.5%	19	47.5%
Sport Status No sports	6	9.8%	4	7.7%	1	2.5%
Sport Status Sport Sometimes	15	24.6%	9	17.3%	6	15.0%
Sport Status Well-trained	30	49.2%	24	46.2%	22	55.0%
Sport Status High Level Competitive	10	16.4%	14	26.9%	9	22.5%
Sport Status Professional	0	0.0%	1	1.9%	0	0.0%
Sport Status Not Specified	0	0.0%	0	0.0%	2	5.0%

PROM Scatterplots

Scatterplots for the PROMs HOOS Pain and NAH Total, shown in *Figure 3*, demonstrated a strong positive linear association between the retrospective and prospective measures in the 6- and 24-month groups, and between test and retest prospective PROMs in the test-retest group. HOOS-Pain and NAH Total, for all groups, showed 'large' to 'very large' correlations between pairs of questionnaires with scatter plots showing 1 to 3 outliers (*Figure 3*; see Appendix A for all scatter plots).

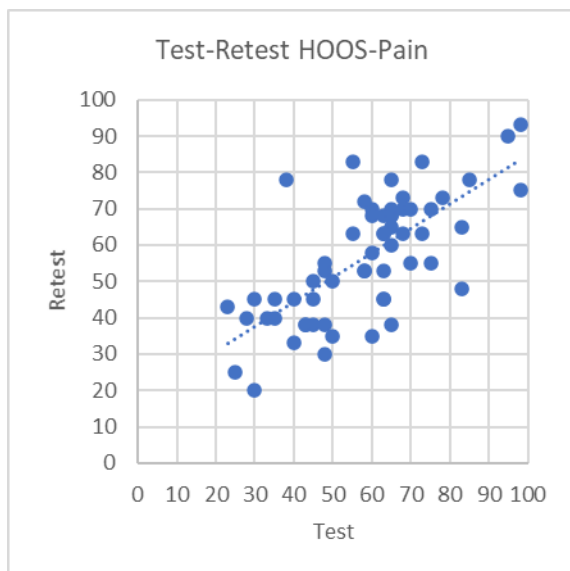


Figure 3a

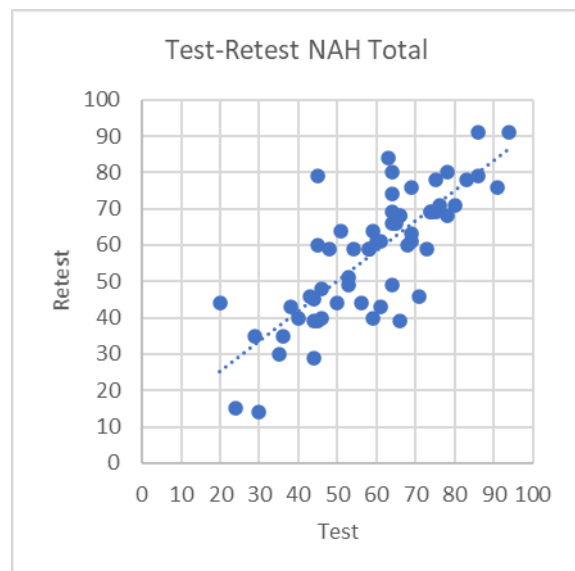


Figure 3b

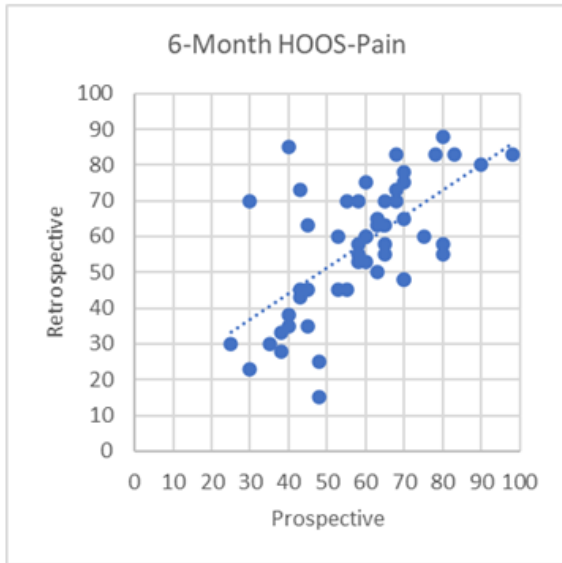


Figure-3c

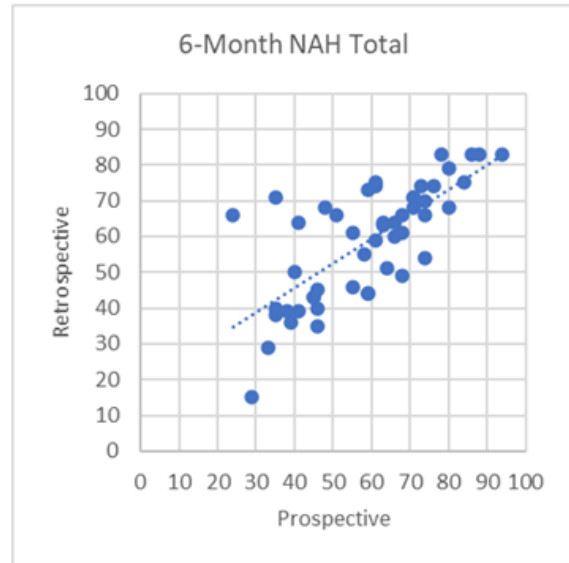


Figure-3d

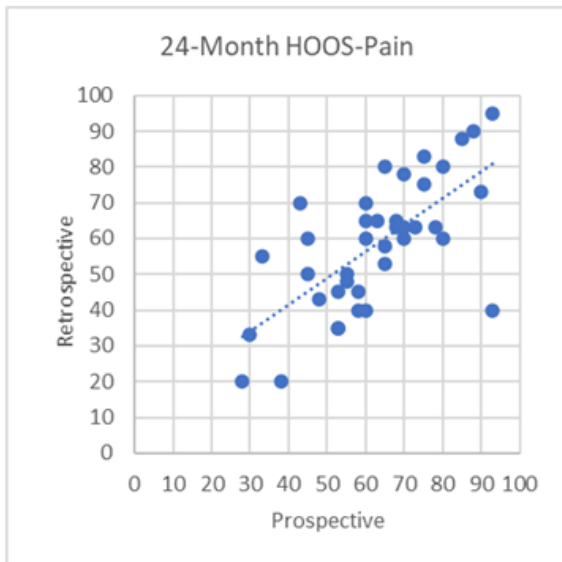


Figure-3e

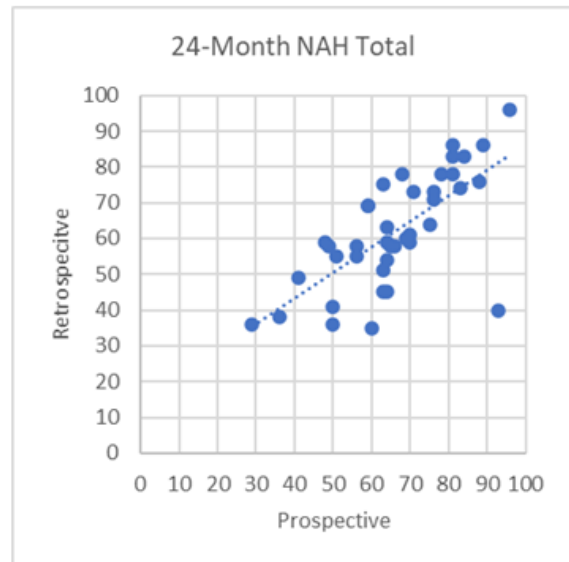


Figure-3f

Figures 3a to 3f. Scatter plots illustrating the associations between scores from retrospective and prospective completions of patient reported outcome measures (PROMs): a, Test-Retest HOOS-Pain; b, Test-retest NAH Total; c, 6-month HOOS-Pain; d, 6-month NAH Total; e, 24-month HOOS-Pain; f, NAH Total. PROMs are Hip Osteoarthritis Outcome Score – Pain Subscale (HOOS-Pain) and Non-Arthritic Hip (NAH) Total.

Intraclass Correlations (Table 2)

In the retrospective 6-month group, PROMs obtained consistency ICCs between prospective and retrospective ranging from 0.42 to 0.72 (Table 2). All ICCs were ‘moderate’ to ‘very large’, excluding VAS Medical Care, Feeling and Pain PROMs, which had an ICC of 0.29, 0.21 and 0.32, respectively.

In the 24-month retrospective group, consistency ICCs ranged from 'moderate' to 'very large'. The ICCs were similar to the 6-month retrospective group, with the same PROM usually obtaining a slightly smaller correlation. The 24-month group had a large variation in the ICCs for the three VAS PROMs, ranging from 'trivial to 'large' effects (Medical Care: ICC = 0.55; Feeling: ICC = 0.09; Pain: ICC = 0.33).

ICCs for test-retest PROMs ranged from 0.44 to 0.79, 'moderate' to 'very large' effect size. HOOS Sports was the only PROM where test-retest did not obtain the largest ICC (0.47), compared to the corresponding ICC from 6-month (0.59) and 24-month (0.55) groups.

The PROM NAH Total had the largest ICCs for each group, with the test-retest obtaining ICC of 0.79, 6-month retrospective group ICC of 0.72, and 24-month retrospective group ICC of 0.70.

Table 2: Intraclass correlation coefficients (ICC) for paired PROM completions for each study group

	Test-Retest Group				
	ICC-consistency		ICC-absolute		Hopkins Descriptor (ICC consistency)
		95%CI		95%CI	
NAH Total	0.79	0.67–0.87	0.79	0.67–0.87	<i>Very large</i>
HOOS Symptoms	0.69	0.53–0.80	0.68	0.52–0.80	<i>Large</i>
HOOS Pain	0.71	0.55–0.81	0.71	0.55–0.81	<i>Very large</i>
HOOS Daily Living	0.68	0.52–0.80	0.67	0.50–0.79	<i>Large</i>
HOOS Sports	0.47	0.24–0.64	0.46	0.24–0.64	<i>Moderate</i>
HOOS Quality of Life	0.67	0.50–0.79	0.67	0.50–0.79	<i>Large</i>
HOOS Short-Form	0.56	0.35–0.72	0.54	0.33–0.70	<i>Large</i>
IHOT-12 Total	0.68	0.52–0.80	0.66	0.48–0.79	<i>Large</i>
VAS Medical Care	0.63	0.45–0.76	0.64	0.45–0.77	<i>Large</i>
VAS Feeling Scale	0.44	0.21–0.62	0.44	0.21–0.62	<i>Moderate</i>
VAS Pain Scale	0.61	0.42–0.75	0.61	0.42–0.74	<i>Large</i>

	6-month Group				
	ICC-consistency		ICC-absolute		Hopkins Descriptor (ICC consistency)
	95%CI		95%CI		
NAH Total	0.72	0.56 – 0.83	0.73	0.57 – 0.84	<i>Very large</i>
HOOS Symptoms	0.43	0.18 – 0.63	0.43	0.18 – 0.63	<i>Moderate</i>
HOOS Pain	0.63	0.43 – 0.77	0.63	0.44 – 0.77	<i>Large</i>
HOOS Daily Living	0.61	0.41 – 0.76	0.62	0.42 – 0.76	<i>Large</i>
HOOS Sports	0.59	0.37 – 0.74	0.59	0.38 – 0.74	<i>Large</i>
HOOS Quality of Life	0.42	0.17 – 0.62	0.41	0.17 – 0.61	<i>Moderate</i>
HOOS Short-Form	0.63	0.42 – 0.77	0.63	0.43 – 0.77	<i>Large</i>
IHOT-12 Total	0.59	0.38 – 0.75	0.60	0.39 – 0.75	<i>Large</i>
VAS Medical Care	0.29	0.02 – 0.53	0.29	0.02 – 0.53	<i>Small</i>
VAS Feeling Scale	0.21	-0.07 – 0.46	0.20	-0.07 – 0.45	<i>Small</i>
VAS Pain Scale	0.32	0.04 – 0.55	0.32	0.05 – 0.55	<i>Moderate</i>

	24-month Group				
	ICC-consistency		ICC-absolute		Hopkins Descriptor (ICC consistency)
	95%CI		95%CI		
NAH Total	0.70	0.50 – 0.83	0.68	0.47 – 0.82	<i>Very large</i>
HOOS Symptoms	0.49	0.21 – 0.70	0.44	0.14 – 0.66	<i>Moderate</i>
HOOS Pain	0.68	0.46 – 0.82	0.66	0.44 – 0.81	<i>Large</i>
HOOS Daily Living	0.55	0.29 – 0.73	0.50	0.21 – 0.71	<i>Large</i>
HOOS Sports	0.55	0.29 – 0.74	0.53	0.27 – 0.72	<i>Large</i>
HOOS Quality of Life	0.49	0.21 – 0.69	0.49	0.21 – 0.70	<i>Moderate</i>
HOOS Short-Form	0.47	0.18 – 0.69	0.45	0.16 – 0.67	<i>Moderate</i>
IHOT-12 Total	0.58	0.32 – 0.75	0.56	0.31 – 0.74	<i>Large</i>
VAS Medical Care	0.55	0.28 – 0.74	0.54	0.27 – 0.73	<i>Large</i>
VAS Feeling Scale	0.09	-0.24 – 0.39	0.09	-0.25 – 0.40	<i>Trivial</i>
VAS Pain Scale	0.33	0.01 – 0.59	0.32	0.02 – 0.58	<i>Moderate</i>

Non-Arthritic Hip Score (NAH) Total; Hip Osteoarthritis Outcome Score (HOOS); International Hip Outcome Tool 12-Item (IHOT-12); Visual Analogue Scale (VAS).

