Designation: F3448 – 20

Standard Guide for Clinical Outcomes for Clinical Trials and/or Clinical Registries for Hip Reconstructive Surgery[1]

This standard is issued under the fixed designation F3448; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide is intended as a resource for individuals and organizations when designing clinical trials and/or clinical registries and addresses the selection of patient-reported outcomes, safety outcomes, imaging outcomes, and other topics related to hip reconstructive surgery (HRS) including: (*1*) hip replacement systems, (*2*) hip fracture surgery, (*3*) acetabular fracture surgery, (*4*) hip arthroscopy and/or labrum repairs, and (*5*) peri-acetabular osteotomies, or other hip surgeries.

1.2 In this guide, methods to measure the efficacy, effectiveness, and safety of HRS devices through standardizing clinical outcome measures are provided for designing, reviewing, and accepting human clinical trial protocols.

1.3 This guide is intended to provide consistency in study design, review, regulatory approval, and health insurance coverage approval for hip reconstructive surgery to the health care market.

1.4 For the purpose of this guide, HRS pertains to any device or tissue-engineered medical product (TEMP) that is intended to replace, resurface, reconstruct, and/or provide fixation of the hip joint, in part or in total, as a treatment for joint disease, trauma, or dysfunction, where long-term improvement in function and pain relief without major adverse events are the desired outcomes.

1.5 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.6 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

F561 Practice for Retrieval and Analysis of Medical Devices, and Associated Tissues and Fluids

F2809 Terminology Relating to Medical and Surgical Materials and Devices (Withdrawn 2019)[3]

F2979 Guide for Characterization of Wear from the Articulating Surfaces in Retrieved Metal-on-Metal and other Hard-on-Hard Hip Prostheses

2.2 *ISO Standards:*[4]

ISO 12891-1 Retrieval and analysis of surgical implants—Part 1: Retrieval and handling

ISO 12891-2:2014 Retrieval and analysis of surgical implants—Part 2: Analysis of retrieved surgical implants

3. Terminology

3.1 *Definitions:*

3.1.1 *level of evidence, n*—strength of clinical evidence for evidence-based medicine (**1**).[5]

3.1.2 *safety, n*—the condition of being protected from or unlikely to cause risk or injury.

3.2 *Acronyms:*

3.2.1 *AAHKS*—American Association of Hip and Knee Surgeons

3.2.2 *AAOS*—American Academy of Orthopaedic Surgeons

3.2.3 *AJRR*—American Joint Replacement Registry

3.2.4 *ANCHOR*—Academic Network of Conservational Hip Outcomes Research

3.2.5 *ASA*—American Society of Anesthesiologists

3.2.6 *CAT*—Computer Adaptive Testing

[1] This guide is under the jurisdiction of ASTM Committee F04 on Medical and Surgical Materials and Devices and is the direct responsibility of Subcommittee F04.39 on Human Clinical Trials.

Current edition approved June 1, 2020. Published August 2020. DOI: 10.1520/F3448-20.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] The last approved version of this historical standard is referenced on www.astm.org.

[4] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

[5] The boldface numbers in parentheses refer to the list of references at the end of this standard.

3.2.7 *CDRH*—Center for Devices and Radiologic Health

3.2.8 *CMS*—Centers for Medicare and Medicaid Services

3.2.9 *EQ-5D*—European Quality of Life – 5 Domains

3.2.10 *FDA*—Food and Drug Administration

3.2.11 *HHS*—Harris Hip Score

3.2.12 *HOOS*—Hip dysfunction and Osteoarthritis Outcome Score

3.2.13 *HOOS JR*—Hip dysfunction and Osteoarthritis Outcome Score Joint Replacement

3.2.14 *HRQL*—Health-related quality of life

3.2.15 *HRS*—Hip Reconstructive Surgery

3.2.16 *HSAS*—Hip Sports Activity Scale

3.2.17 *ICD*—International Classification of Diseases

3.2.18 *iHOT-12*—international Hip Outcome Tool (12 questions)