Bland-Altman Plots

Bland-Altman charts plot difference in the paired PROM scores against their average. Four Bland-Altman plots in the two retrospective groups displayed a heteroscedastic pattern. For 6-month retrospective VAS Feeling, 24-month IHOT-12 Total, 24-month VAS Feeling, and 24-month VAS Pain patients typically overestimated their preoperative state at the post-surgery time point, this overestimation increasing as average score increased (Figure 4a – d.). For 24-month IHOT-12 Total and 24-month VAS Pain patients tended to underestimate their preoperative state when average scores were low (Figure 4b and d). An increasing spread, or difference between prospective and retrospective scores, with higher average score was also noted for 6-month and 24-month VAS Feeling (Figures 4 a and d).

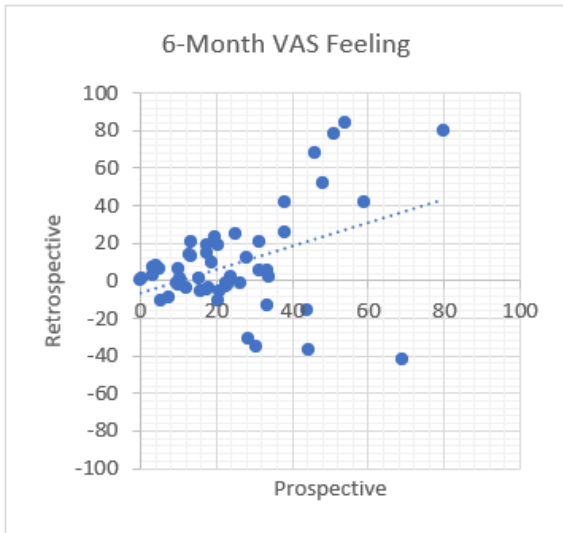


Figure 4a

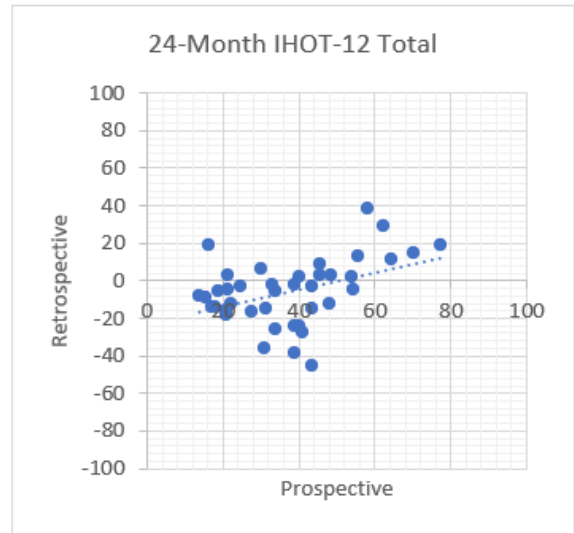


Figure 4b

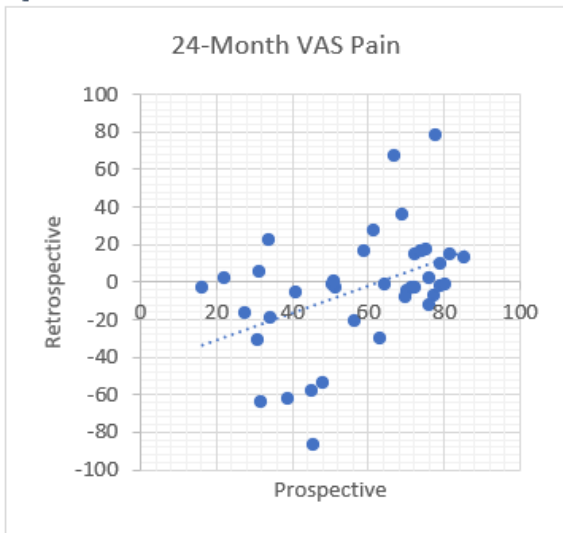


Figure 4c

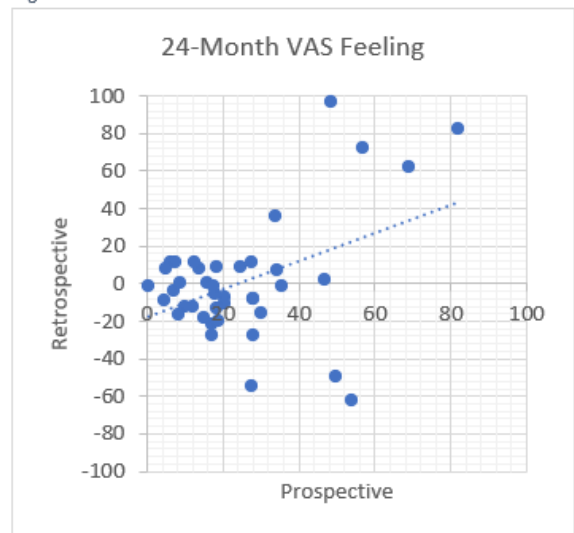


Figure 4d

Figure 4a-d. Bland-Altman plots showing difference versus average of the scores for paired PROM questionnaire completions: a, 6-month visual analogue scale (VAS) feeling; b, 24-month International Hip Outcome Tool 12-Item (IHOT-12) total; c, 24-month VAS pain; d, 24-month VAS feeling.

Changes in PROM Scores (Table 3)

For most of the PROMs in this study, the average change from prospective to retrospective questionnaire completion in the retrospective groups, and from test to retest, was negative. This means that in most cases, participants underestimated their preoperative score retrospectively or, in the case of the test-retest group, reported that their condition worsened before surgery. However, only a few PROMs showed changes that were statistically significant.

In the 24-month retrospective group, differences for NAH Total (-4.10 [-7.93--0.27; (95%CI)]; p=0.03), HOOS Symptoms (-8.21 [-13.44--2.97]; p=0.003), HOOS Daily Living (-7.53 [-12.71--2.36]; p=0.005), HOOS Sports (-6.49 [-12.97--0.00]; p=0.05), and HOOS Short-Form (-5.16 [-9.83--0.49]; p=0.03) attained statistical significance (p<0.05). However, the degree of retrospective underestimation of pre-operative state compared to prospective, ranging from -4.10 to -8.21, was within previously reported thresholds of minimal clinically important differences (MCIDs), the smallest change in an outcome that a patient can recognise or appreciate (20). Previously reported MCIDs for the PROMs used in the study are as follows: 9 for all of the HOOS PROMs (21); 13 for IHOT-12 Total (22); 8.8 for NAH-Total (23); and 22.7 for VAS (24).

In some of the PROMs in the 6-month retrospective group, similar non-significant reductions were also observed, except for VAS Feeling which increased retrospectively attaining borderline statistical significance (7.47 [0.13--14.81]; p = 0.046).

In the test-retest group, the only PROMs that obtained a statistically significant change were HOOS Daily Living (-3.87 [-7.73--0.00]; p=0.05), HOOS Short-Form (-4.49 [-8.41--0.53]; p=0.03), and IHOT-12 Total (-4.91 [-8.7--1.23]; p=0.01). These changes were also below the MCID.

Table 3. Mean and SD scores for each time-point

	Test-retest Patient Group							
	n	T1		T2		Difference		p
		mean	SD	mean	SD	mean	95%CI	
NAH Total	60	59.2	16.7	57.6	17.5	-1.6	-4.4 – 1.3	0.278
HOOS Symptoms	60	60.1	16.2	56.8	17.3	-3.3	-6.7 – 0.2	0.061
HOOS Pain	60	58.1	17.5	56.5	16.7	-1.7	-5.1 – 1.7	0.329
HOOS Daily Living	60	69.1	18.1	65.2	19.2	-3.9	-7.7 – 0.0	0.050
HOOS Sports	58	47.9	26.0	43.5	23.1	-4.5	-11.1 – 2.2	0.188
HOOS Quality of Life	60	31.3	17.6	32.9	17.7	1.7	-2.1 – 5.4	0.380
HOOS Short-Form	56	69.6	16.0	65.1	15.5	-4.5	-8.5 – -0.5	0.027
IHOT-12 Total	61	38.5	20.0	33.5	16.5	-5.0	-8.7 – -1.2	0.010
VAS Medical Care	59	80.7	19.7	82.3	20.0	1.6	-2.8 – 6.0	0.469
VAS Feeling Scale	61	23.9	20.3	25.1	21.3	1.2	-4.5 – 6.8	0.681
VAS Pain Scale	61	63.3	21.3	60.1	22.8	-3.2	-8.2 – 1.8	0.203

6-month Patient Group								
	n	T1		T2		Difference		
		mean	SD	mean	SD	mean	95%CI	p
NAH Total	50	58.9	17.0	58.8	16.1	-0.1	-3.6 – 3.4	0.936
HOOS Symptoms	52	57.2	16.1	53.9	18.8	-3.4	-8.5 – 1.8	0.198
HOOS Pain	52	58.3	16.0	57.2	18.2	-1.1	5.2 – 3.0	0.594
HOOS Daily Living	52	67.5	18.8	67.0	20.3	-0.5	5.2 – 4.3	0.834
HOOS Sports	51	44.1	20.9	43.3	20.5	-0.9	6.2 – 4.4	0.744
HOOS Quality of Life	52	30.1	16.3	34.4	16.2	4.3	-0.6 – 9.1	0.085
HOOS Short-Form	50	66.3	15.7	65.5	17.0	-0.8	-4.8 – 3.2	0.701
IHOT-12 Total	51	34.0	16.9	32.7	14.7	-1.3	-5.4 – 2.7	0.507
VAS Medical Care	49	77.6	25.1	81.3	18.8	3.7	-3.9 – 11.3	0.329
VAS Feeling Scale	49	20.5	17.2	27.9	23.1	7.5	0.1 – 14.8	0.046
VAS Pain Scale	49	61.6	20.1	62.7	23.1	1.0	-6.2 – 8.3	0.774

24-month Patient Group								
	n	T1		T2		Difference		
		mean	SD	mean	SD	mean	95%CI	p
NAH Total	40	66.2	15.2	62.1	15.7	-4.1	-7.9 – -0.3	0.037
HOOS Symptoms	39	63.3	13.1	55.1	18.5	-8.2	-13.4 – -3.0	0.003
HOOS Pain	39	62.8	16.8	58.4	18.3	-4.4	-8.9 – 0.2	0.060
HOOS Daily Living	39	74.3	15.5	66.8	18.0	-7.5	-12.7 – -2.4	0.005
HOOS Sports	39	51.6	21.7	45.1	20.6	-6.5	-13.0 – 0.0	0.050
HOOS Quality of Life	39	36.7	17.9	36.9	18.8	0.2	-5.9 – 6.2	0.952
HOOS Short-Form	38	73.4	12.7	68.2	14.9	-5.2	-9.8 – -0.5	0.031
IHOT-12 Total	39	40.1	14.7	35.1	3.5	-5.0	-10.5 – 0.6	0.076
VAS Medical Care	36	78.4	21.8	72.1	24.3	-6.4	-13.8 – 1.0	0.090
VAS Feeling Scale	37	22.4	16.8	23.0	25.6	0.6	-9.1 – 10.4	0.898
VAS Pain Scale	37	61.8	18.6	54.9	27.6	-6.9	-16.0 – 2.2	0.132

Non-Arthritic Hip Score (NAH); Hip Osteoarthritis Outcome Score (HOOS); International Hip Outcome Tool 12-Item (IHOT-12); Visual Analogue Scale (VAS); T1 = first time point (test for test-retest, prospective for retrospective groups); T2 = second time point (retest for test-retest, retrospective for 6-month and 24-month groups).

Discussion

The aim of this study was to evaluate the validity of retrospective PROMs 6-months and 24-months after a hip arthroscopy. To our knowledge, this was the first study that assessed the reliability and validity of PROMs before surgery and retrospectively, involving hip arthroscopy.

As part of the investigation into retrospective PROMs, the reliability of prospective test-retest PROMs was also considered. The reason for this was to have the validity of retrospective PROMs determined partly by reference to the reliability of the test-retest PROMs. Therefore,

data obtained from the test-retest PROMs served as a benchmark in the assessment of the validity of the retrospective PROMs.

In summary, in both the 6-month and 24-month retrospective groups, participants for all PROMs were able to recall their pre-operative status with a moderate to very high degree of accuracy, based on degrees of correlation. These results exclude the VAS Feeling, Pain and Patient Satisfaction PROMs, for the reason that some of them obtained a trivial and small correlation, possibly due to the nature of the questionnaire. The results for the test-retest PROMs showed that patients were able to reproduce prospective PROMs with a moderate to very high degree of reliability, which was similar but slightly higher than in the case of most of the retrospective PROMs.

A major strength of this study was the examination of test-retest reliability in conjunction with retrospective validity. This provided useful insight into the applicability of retrospective PROMs in clinical studies and practice, due to a lack of significant change in the intraclass correlations between each group. This small degree of change enhances the validity of the use of retrospective PROMs, in clinical practice, as there was no major difference in reliability between the different groups.

Comparison to Other Studies of Retrospective PROMs

Several previous studies have been conducted on retrospective and prospective PROMs in respect of other different surgeries. The most relevant of these investigate hip arthroplasties. Marsh et al., for some of the retrospective PROMs assessed, reported a very high accuracy 3-months after a hip arthroplasty (ICC = 0.86 – 0.89) (3). Howell et al., also accuracy high reliability for retrospective PROMs at 6-weeks (ICC = 0.68 – 0.88) and 3-months (ICC = 0.74 – 0.92) after hip arthroplasty (8).

By way of comparison, the results of this study were as follows: in the 6-month group the ICC was 0.42 to 0.73, and in the 24-month group the ICC was 0.48 to 0.70, both of which excludes the VAS PROMs. Whilst the accuracy in correlation between prospective and retrospective PROMs in this study was not as high as the studies by Marsh et al. (3) and Howell et al. (8), it was fairly consistent with the test-retest group (ICC = 0.47 – 0.79).

One possible reason for the differences in correlations between the prospective and retrospective PROMs in this study, and those correlations found in the studies of Marsh et al. (3) and Howell et al. (8), may be due to the much longer 6-month and 24-month time-periods that applied in this study. Such an explanation would be consistent with the fact that studies which have looked into longer time-periods have generally not had favourable results

concerning retrospective PROMs (9,25–27). However, the results of these particular studies, whilst perhaps affirming a possible deficiency in longer time periods, were nevertheless quite different from the overall results of this study, which tend to support the validity of retrospective PROMs after hip arthroscopy. Although retrospective PROMs at the longest time-point of 24-months were indeed not as reliable as prospective PROMs, as was apparent from a comparison with the test-retest group, they were still moderately to strongly consistent with their pre-surgical scores, once again excluding the VAS PROMs.

The difference between the results of this study and those studies (9,25–27) looking at similar to longer time periods may be due to the fact that the latter concerned older patients, whereas this study dealt with younger patients who had undergone hip arthroscopic surgery. The average age of participants in these studies ranged from 36.63 to 39.95 years old. Even though a few studies have denied the influence of age on a patient's recall (8,25,27), these studies have in fact focused solely on older populations. However, a recent study by Gotlin et al. contrasted these findings and observed, in a cohort of shoulder and elbow arthroscopy patients aged 57 ± 10 years, that younger age was associated with more accuracy in retrospective PROMs (28). This particular study looked into long time frames and, in that context, concluded that a younger cohort could potentially increase the reliability of the retrospective PROMs. The findings from this current study tend to support the conclusion reached by Gotlin et al., namely that younger patients may better recall their preoperative status at longer time-periods (28).

A further feature of this study was that the sample size of cohorts in all of the groups was larger than some comparable studies. The test-retest group had 61; the 6-month group had 52; and the 24-month group had 40 participants. This means that the total participants numbered 153. In comparison, and by way of example, Widnall et al. (10), and Reynolds and Thirkannad (29), had 36 and 38 patients respectively. Neither of these studies commented on the relatively small size of their patient cohort and the impact it could have had on the reliability of their findings.

Variability between Test-retest and Retrospective PROMs

In this study, test-retest results for PROMs showed ICCs ranging from 0.44 to 0.79. Hopkins descriptor would classify these results as having a 'moderate' to 'very large' reliability. The range of these correlations indicates that there is some variability in a patient's response to PROM questionnaire items prior to surgery. This could be due to a patient's condition fluctuating from a day-to-day basis prior to surgery. Variability may also be affected by changes in patient mood or judgment at the time.

Retrospective PROMs can somewhat circumvent this variability and limitation in prospective PROMs, by asking patients to recall their condition over the broader pre-surgery period.

Another influence on retrospective PROMs is that, unlike prospective PROMs, they rely on a patient's memory and understanding of their condition and events that transpired, both before and after the operation. Unfortunately, this makes retrospective PROMs susceptible to recall bias and response-drift. However, putting these possible faults to one side, the very nature of retrospective PROMs can give a patient the opportunity to encapsulate their understanding and their memory of the time before surgery in its entirety.

This fundamental difference between prospective PROMs, which look at the present moment, and retrospective PROMs, which consider the period before surgery, highlights the rationale as to why both types of PROMs are of benefit to the practitioner. One provides a simple metric about the patient's pre-surgical condition, and the other a measure which is based on the sum of a patient's memory.

In addition to potentially being subject to recall bias or response shift, it is also possible that a patient's memory of their condition prior to surgery may be influenced by their current health status. In one particular study, it was theorised that if patients could not recall their preoperative state, then they resort to deductive reasoning, which is based on their current status, to deduce their preoperative state (3).

In assessing the validity of retrospective PROMs in this study, the intention has been to elicit an accurate response from participants about a specific moment in time, which was when they filled out their prospective PROMs before surgery. The study is based upon the premise that prospective PROMs served to provide participants with a snapshot or 'anchor-point' to assist them in remembering their pre-surgical condition more accurately, and so avoid the need to engage in deductive reasoning based on their current health.

A further consideration, concerning the reliability of retrospective PROMs after hip arthroscopy, arises from studies that have proven there is a placebo effect involved in surgery, even in hip arthroscopy (30). This means that the mere involvement of surgery could potentially influence a patients' perceived benefit of surgery. This could exaggerate their understanding of their pre-surgical state or the effectiveness of treatment. This effect may contribute to recall bias, due to patients overestimating their previous condition, leading to worse scores in retrospective PROMs, a trend which was noted in this study.

Any such placebo effect should have been minor in this study since the nature of the PROMs that were used examined physical events, such as injuries or barriers affecting daily

quality of life. Therefore, these PROMs allowed patients to measure outcomes by reference to daily lifestyle markers. This arguably minimised any potential bias the placebo effect might have produced retrospectively and prospectively.

Retrospective PROM Findings

As noted, in the retrospective groups, both the 6-month and 24-month cohorts reported high and very high correlations in several retrospective PROMs, almost maintaining similar scores to the test-retest PROMs. The PROM NAH Total maintained a very large intraclass correlation across all groups. However, with longer timeframes the correlation across the groups from test-retest, 6-month, and 24-month cohorts, slowly decreased (ICCs: 0.79, 0.72 and 0.70 respectively). Most of the PROMs had a similar pattern across all groups, with reliability dropping as the time period increased.

However, this pattern differed in the PROM HOOS Sports, for which the correlation of the test-retest group (ICC = 0.47) was not as good as the 6-month and 24-month retrospective groups (ICC = 0.59; 0.55). This anomalous finding could possibly be due to random error and might be rectified if a larger patient pool is utilised in future studies. Nevertheless, the available data from this study suggests that patients may more reliably recall their pre-operative sporting status 6-months after surgery, in a manner which does not contain the same degree of variability apparent in the test-retest prospective assessment. The mean difference for the test-retest group was -4.45 and the 6-month group -0.86. However, it is considered that, because of the high p-value ($p = 0.19$; 0.74), this finding is not statistically significant enough to necessarily justify the use of retrospective PROMs in sporting surgeries instead of prospective PROMs.

A systematic change, potentially caused by recall bias, was noted in the 24-month group with retrospective PROMs. The statistical change was significant in several retrospective PROMs being NAH total (-4.10), HOOS symptoms (-8.21), HOOS Daily Living (-7.54) HOOS Sports (-6.49), HOOS Short Form (-5.16). The negative differences mean that patients recalling their preoperative rating underestimated their previous prospectively-completed PROM. This finding was relatively consistent with the study by Gotlin et al. (2020) (28), which noted an average change of -20.73 in retrospective PROMs one year after surgery. In comparison, the average change in the current was -4.10 to -8.12 in the scores that attained statistical significance. Other scores in the study showed non-significant trends, especially the HOOS Quality of Life and VAS Pain and Feeling Scales in the 24-month groups.

The average difference in the 24-month group for the HOOS Quality of Life and VAS Feeling Scale PROMs was smaller than the change found in the test-retest PROMs. This small

change could indicate that patients may maintain, after 24-months, a more accurate recollection of their quality of life and feeling of normality before surgery. However, as noted, the results were not statistically significant and, for that reason, no firm conclusion can be drawn.

As to other patterns found in the data, visual inspection of Bland-Altman plots revealed heteroscedasticity in a few retrospective PROMs. This means that there was a development of a noticeable heteroscedastic pattern in a patient's ability to recall their prior state which depended on their preoperative PROM score. This in turn affects the level of error and difference according to the severity of the preoperative PROM (31). The PROMs that had a heteroscedastic pattern in the study were the 6-month VAS Feeling score, 24-month iHOT-12 total score, and the 24-month VAS Feeling score.

Test-retest Group

As the test-retest PROMs demonstrate, PROMs are a good way to quantify otherwise unquantifiable data and provide a means of understanding heavily subjective outcomes and measures. They can provide insight into a patient's condition and an understanding of the progression of a patient's symptoms. However, as this study shows, prospective PROMs do not necessarily provide a complete picture on a patient's pre-operative status.

This is apparent from the fact that the test-retest group did not have consistently large reliability across all PROMs. Arguably, as the test-retest group dealt with two prospective PROMs taken at different time-points before surgery, it could perhaps have been expected that each PROM would have been consistent across all questionnaires and would have retained a very high correlation. However, only NAH total, HOOS pain and iHOT-12 total PROMs obtained this outcome, namely a very high correlation (ICC 0.67 – 0.79).

In the test-retest group, all the PROMs, except for the HOOS Quality of Life, Patient Satisfaction Medical Care and VAS Feeling Scale, had a negative change in their scores. This could potentially be explained by patients' worsening condition approaching surgery. However, only the HOOS Daily Living, HOOS Short form and iHOT-12 Total test-retest PROMs showed changes that were statistically significant. Due to these PROMs being statistically significant, we are therefore able to draw the conclusion that there was a deterioration in these patients' condition and symptoms as they approached surgery. This means that these particular preoperative PROMs are likely to exhibit worse scores for patients as time approaches a hip arthroscopic surgery.

This observation is consistent with the fact that, although preoperative PROMs attempt to increase the understanding of a patient's overall condition by gaining quantifiable data, they still, by their very nature, focus on a patient's current perception of their condition. This limitation means that prospective PROMs may not effectively gain all the information necessary to fully understand a patient's pre-surgical condition. This is partly due to the fact that symptoms can fluctuate on a daily basis and deteriorate rapidly approaching surgery, and so affect a patient's response to the PROMs (32).

Limitations

A limitation of this study was that there was a sizable exclusion rate amongst participants. This was primarily due to a number of participants apparently misunderstanding study instructions. To some extent, the confusion may have arisen from the use of Socrates, which was the program responsible for sending patients PROMs via email. Socrates did not mention that the questionnaires required patients to answer retrospectively by recalling their pre-operative condition and instructions in the questionnaire were unable to be adjusted to past tense.

Another possible reason for misunderstanding of instructions was that the retrospective PROMs, with the exception of the VAS PROM, were identical to the questionnaires that had been very recently sent to participants as part of routine postoperative follow-up. Consequently, a number of patients may simply have answered retrospective PROM questionnaires by replicating follow-up postoperative PROMs which they had recently completed. It appears that, in such cases, instructions to participants contained in accompanying emails may have been either overlooked or misunderstood.

When completed PROMs were virtually identical to a participant's recent follow-up PROMs, and were substantially different from their prospective PROMs, data from these participants were excluded from analysis as invalid. This exclusion potentially introduced bias into the study and may have resulted in an overestimation of the correlation between prospective and retrospective PROMs scores. However, even if it did exist, such bias is fairly unlikely to have greatly impacted the correlation. This is because the considerable difference patients typically had between their follow-up and prospective PROMs suggested that exclusion of the data was very likely to mean that the retrospective PROMs were not completed by remembering their pre-surgery condition as instructed.

Conclusion

In conclusion, this study supports the validity of using retrospective PROMs 6-months and 24-months after a hip arthroscopy, although not always in place of prospective PROMs.

Even though retrospective PROMs may in the end have been influenced by recall bias, the correlation of retrospective to prospective PROMs did not, in this study, greatly differ in comparison to the test-retest PROMs. This means that in cases where prospective data cannot be gathered for any reason retrospective PROMs undertaken 6-months or 24-months after a hip arthroscopy can be substituted. Not only may this reduce administrative burden, but it could also have the advantage of reducing the effect of day-to-day fluctuations in patient status prior to surgery.

In terms of future studies, it is suggested that the use of larger patient cohorts than those adopted in this study could assist in further consideration of the validity of retrospective PROMs in hip arthroscopy. In addition, changing the phrasing of PROMs to fit a retrospective narrative would assist in reducing participant confusion, exclusion rate and, as a result, improve and increase the sample size.

References

1. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, De Vet HCW. Outcome instruments: Rationale for their use. *J Bone Jt Surg.* 2009;91(3):41–9.
2. Jackowski D, Guyatt G. A Guide to Health Measurement. *Clin Orthop Relat Res.* 2003;413:80–9.
3. Marsh J, Bryant D, MacDonald SJ. Older Patients Can Accurately Recall Their Preoperative Health Status Six Weeks Following Total Hip Arthroplasty. *J Bone Jt Surgery-American Vol [Internet].* 2009;91(12):2827–37. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004623-200912000-00007>
4. Kwong E, Neuburger J, Black N. Agreement between retrospectively and contemporaneously collected patient-reported outcome measures (PROMs) in hip and knee replacement patients. *Qual Life Res.* 2018;27(7):1845-1854.
5. Lawson A, Tan AC, Naylor J, Harris IA. Is retrospective assessment of health-related quality of life valid? *BMC Musculoskelet Disord.* 2020;21(1):1–10.