3.2.19 *iHOT-33*—international Hip Outcome Tool (33 questions)

3.2.20 *LEAS*—Lower Extremity Activity Scale

3.2.21 *MCID*—Minimal clinically important difference

3.2.22 *MDC*—Minimum detectable change

3.2.23 *MRI*—Magnetic Resonance Imaging

3.2.24 *NPRS*—Numeric Pain Rating Scale

3.2.25 *OHS*—Oxford Hip Score

3.2.26 *PRO*—Patient-reported outcome

3.2.27 *PROMIS*—Patient-Reported Outcomes Measurement Information System

3.2.28 *QALY*—Quality Adjusted Life Year

3.2.29 *RSA*—Radiostereometric analysis

3.2.30 *SAE*—Serious adverse event

3.2.31 *SD*—Standard deviation

3.2.32 *SEM*—Standard error of the measurement

3.2.33 *SF-6D*—Short Form (6 dimensions)

3.2.34 *SF-12*—Short Form (12 questions)

3.2.35 *SF-36*—Short Form (36 questions)

3.2.36 *SMFA*—Short Musculoskeletal Functional Assessment

3.2.37 *TEMP*—Tissue Engineered Medical Products (ASTM Subcommittee F04.40)

3.2.38 *THA*—Total Hip Arthroplasty

3.2.39 *TUG*—Timed Up and Go

3.2.40 *UCLA*—University of California Los Angeles

3.2.41 *VAS*—Visual Analog Scale

3.2.42 *VR-6D*—Veterans Rand (6 dimensions)

3.2.43 *VR-12*—Veterans Rand (12 questions)

3.2.44 *VR-36*—Veterans Rand (36 questions)

3.2.45 *WOMAC*—Western Ontario and McMaster Universities Osteoarthritis Index

4. Summary of Guide

4.1 It is the intent of this guide to provide an overview of appropriate outcomes that are to be addressed in human clinical trials of hip reconstructive surgery (HRS). Depending on the requirements of the clinical trial, the outcomes to be addressed include hip-specific patient-reported outcomes, health-related quality-of-life patient-reported outcomes, activity level scales, pain relief (that is, VAS, NPRS), patient preference data, and adverse events collection and reporting.

4.2 Because of the broad range of indications for HRS, patient comorbidities, and functional/activity levels, it is impossible to identify or specify a single instrument score that measures the "success" of HRS. Instead, a clinically significant improvement (minimum clinically important difference [MCID]) in a joint-specific, disease-specific, or quality-of-life instrument should be used as a measure of clinical "success" (**2**). Clinical success measured with patient-reported outcomes may be defined through clinical improvement in terms of MCIDs and/or achieving a clinical success threshold value defined and justified in the study protocol or literature. The MCID can be calculated using consensus methods (also known as Delphi methods), anchor-based methods, and distribution methods.

4.2.1 Consensus methods use clinical and domain experts to define the MCID (**3**). Anchor-based approaches compare the change in the patient-reported outcome (PRO) score to some other measure of change, considered an anchor or external criterion, to determine whether or not a magnitude of change is significant. The anchor may consist of a clinical measure or a Global Assessment Rating in which the patients rate themselves to some extent as "better," "unchanged," or "worse." Distribution-based approaches compare the change in PRO scores to some measure of variability such as the standard error of measurement (SEM), the standard deviation (SD), the effect size, or the minimum detectable change (MDC) (**4**). Although there is no consensus as to the superior method to determine the MCID, it is recommended that the MCID be based primarily on relevant patient-based and clinical anchors. Distribution-based methods should be used to support the estimates from anchor-based approaches and can be used in situations in which anchor-based estimates are unavailable (**5**). Whenever possible, investigators should use validated scores with established MCID values.

4.3 The application of this guide does not guarantee clinical success of a finished product but will help to ensure consistency and adequacy of the data collected based on the clinical trial protocol.