6. Bryant D, Norman G, Stratford P, Marx RG, Walter SD, Guyatt G. Patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery. *J Clin Epidemiol*. 2006;59(9):984–93.
7. Kuittinen P, Aalto T, Heikkilä T, Leinonen V, Savolainen S, Sipola P, et al. Accuracy and reproducibility of a retrospective outcome assessment for lumbar spinal stenosis surgery. *BMC Musculoskelet Disord*. 2012;13(1):83.
8. Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS. A Comparison Between Patient Recall and Concurrent Measurement of Preoperative Quality of Life Outcome in Total Hip Arthroplasty. *J Arthroplasty* [Internet]. 2008 [cited 2017 May 4];23(6):843–9. Available from: <http://www.sciencedirect.com.libproxy.unitec.ac.nz/science/article/pii/S0883540307004627>
9. Murphy MT, Vardi R, Journeaux SF, Whitehouse SL. A patient’s recollection of pre-operative status is not accurate one year after arthroplasty of the hip or knee. *Bone Jt J*. 2015;97-B(8):1070–5.
10. Widnall J, Ralte P, Selvan D, Molloy A. An evaluation of retrospective outcome scores in elective foot and ankle surgery. *Foot* [Internet]. 2014 [cited 2017 Apr 18];24(1):28–30. Available from: <http://www.sciencedirect.com.libproxy.unitec.ac.nz/science/article/pii/S0958259214000078?np=y&npKey=5556adeb3d2d9d4bfae8ab1c57508794a85884bde2fd1b394f7299404872b9a0>
11. Grant S, St John W, Patterson E. Recovery from total hip replacement surgery: “It’s not just physical.” *Qual Health Res*. 2009;19(11):1612–20.
12. Jayakumar P, Ramachandran M, Youm T, Achan P. Arthroscopy of the hip for paediatric and adolescent disorders: Current concepts. *Journal of Bone and Joint Surgery - Series B*. 2012.
13. Christensen CP, Althausen PL, Mittleman MA, Lee J, McCarthy JC. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res* [Internet]. 2003;406(406):75–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12579003>
14. Klässbo 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.
15. Griffin DR, Parsons N, Mohtadi NGH, Safran MR. A short version of the International

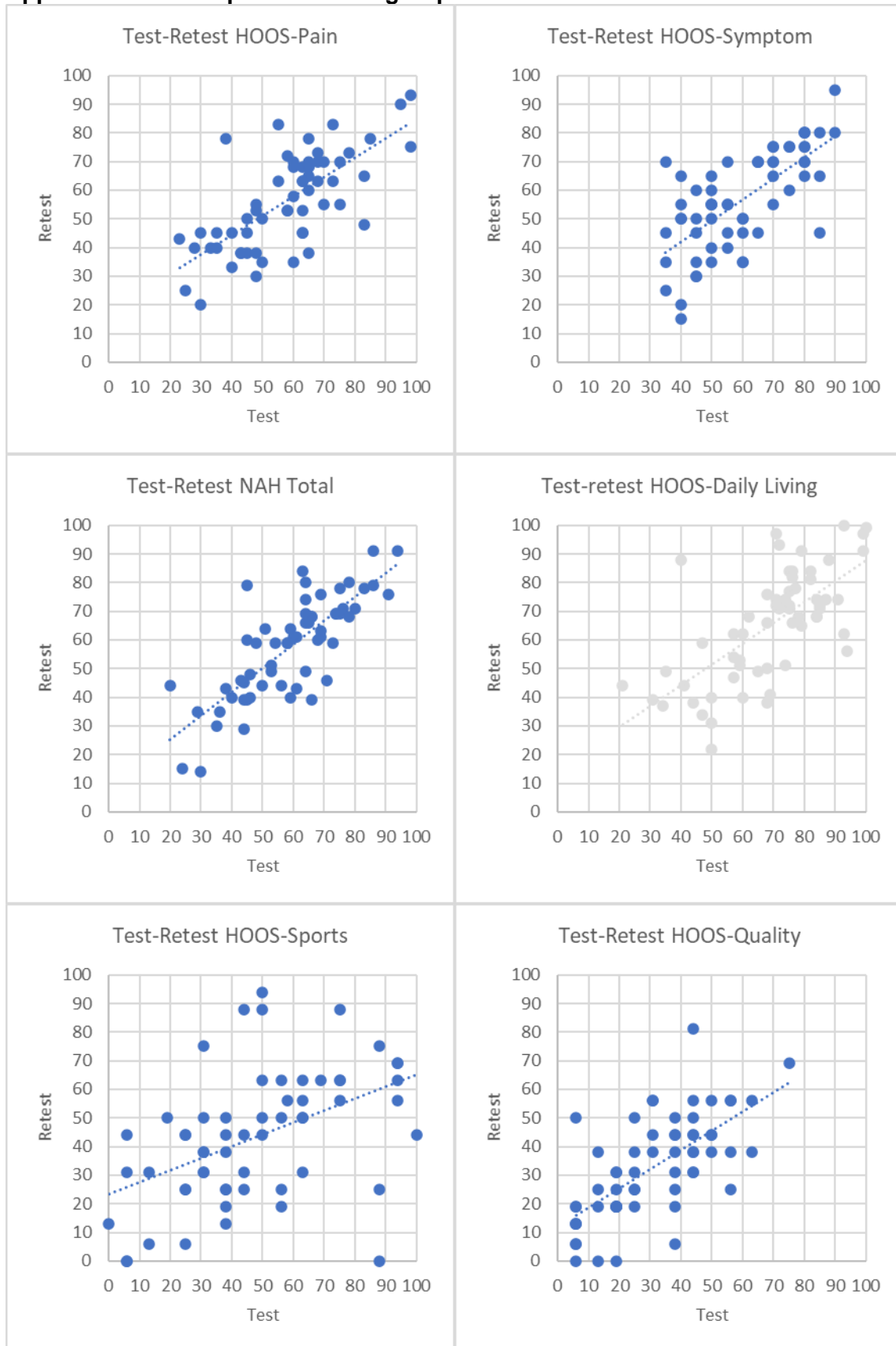
- Hip Outcome Tool (iHOT-12) for use in routine clinical practice. *Arthrosc - J Arthrosc Relat Surg*. 2012;28(5):611–8.
16. McCormack HM, Horne DJ d. L, Sheather S. Clinical applications of visual analogue scales: A critical review. *Psychol Med*. 1988;18(4):1007–19.
 17. Huskisson EC. Measurement fo pain. *Lancet*. 1974;304(7889):1127–31.
 18. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain. *Arthritis Care Res (Hoboken)*. 2011;63(11):S210–52.
 19. Will G Hopkins. A New View of Statistics [Internet]. 2002 [cited 2021 Feb 15]. p. 1–4. Available from: <https://www.sportsci.org/resource/stats/effectmag.html>
 20. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg Res*. 2015;10(24):1–11.
 21. Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res*. 2018;476(12):2432–2441.
 22. Martin RRL, Kivlan BR, Christoforetti JJ, Wolff AB, Nho SJ, Salvo JP, et al. Minimal Clinically Important Difference and Substantial Clinical Benefit Values for the 12-Item International Hip Outcome Tool. *Arthrosc - J Arthrosc Relat Surg*. 2019;35(2):411–6.
 23. Meghpara MB, Yelton MJ, Annin S, Shapira J, Rosinsky PJ, Maldonado DR, et al. Mid-Term Outcomes of Endoscopic Gluteus Medius Repair With Concomitant Arthroscopic Labral Treatment: A Propensity-Matched Controlled Study. *Arthrosc - J Arthrosc Relat Surg*. 2020;36(11):2856–65.
 24. Martin RRL, Kivlan BR, Christoforetti JJ, Wolff AB, Nho SJ, Salvo JP, et al. Minimal Clinically Important Difference and Substantial Clinical Benefit Values for a Pain Visual Analog Scale After Hip Arthroscopy. *Arthrosc - J Arthrosc Relat Surg*. 2019;35(7):2064–9.
 25. Pellisé F, Vidal X, Hernández A, Cedraschi C, Bagó J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine (Phila Pa 1976)*. 2005;30(3):365–8.
 26. Aleem IS, Duncan J, Ahmed AM, Zarrabian M, Eck J, Rhee J, et al. Do Lumbar Decompression and Fusion Patients Recall Their Preoperative Status? *Spine (Phila Pa 1976)* [Internet]. 2017;42(2):128–34. Available from:

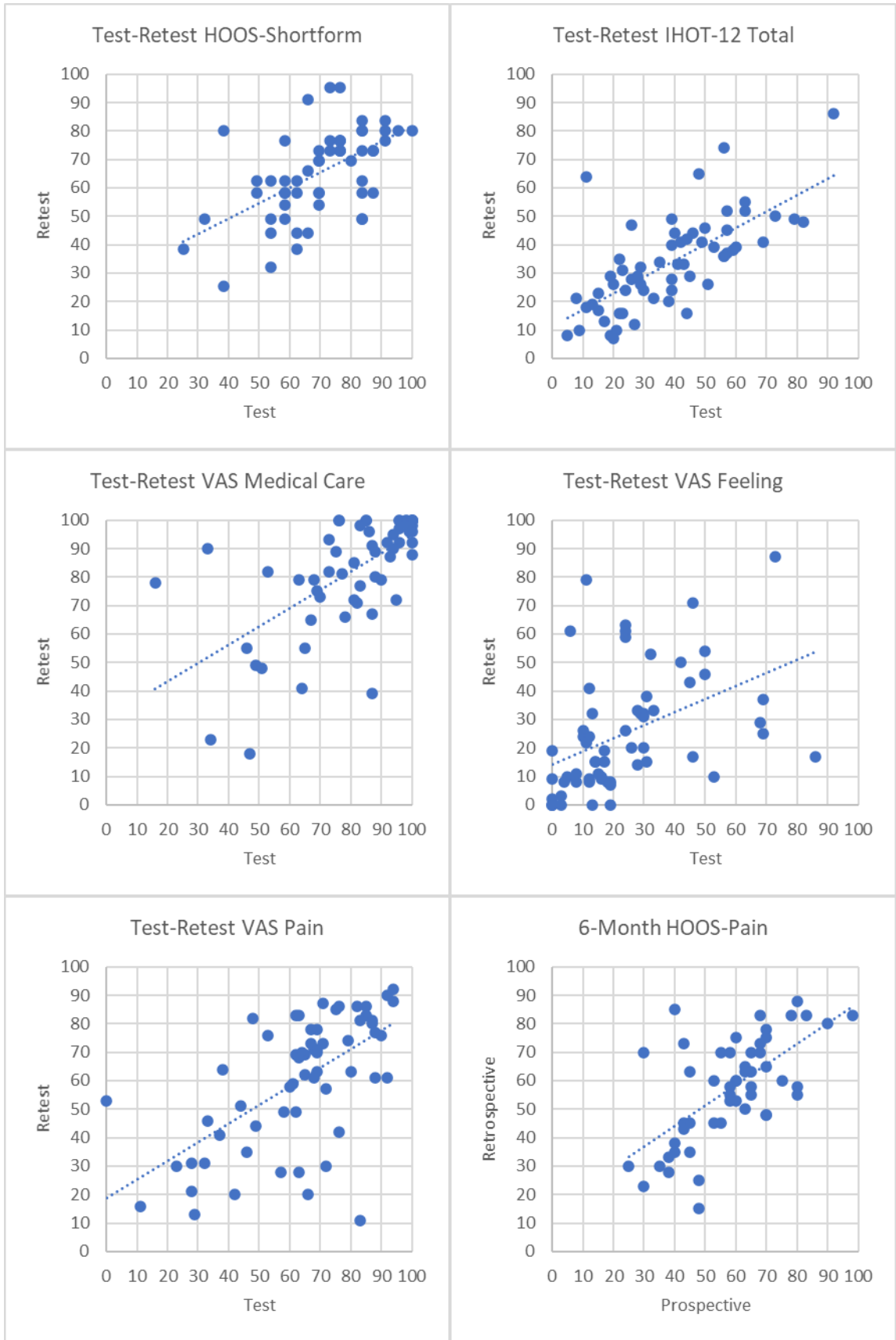
<http://insights.ovid.com/crossref?an=00007632-201701150-00018>

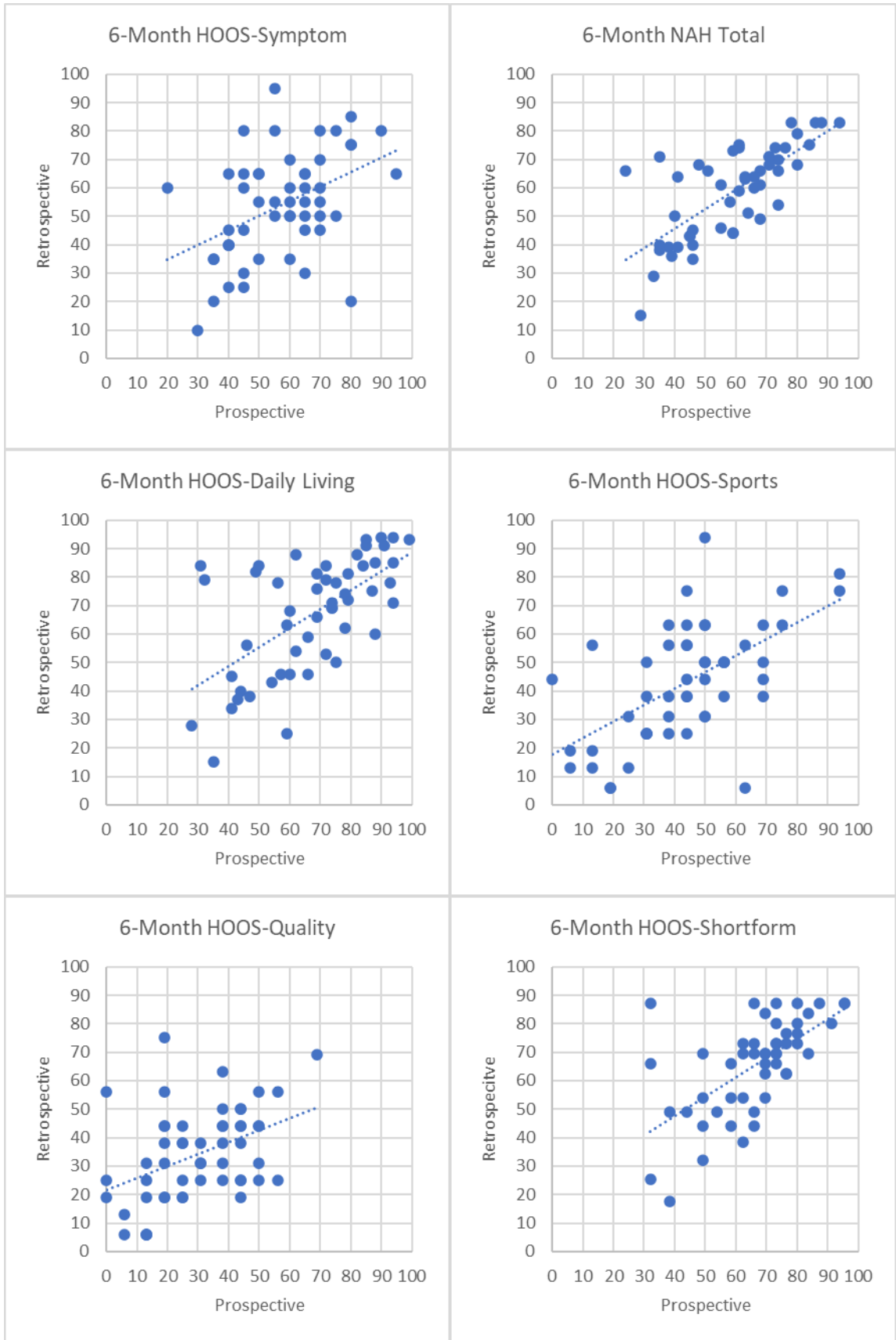
27. Wormdal AH, Tallaksen SH, Hagen Ø, Foss OA, Finsen V. Remembered preoperative Quick disabilities of the arm, shoulder and hand (QuickDASH) scores. *J Hand Surg (European Vol [Internet]*. 2017;42(1):78–83. Available from: <http://journals.sagepub.com/doi/10.1177/1753193416665299>
28. Gotlin MJ, Kingery MT, Baron SL, McCafferty J, Jazrawi LM, Meislin RJ. Recall Bias in Retrospective Assessment of Preoperative Patient-Reported American Shoulder and Elbow Surgeons Scores in Arthroscopic Rotator Cuff Repair Surgery. *Am J Sports Med [Internet]*. 2020;48(6):1471–5. Available from: <https://doi.org/10.1177/0363546520913491>
29. Reynolds N, Thirkannad S. The recall DASH score - a novel research tool. *Hand Surg*. 2013;18(1):11–4.
30. Griffin DR, Dickenson EJ, Wall PDH, Achana F, Donovan JL, Griffin J, et al. Hip arthroscopy versus best conservative care for the treatment of femoroacetabular impingement syndrome (UK FASHIoN): a multicentre randomised controlled trial. *Lancet*. 2018;391(10136):2225–35.
31. Chhapola V, Kanwal SK, Brar R. Reporting standards for Bland–Altman agreement analysis in laboratory research: a cross-sectional survey of current practice. *Ann Clin Biochem*. 2015;52(3):382–6.
32. Johnson EC, Horwood J, Gooberman-Hill R. Conceptualising time before surgery: The experience of patients waiting for hip replacement. *Soc Sci Med*. 2014;116:126–33.

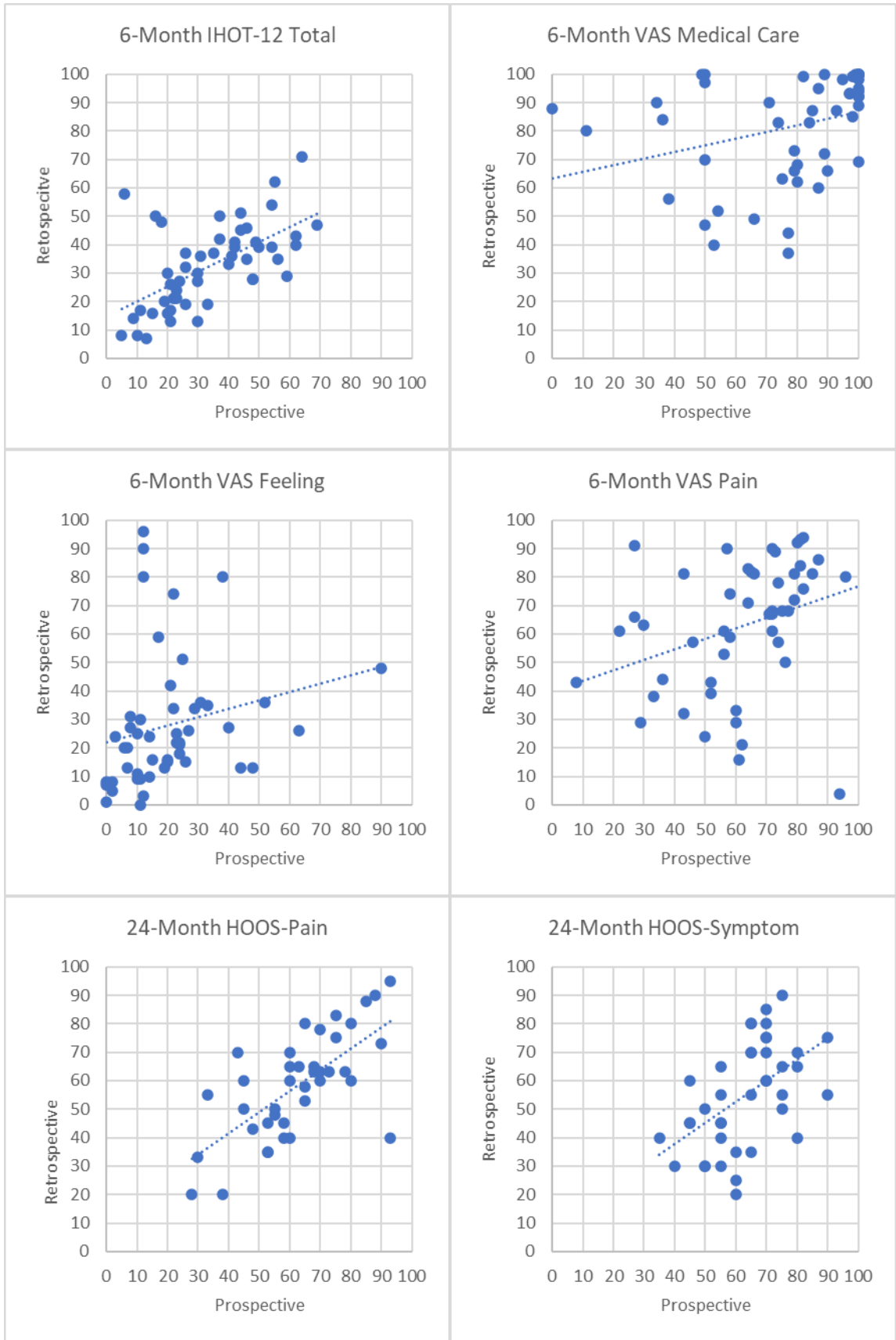
SECTION 3: Appendices

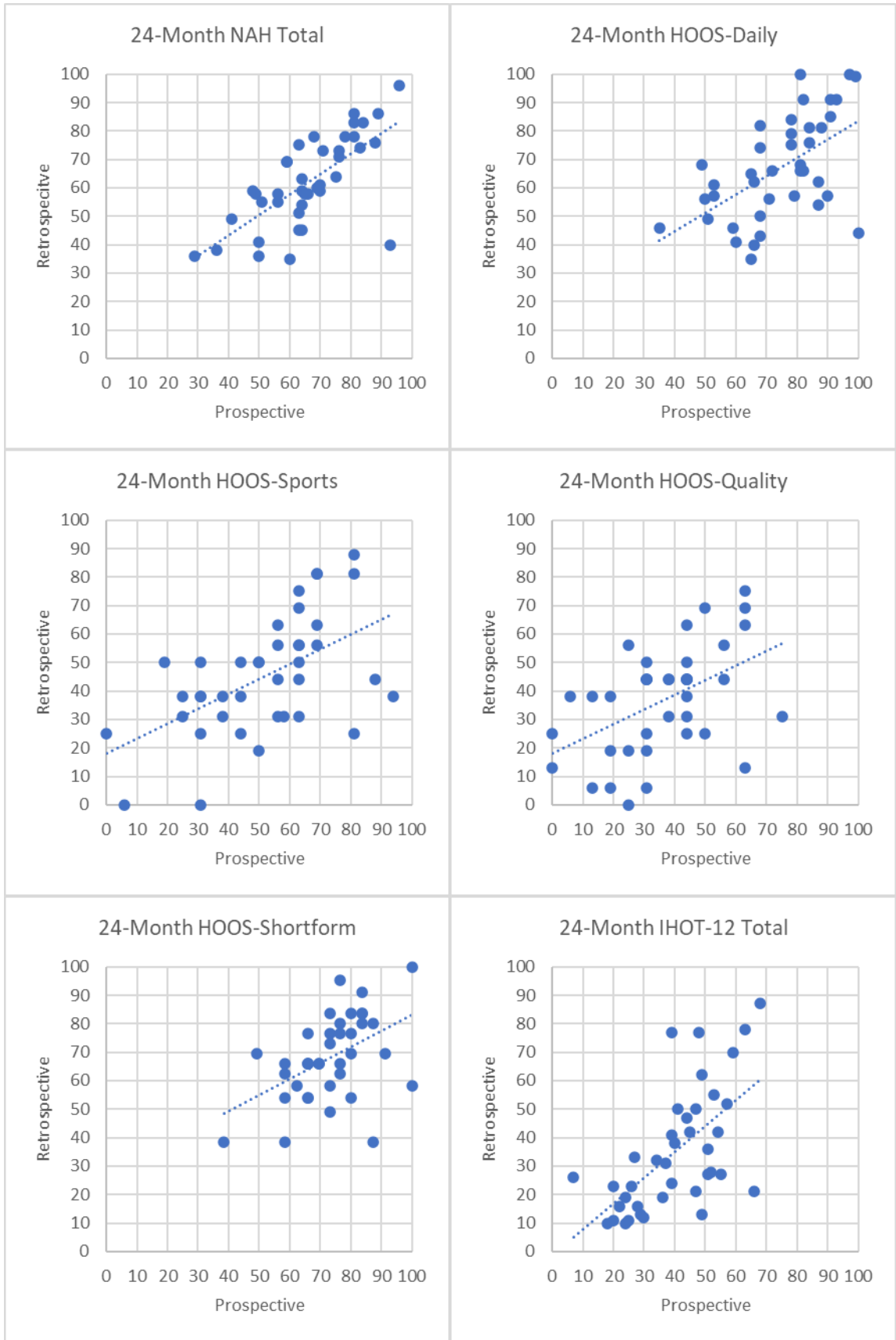
Appendix A: Scatterplots for each group



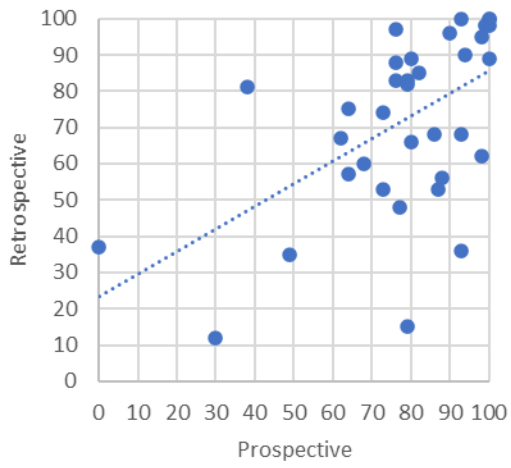




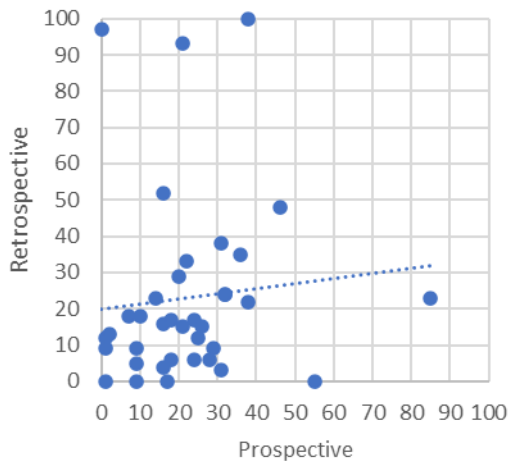




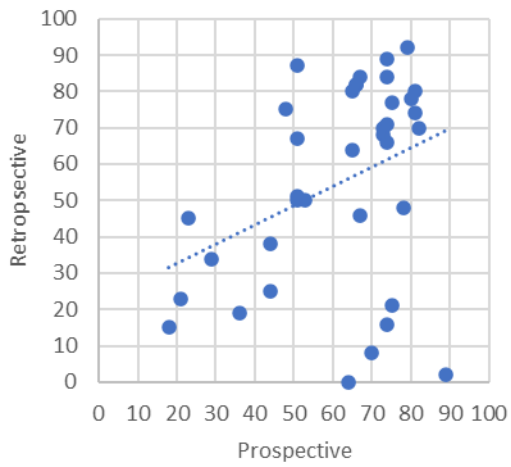
24-Month VAS Medical Care



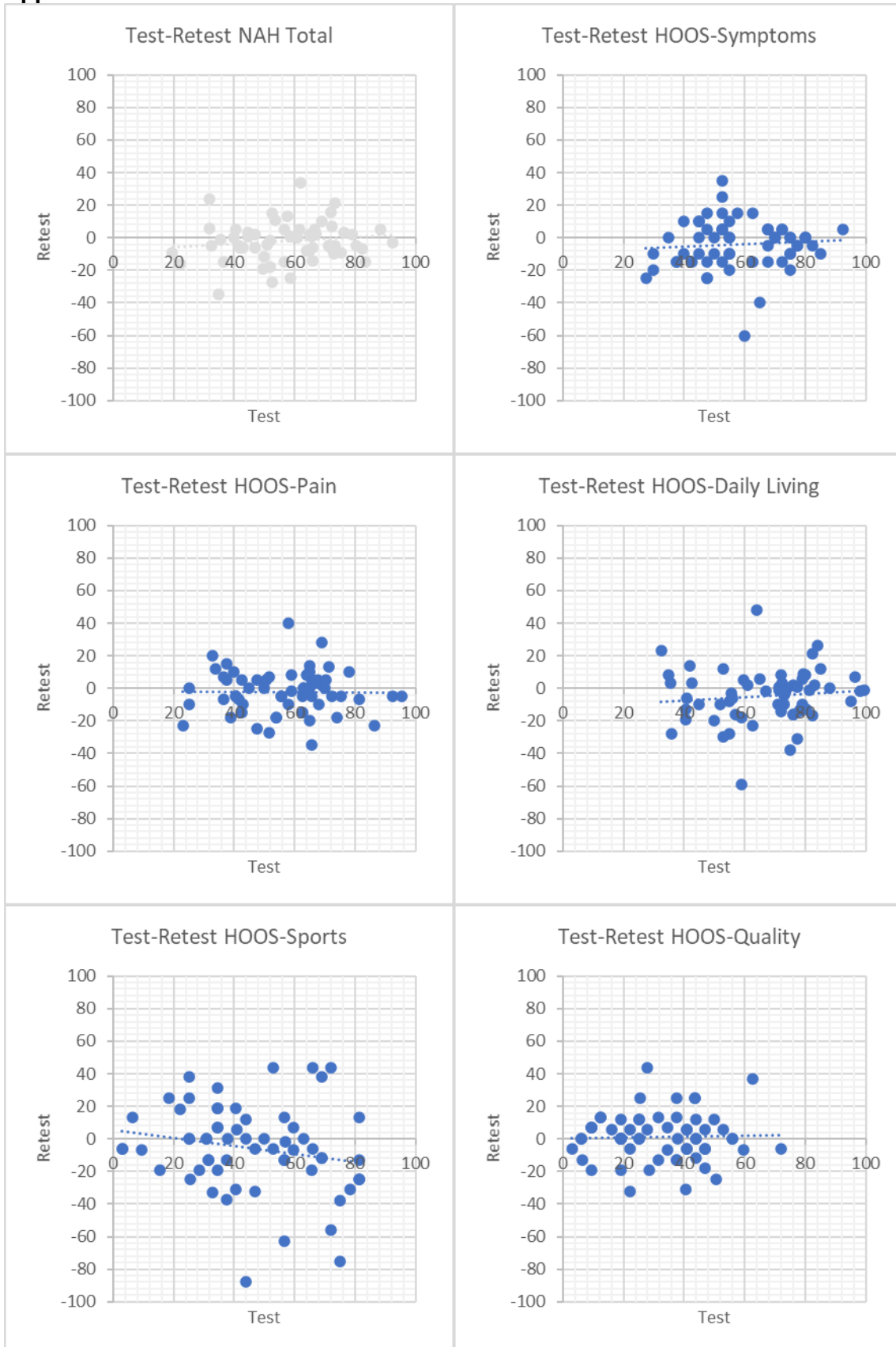
24-Month VAS Feeling

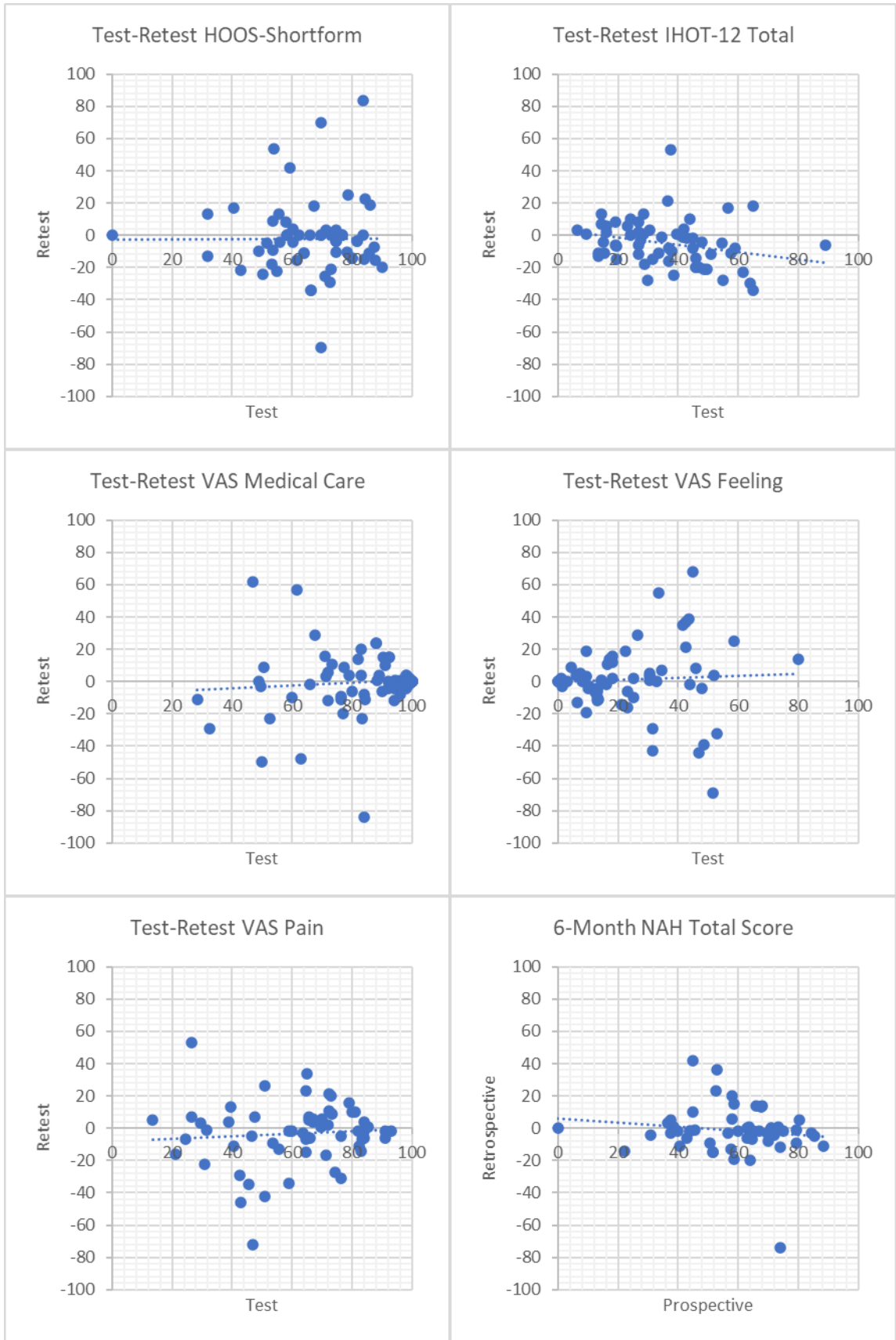


24-Month VAS Pain

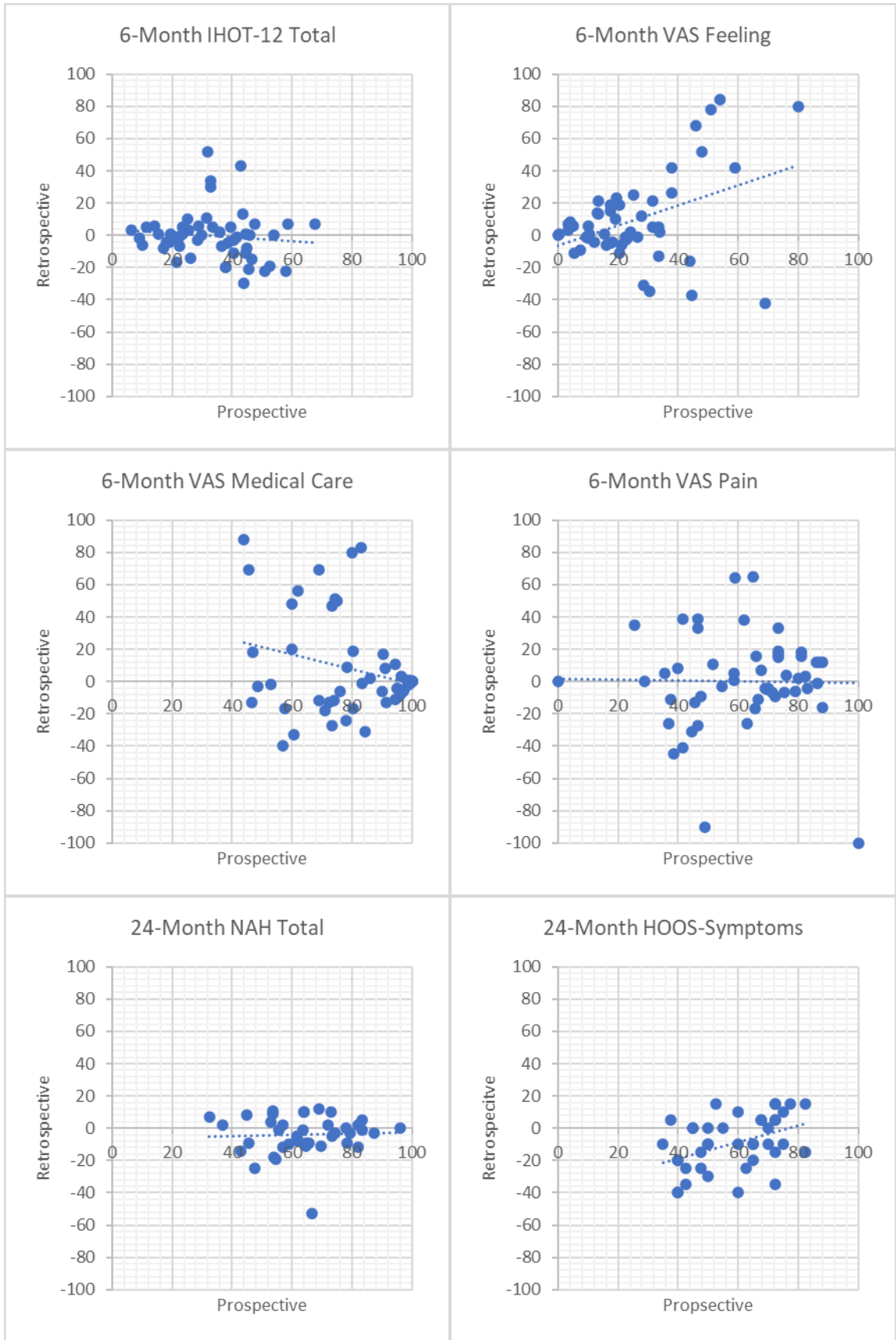


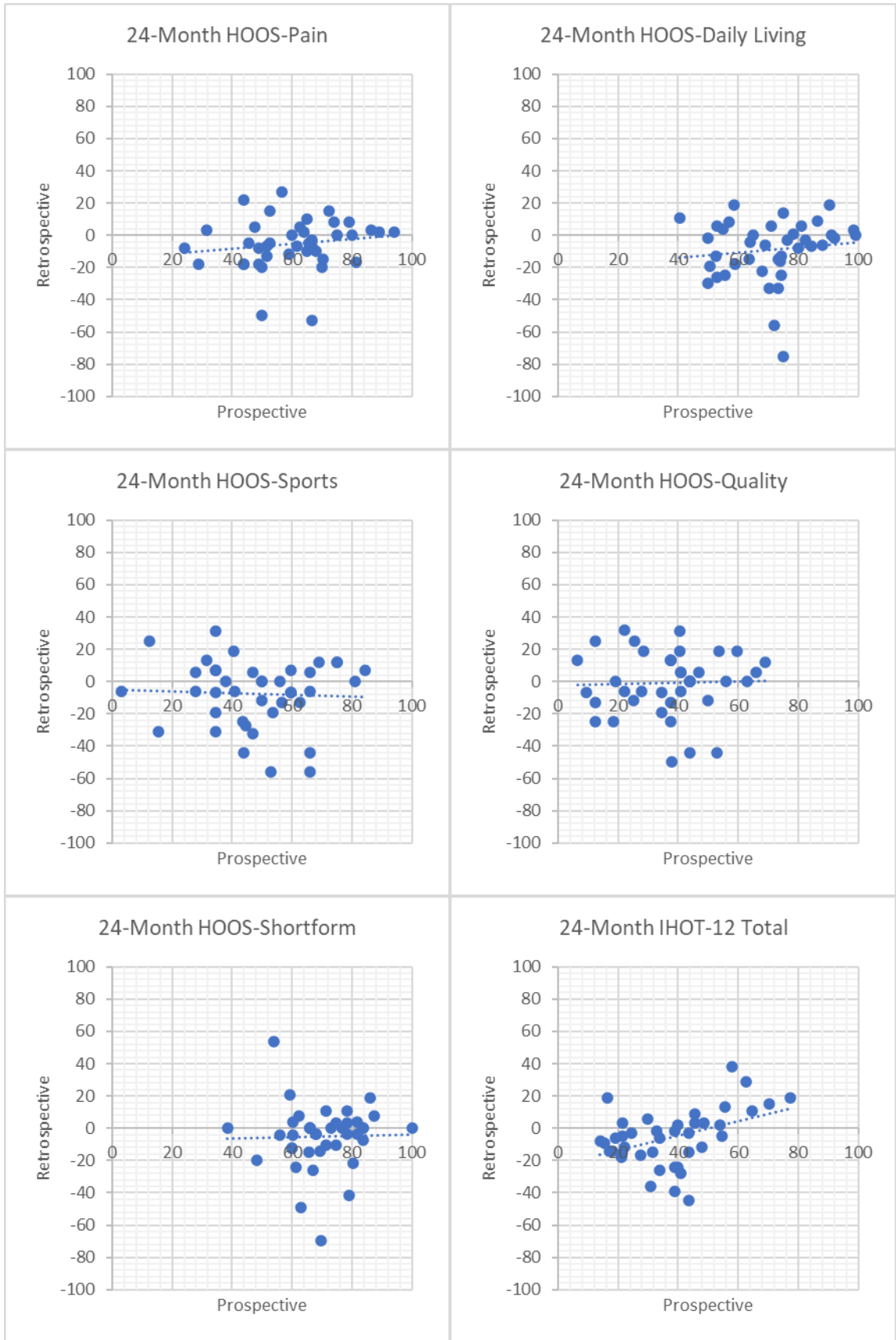
Appendix B: Bland-Altman Plots



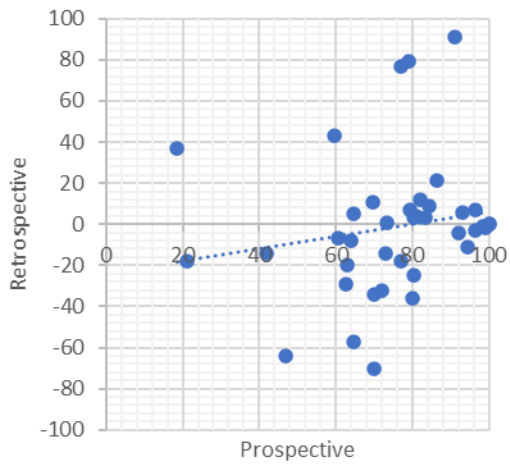




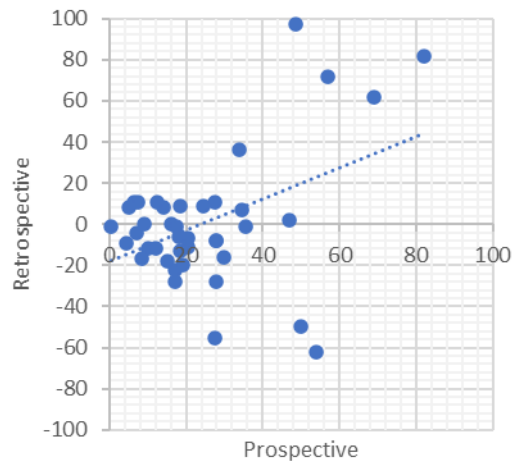




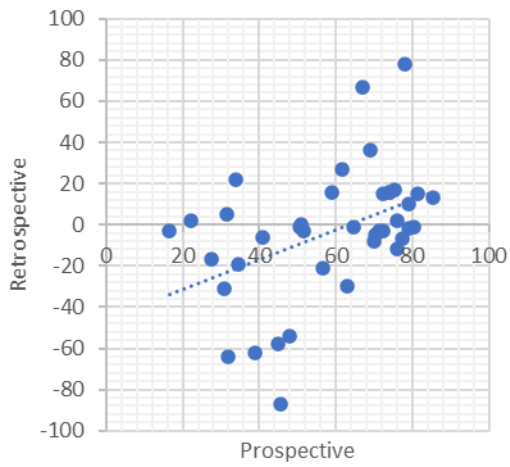
24-Month VAS Medical Care



24-Month VAS Feeling



24-Month VAS Pain



Appendix C: Hip Osteoarthritis Outcome Score (HOOS)

HOOS Hip Score

Patient Name _____ ID _____ Side Right
 Left

Date of review: ____/____/____

Follow up period: **PreOp** OR _____ weeks/months/years (circle one)

INSTRUCTIONS: This survey asks for your view about your hip. This information will help us keep track of how you feel about your hip and how well you are able to do your usual activities. Answer every question by ticking the appropriate box, only one box for each question. If you are unsure about how to answer a question, please give the best answer you can.

Symptoms

These questions should be answered thinking of your hip symptoms and difficulties during the **last week**.

S1. Do you feel grinding, hear clicking or any other noise from your hip?

Never Rarely Sometimes Often

Always

S2. Difficulties spreading legs wide apart?

None Mild Moderate Severe

Extreme

S3. Difficulties to stride out when walking?

None Mild Moderate Severe

Extreme

Stiffness

The following questions concern the amount of joint stiffness you have experienced during the **last week** in your hip. Stiffness is a sensation of restriction or slowness in the ease with which you move your hip joint.

S6. How severe is your hip joint stiffness after first wakening in the morning?

None Mild Moderate Severe

Extreme

S7. How severe is your hip stiffness after sitting, lying or resting later in the day?

None Mild Moderate Severe

Extreme

Pain

P1. How often is your hip painful?

Never Monthly Weekly Daily

Always

What amount of hip pain have you experienced the last week during the following activities?

P2. Straightening hip fully

None Mild Moderate Severe

Extreme

P3. Bending hip fully

None Mild Moderate Severe

Extreme

P4. Walking on flat surface

None Mild Moderate Severe

Extreme

P5. Going up or down stairs

None Mild Moderate Severe

Extreme

P6. At night while in bed

None Mild Moderate Severe

Extreme

P7. Sitting or lying

None Mild Moderate Severe

Extreme

P8. Standing upright

None Mild Moderate Severe

Extreme

P9. Walking on a hard surface (asphalt, concrete etc)

None Mild Moderate Severe

Extreme

P10. Walking on an uneven surface

None Mild Moderate Severe

Extreme

Function, daily living-

The following questions concern your physical function. By this we mean your ability to move around and to look after yourself. For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your hip.

A1. Descending stairs

None Mild Moderate Severe

Extreme

A2. Ascending stairs

None Mild Moderate Severe

Extreme

A3. Rising from sitting

None Mild Moderate Severe