4.4 The insurance coverage criteria for medical treatments include: (*1*) that a net health outcome is achieved, (*2*) the clinical trial results are applicable (generalizable) to the patient population, and (*3*) the clinical trial results are applicable (generalizable) to medical providers. Therefore, subgroup analyses based on patient characteristics (age, sex) and provider characteristics (academic medical center practice versus community orthopedic practice setting, high versus low surgical volume centers, urban versus rural geographic practice locations) should be included. Financial disclosures of clinical

investigators should be provided based on Code of Federal Regulations Title 21 Part 54 "Financial Disclosure by Clinical Investigators."[6]

4.5 This guide does not suggest that all patient-reported outcome instruments be used for each HRS. However, inclusion of an outcome measure from each section will provide a thorough description of the benefits of HRS, including hip function, pain relief, health-related quality of life including a health utility measure with the ability to calculate Quality Adjusted Life Years (QALYs) (**6**), and mobility/activity level.

5. Significance and Use

5.1 Approximately 500 000 primary total hip arthroplasties (THAs) and 66 000 revision THAs are predicted to be performed in the United States in 2020 (**7**). There are an estimated 340 000 hip fractures per year in the United States (**8**).

6. Use (Outcome Measures)

6.1 *Patient-Reported Outcomes (PROs):*

6.1.1 Patient-reported outcomes (PROs) are vital to understanding the value patients receive from health care. Value can be defined as the change in quality of life and function divided by the total cost of care. Improvement in quality of life is most commonly measured by Quality Adjusted Life Years (QALYs) (**6**). QALYs are required for cost-effectiveness analyses and comparative effectiveness analyses used in coverage decisions. Standardization of PRO measures is necessary to compare outcomes of procedures (**9**). Standardizing PRO measures for implant and outcome registries will make comparative effectiveness data available to the clinical and regulatory communities.

6.1.2 *PRO Measure Selection*—PRO measure selection shall be pragmatic. A high-respondent burden (too many questions) will result in poor rates of patient completion. High licensing fees make it difficult for not-for-profit registries to license the measure. Selection of PRO measures should be based on whether they serve as primary or secondary outcomes in clinical trials, as different PRO measures have strengths and weaknesses.

6.1.3 *Hip-Specific or Disease-Specific Outcome Instruments* (Table 1)*:*

6.1.3.1 *Hip Osteoarthritis/Arthroplasty*—The hip-specific PRO recommendation measure consensus for total hip arthroplasty (THA) from the American Academy of Orthopaedic Surgeons (AAOS), the American Association of Hip and Knee Surgeons (AAHKS), and the American Joint Replacement Registry (AJRR) is the Hip dysfunction and Osteoarthritis Outcome Score JR (HOOS JR) (**10**). The HOOS JR has been validated for total hip replacement surgery (**11, 12**). The AJRR will present national bench marking data for the HOOS JR.

6.1.3.2 The most frequently used PROs for hip replacement surgery are the Oxford Hip Score (OHS) (**13**) and Hip dysfunction and Osteoarthritis Outcome Score (HOOS) (**14**). The OHS is used in the New Zealand Joint Registry (**15**) and the National Joint Registry of England, Wales, and Northern Ireland (**16**).

6.1.3.3 In addition to the HOOS JR, the AJRR accepts the full HOOS (**14**), Oxford Hip Score (OHS) (**13**), Western Ontario and McMaster Universities Arthritis Index (WOMAC) (**17**), and Harris Hip Score (**18**) as Level 3 data on patient-reported outcomes. The Western Ontario McMaster Osteoarthritis Index (WOMAC) (**17**) is a lower extremity osteoarthritis disease-specific outcome instrument used for hip osteoarthritis. A more recent validated PRO for lower extremity osteoarthritis is the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function instrument (**19**).