Extreme

Indicate the degree of difficulty you have experienced in the last week due to your hip.

A4. Standing

None Mild Moderate Severe

Extreme

A5. Bending to floor/pick up an object

None Mild Moderate Severe

Extreme

A6. Walking on flat surface

None Mild Moderate Severe

Extreme

A7. Getting in/out of car

None Mild Moderate Severe

Extreme

A8. Going shopping

None Mild Moderate Severe

Extreme

A9. Putting on socks/stockings

None Mild Moderate Severe

Extreme

A10. Rising from bed

None Mild Moderate Severe

Extreme

A11. Taking off socks/stockings

None Mild Moderate Severe

Extreme

A12. Lying in bed (turning over, maintaining hip position)

None Mild Moderate Severe

Extreme

A13. Getting in/out of bath

None Mild Moderate Severe

Extreme

A14. Sitting

None Mild Moderate Severe

Extreme

A15. Getting on/off toilet

None Mild Moderate Severe

Extreme

A16. Heavy domestic duties (moving heavy boxes, scrubbing floors, etc)

None Mild Moderate Severe

Extreme

A17. Light domestic duties (cooking, dusting, etc)

None Mild Moderate Severe

Extreme

Function, sports and recreational activities

The following questions concern your physical function when being active on a higher level. The questions should be answered thinking of what degree of difficulty you have experienced during the **last week** due to your hip.

SP1. Squatting

None Mild Moderate Severe

Extreme

SP2. Running

None Mild Moderate Severe

Extreme

SP3. Twisting/pivoting on loaded leg

None Mild Moderate Severe

Extreme

SP4. Walking on uneven surface

None Mild Moderate Severe

Extreme

Quality of Life

Q1. How often are you aware of your hip problem?

Never Monthly Weekly Daily

Constantly

Q2. Have you modified your life style to avoid potentially damaging activities to your hip?

Not at all Mildly Moderately Severely

Totally

Q3. How much are you troubled with lack of confidence in your hip?

Not at all Mildly Moderately Severely

Extremely

Q4. In general, how much difficulty do you have with your hip?

None Mild Moderate Severe

Extreme

Thank you very much for completing all the questions in this questionnaire.

Appendix D: Non-Arthritic Hip (NAH) Score

S5 Non Arthritic Hip score

Patient Name _____ ID _____ Side Right

Left

Date of review: ____/____/____

Follow up period: PreOp OR _____ weeks/months/years (circle one)

INSTRUCTIONS: The following 5 questions concern the amount of pain you are currently experiencing in the hip that you are having evaluated today. For each situation, please select the response that most accurately reflects the amount of pain experienced in the **past 48 hours**. Please select one answer that best describes your situation.

How much pain do you have -

Walking on a flat surface

None Mild Moderate Severe Extreme

Going up or down stairs

None Mild Moderate Severe Extreme

At night while in bed

None Mild Moderate Severe Extreme

Sitting or lying

None Mild Moderate Severe Extreme

Standing upright

None Mild Moderate Severe Extreme

INSTRUCTIONS: The following 4 questions concern the symptoms that you are currently experiencing in the hip that you are having evaluated today. For each situation, please select the response that most accurately reflects the symptoms experienced in the **past 48 hours**. Please select one answer that best describes your situation.

How much trouble do you have with

Catching or locking of your hip

None Mild Moderate Severe Extreme

Your hip giving out on you?

None Mild Moderate Severe Extreme

Stiffness in your hip

None Mild Moderate Severe Extreme

Extreme

Decreased motion in your hip

None Mild Moderate Severe

Extreme

INSTRUCTIONS: The following 5 questions concern your physical function. For each of the following activities, please select the response that most accurately reflects the difficulty that you have experienced in the **past 48 hours** because of your hip pain. Please select one answer that best describes your situation.

What degree of difficulty do you have with
Descending stairs

None Mild Moderate Severe

Extreme

Ascending stairs

None Mild Moderate Severe

Extreme

Rising from sitting

None Mild Moderate Severe

Extreme

Putting on socks and stockings

None Mild Moderate Severe

Extreme

Rising from bed

None Mild Moderate Severe

Extreme

INSTRUCTIONS: The following 6 questions concern your ability to participate in certain types of activities. For each of the following activities, please select the response that most accurately reflects the difficulty that you have experienced **in the past month** because of your hip pain. If you do not participate in a certain type of activity, please estimate how much trouble your hip would cause you if you had to perform that type of activity. Please select one answer that best describes your situation.

How much trouble does your hip cause you when you participate in

High demand sports involving sprinting or cutting (for example, football, basketball, tennis, and exercise aerobics)?

None Mild Moderate Severe

Extreme

Low demand sports (for example, golfing and bowling)

None Mild Moderate Severe

Extreme

Jogging for exercise?

None Mild Moderate Severe

Extreme

Walking for exercise?

None Mild Moderate Severe

Extreme

Heavy household duties (for example, lifting firewood and moving furniture)

None Mild Moderate Severe

Extreme

Light household duties (for example, cooking, dusting, vacuuming, and doing laundry)

None Mild Moderate Severe

Extreme

Appendix E: International Hip Outcome Tool 12-Item (IHOT-12)

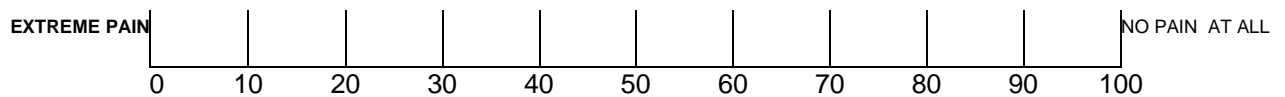
International Hip Outcome Tool Item-12

Patient Name _____ ID _____ Side Right
 Left

Date of review: ____/____/____ OR Follow up period: PreOp OR _____
weeks/months/years (circle one)

The following questions ask about symptoms that you may experience in your **hip** and about the function of your **hip** with respect to daily activities. Please think about how you have felt most of the time over the past **month** and answer accordingly by marking the line.

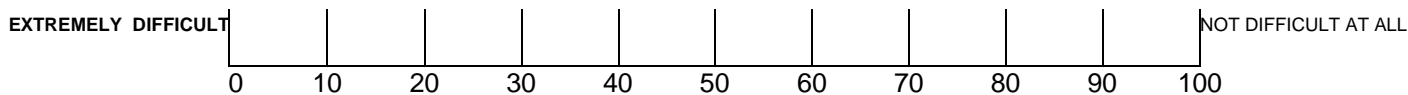
1. Overall, how much pain do you have in your hip/groin?



2. How difficult is it for you to get up and down off the floor/ground?



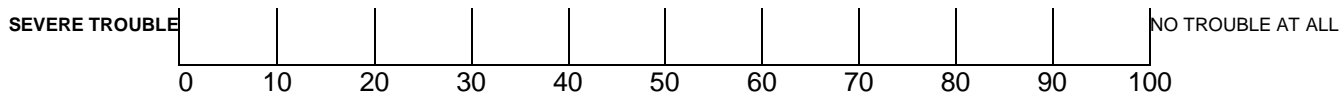
3. How difficult is it for you to walk long distances?



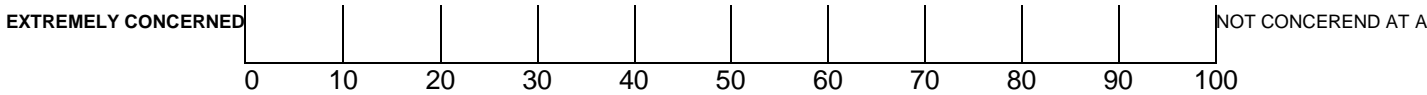
4. How much trouble do you have with grinding, catching or clicking in your hip?



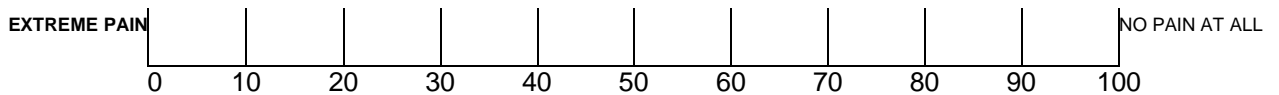
5. How much trouble do you have pushing, pulling, lifting or carrying heavy objects?



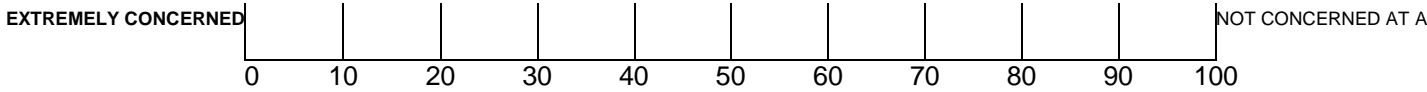
6. How concerned are you about cutting/changing directions during your sport or recreational activities?



7. How much pain do you experience in your hip after activity?

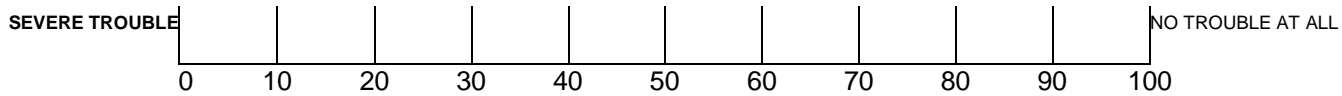


8. How concerned are you about picking up or carrying children because of your hip?

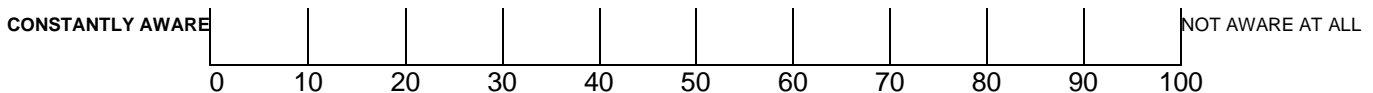


9. How much trouble do you have with sexual activity because of your hip?

This is not relevant to me



10. How much of the time are you aware of the disability in your hip?



11. How concerned are you about your ability to maintain your desired fitness level?



0 10 20 30 40 50 60 70 80 90 100

12. How much of a distraction is your hip problem?



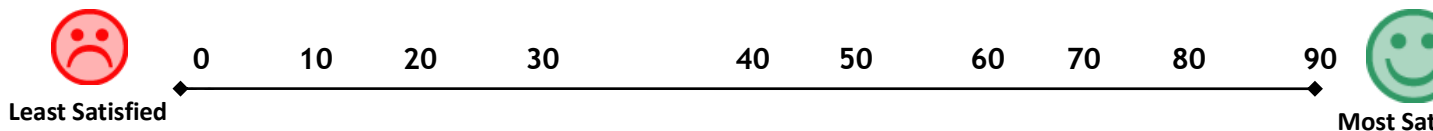
Appendix F: Preoperative Visual Analogue Scale (VAS) for Medical Care, Feeling and Pain

VAS Patient Satisfaction and Pain

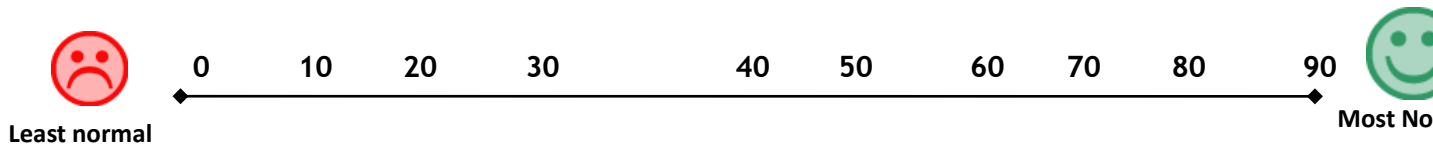
Patient Name _____ ID _____ Side Right
 Left

Date of review: ____/____/____

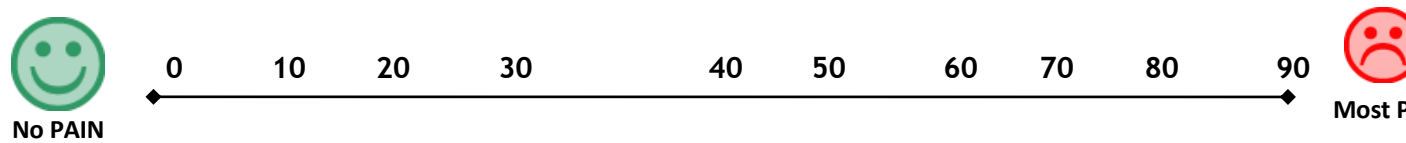
Please circle the number on the scale at the level which most corresponds to **how satisfied** you are with your medical care.



Please circle the number on the scale at the level which most corresponds to **how normal** your operated joint feels compared to the other one.



Please circle the number on the scale at the level which most corresponds to **how much pain** you feel in your operated joint compared to the other one.



Appendix G: Arthroscopy: The Journal of Arthroscopic and Related Surgery Guide for Authors

INTRODUCTION

All submissions to *Arthroscopy: The Journal of Arthroscopic and Related Surgery* must comply with these Instructions for Authors. Studies should be in compliance with human studies committees and animal welfare regulations at the authors' institutions and also in compliance with Food and Drug Administration guidelines. All manuscripts will be subject to peer review. Letters to the Editor and comments on the Journal's content or policies are always welcome and encouraged.

Separate Title Page

A separate (unblinded) title page of each manuscript should include the following essential information:

1. **Title.** Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
2. **All Authors' full names, degrees, and affiliations.** Where the family name may be ambiguous (e.g., a double name), please indicate this clearly. Present each author's affiliation and address below the names.
3. **Corresponding Author.** Clearly indicate who will handle correspondence at all stages of reviewing and publication, and after publication. Ensure that telephone numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address. Contact details must be kept up to date by the corresponding author.
4. **In addition,** include IRB and RCT information, as well as a short running title (maximum of 45 characters and spaces). Include any acknowledgment of persons who provided help during the research/writing (e.g., language help, writing assistance, or proof reading the manuscript, etc.).

Blinding the Manuscript

Because all manuscripts are blinded to reviewers, the first page of the blinded manuscript must be a blinded title page that lists *only the title*. Likewise, in the text, do not include any identifying information, such as an author's initials or the names of institutions where the study was done, or a phrase such as "our study" that, when followed by a citation, reveals authorship of the present manuscript in the reference list.

Manuscript Structure

1. Abstract

Original Articles, abstracts should be a *maximum of 300 words* and structured to include the following sections: *Purpose:* One or 2 sentences that simply state the purpose with no background information or hypothesis. *Methods:* Provide, with sufficient detail, the methods of the study including selection criteria. *Results:* Provide results, with data, P values, and standard deviation of mean or 95% confidence intervals. Present most important findings first. Please provide exact P values (not $P <$) and numbers to support your methods findings. *Conclusions:* State only what your study found; do not include extraneous information not backed up by the results. *Level of Evidence* (for human studies) or *Clinical*

Relevance (basic science or in vitro study: why is this study important from a clinical standpoint?).

2. Introduction

The introduction of an Original Article should succinctly state the problem or controversy that led you to undertake the study, including a concise review of only the most relevant literature. Conclude the introduction by stating the *purpose* of the study and your *hypothesis*. The purpose in the Introduction should match that of the Abstract.

3. Methods

Describe the study design (prospective or retrospective, inclusion and exclusion criteria, duration). If prospective or a cadaver study, the number of enrolled subjects is reported in Methods. If retrospective, the study population (numbers, demographics, length of follow-up) should be in Results.

Include IRB and animal studies information. IRB approval is required for all human studies except retrospective and cadaver studies (unless the institution where the study was performed requires it).

The statistics that you have used to analyze the data should be described in detail. You cannot make the statement, "We found no significant difference between the two groups" unless a power study was done, and you include in the text the value of alpha, beta, and standard deviation. Use of the word *significant* requires your reporting an exact *P* value. Confidence intervals of 95% are required whenever the results of survivorship analysis are given in the text, tables, or figures. Use of the word *correlation* requires you to report the correlation coefficient.

Arthroscopy encourages the use of validated outcome instruments. The use of both a general health outcome measure and a joint-specific, limb-specific, or condition-specific measure is encouraged. If an outcome instrument leads to a categorical ranking (e.g., excellent or good or poor), the aggregate outcome score for each patient should be provided.

4. Results

Describe in detail the data obtained during the study following the order of the Methods to include final number of subjects, demographics, length of follow-up (mean and range). The overall final patient follow-up should be 80% or greater (less than 20% drop-out) in order to minimize follow-up bias. In general, scientific studies will not be accepted for publication without meeting this criterion. **Results obtained with less than two years of follow-up are rarely accepted for publication by the Journal.** All data in the text must be consistent with the rest of the manuscript, including data in tables, figures, and legends. Present comparison data in tables and present as mean \pm standard error of the mean with confidence intervals.

5. Discussion

Be concise. The Discussion should start with the most important findings of your study. Is your hypothesis affirmed or refuted? Compare and contrast your study with others in the most relevant world literature, particularly the recent literature. A complete literature review is unnecessary.

At the end of the Discussion, under the subheading "Limitations," review the limitations of your study.

6. Conclusions

Briefly state your new (or verified) view of the problem you outlined in the Introduction. Take special care to draw your conclusions only from your results and verify that your conclusions are firmly supported by your data. Most importantly, do not make concluding statements that are not supported by your data, lie beyond the scope of your study, or are unnecessary (e.g., "further studies are warranted"). **The conclusions in the text must match those in the abstract.**

7. References

The Journal follows the reference style in "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" (see <http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html#g>). Provide all authors' names when 6 or fewer; when 7 or more, list the first 3 and add et al. Provide article titles and inclusive page numbers (321-328, not 321-8). References to online-only material must list author, title, the URL, and the date accessed by the author. For abbreviations of journal names, refer to PubMed. Please ensure that every reference cited in the text is present in the reference list (and vice versa). **The accuracy of reference data is the responsibility of all authors.**

Data References

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. This identifier will not appear in your published article.

Reference style

In text: Number references in the order in which they appear in the text. Unpublished results and personal communications (only if essential to your message) should be mentioned in the body of the text at the end the statement with the appropriate information in parentheses. For example: (J. Karlsson, M.D., personal communication, [month and year of communication]).

Formatting Examples

Periodical

Jackson TJ, Lindner D, El-Bitar YF, Domb BG. Effect of femoral anteversion on clinical outcomes after hip arthroscopy. *Arthroscopy* 2015;31:35-41.

Chapter in a book

Ruch DS, Poehling GG. Operative arthroscopy of the wrist. In: Andrews JR, Timmerman LA, eds. *Diagnostic and operative arthroscopy*. Philadelphia: WB Saunders, 1997;199-205.

Book

Burkhart SS, Lo IKY, Brady PC, Denard PJ. *The cowboy's companion: A trail guide for the arthroscopic shoulder surgeon*. Philadelphia: Lippincott Williams & Williams, 2012.

Article in Press

Note: Citation of an 'in press' article is permitted only if it has been accepted for publication. Rosso F, Bisicchia S, Bonasia DE, Amendola A. Meniscal allograft transplantation: A systematic review. *Am J Sports Med* in press, available online 13 June, 2014. doi:10.1177/0363546514536021.

Dataset

[dataset] Oguro, M, Imahiro, S, Saito, S, Nakashizuka, T. Mortality data for Japanese oak wilt disease and surrounding forest compositions, Mendeley Data, v1; 2015. <https://doi.org/10.17632/xwj98nb39r.1>.

8. Figure and Video Legends

Ensure that each illustration and each part of a multipart illustration has a legend (caption). Supply legends separately, not attached to the figure. Figure legends must be robust and "stand alone" (i.e., contain a complete, take-home, educational message, as if a reader viewed only that Figure without looking at any other Figure or without reading the text). Orient the reader to the image by mentioning patient position, side, and viewing portal or MRI orientation as appropriate. Keep text in the illustrations themselves to a minimum but explain in the legend all symbols and abbreviations used.

9. Tables

Number tables consecutively in accordance with their appearance in the text. Include a short descriptive title with the table number. Place footnotes to tables below the table body and indicate them according to the symbol hierarchy (i.e., asterisk, dagger, double dagger, etc.). Define all abbreviations. Avoid vertical rules. Do not give the same information in tables that you give in the text or in figures.

10. Figures

Number figures consecutively in accordance with their appearance in the text. Figures must be submitted separately from the text. Arrows and labels should be added to figures as appropriate to orient the reader to the arthroscopic images. Previously published figures may be used if permission has been received from the source publisher.

11. Disclosures

After the figures, you will upload each author's completed *Arthroscopy* ICMJE form. These forms must be completed, signed, and submitted with the manuscript.

Discussion and Conclusion should follow the Journal's guidelines for original research.

Common Errors

- **Including studies with duplicate patient populations.** In some instances, a SR turns up studies on the same patient group. Including these studies in any statistical analysis artificially inflates the number of patients and should be avoided.
- **Pooling diverse, heterogeneous studies with different designs.** Combining non-randomized studies with randomized trials is typically not appropriate as these designs carry different risks of bias and are apt to distort the results. If a SR includes studies with different designs (randomize trials, cohort studies, etc.) these should be pooled separately. Typically, these are level III or IV evidence studies.
- **No rationale for provided for pooling non-randomized studies.** If the available literature is limited to observational studies, a rationale for why a meta-analysis will produce valid results that contribute to the understanding of the problem under question is needed. If one can not be reached, a meta-analysis should be avoided.
- **Quantifying heterogeneity but not failing to explore or discuss it.** Reporting of the I^2 statistic has become more frequent however it's important to discuss its impact on the results. If the results are heterogeneous efforts should be undertaken to explore this inconsistency. Techniques like subgroup analysis can be used to determine if I^2 values change when grouped according to co-variants. For example, I^2 values may change when the studies are analysed according to a clinical characteristic (those that included patients with bone loss vs. those that did not) or a risk of bias item (those that adequately randomized patients versus those that did not). Lastly, I^2 is a relative measure. As recommended above, providing a prediction interval will assist in interpreting the effect of heterogeneity. A prediction interval provides a range of probable effects that reflects the variation in the different studies and settings, including what would be expected in future patients.

SUBMISSION CHECKLIST

The following checklist will be useful before sending a manuscript to the journal for review. Ensure that the following items are present:

One author has been designated as the corresponding author with the following contact details:

- E-mail address
- Full postal address
- Telephone numbers

All necessary files have been uploaded, and contain:

- All figure legends
- All tables (including title, description, footnotes)
- Separate files for figures
- ICMJE forms for all authors

Further considerations:

- Manuscript has been spell-checked and grammar-checked
- References are in the correct format for *Arthroscopy*
- All references included in the reference list are cited in the text, and vice versa
- Permission has been obtained for use of copyrighted material from other sources, including the Web