6.1.3.4 *Hip Fractures*—Because hip arthroplasty (hemiarthroplasty or THA) is the recommended treatment for displaced femoral neck fractures (**20**), any of the hip-specific PROs used for hip osteoarthritis can be used for hip-specific functional outcome measures. However, the Swedish National Hip Fracture Register used only HRQL (EQ-5D) and pain VAS for PROs (**21**). This is likely because most hip fractures occur in the elderly and medical comorbidities and age may have greater impact on physical function than can be measured by a hip-specific PRO.

6.1.3.5 Although not a PRO, the Timed Up and Go (TUG) test assesses ambulatory function (**22**). Because TUG is not a PRO, TUG can be used in patients with cognitive impairment who cannot complete PRO questions.

6.1.3.6 *Acetabular Fractures*—Dodd reviewed patient-reported outcome measures are acetabular fractures and noted measures were not validated for acetabular fractures (**23**). Multiple PROs have been used for acetabular studies including short musculoskeletal function assessment (SMFA) (**24**), Majeed Pelvic Score (**25**), HOOS (**26**), Oxford Hip Score (**27**), and the Harris Hip Score (**28**). Interesting, the Harris Hip Score was originally developed for post-traumatic arthritis after acetabular fracture (**29**).

6.1.3.7 *Hip Arthroscopy/Labrum Repairs*—HOOS has been used for hip arthroscopy, femoroacetabular impingement, and labral repairs (**30, 31**). The international Hip Outcome Tool (iHOT-33) was developed as a patient-reported outcome measure for young, active patients because of the ceiling effects of hip-specific PROs developed for osteoarthritis (**32**). A shorter version (iHOT-12) was also developed (**33**).

6.1.3.8 *Peri-acetabular Osteotomies*—The Academic Network of Conservational Hip Outcomes Research ((ANCHOR) uses the modified Harris Hip Score and WOMAC for hip-specific outcome measures (**34, 35**).

6.1.4 *General Health-Related Quality of Life (HRQL) Outcome Instruments* (Table 1)*:*

6.1.4.1 The AAOS, AAHKS, and Centers for Medicare and Medicaid Services (CMS) have published their consensus recommendations for health-related quality of life PROs for total hip arthroplasty: PROMIS Global Health 10 (**19, 36**) and Veterans Rand-12 (VR-12) (**10**). The AJRR will present national bench marking data for PROMIS Global Health and VR-12. Computer Adaptive Testing (CAT) for PROMIS Global Health can reduce the respondent burden.

[6] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=54

TABLE 1 Recommended Patient Reported Outcome Measures for Hip Reconstructive Surgery

	Health-Related Quality of Life	Hip or Disease Specific	Activity Level	Pain
Hip Arthroplasty	PROMIS Global Health[A]	HOOS[E] JR	LEAS[I]	NPRS[J]
	Veterans Rand 12	HOOS	UCLA	VAS, Likert
	EQ-5D[B]	Oxford Hip Score		
	Veterans Rand 36	Harris Hip Score		
	SF-12[C]	WOMAC[F]		
	SF-36[D]	PROMIS Physical Function		
Hip Fractures	PROMIS Global Health	HOOS JR	LEAS	NPRS
Femoral neck fractures	Veterans Rand 12	HOOS	UCLA	VAS, Likert
Intertrochanteric fractures	EQ-5D	Oxford Hip Score		
Subtrochanteric fracture	Veterans Rand 36	Harris Hip Score		
	SF-12	WOMAC		
	SF-36	PROMIS Physical Function		
Acetabular Fractures	PROMIS Global Health	SMFA[G]	LEAS	NPRS
	Veterans Rand 12	Majeed	UCLA	VAS, Likert
	EQ-5D	HOOS		
		Oxford Hip Score		
		Harris Hip Score		
		WOMAC		
		PROMIS Physical Function		
Hip Arthroscopy/Labrum Repair	PROMIS Global Health	HOOS	Modified Tegner	NPRS
Femoroacetabular impingement	Veterans Rand 12	iHOT-33[H]	LEAS	VAS, Likert
	EQ-5D	iHOT-12		
Peri-articular Osteotomies	PROMIS Global Health	HOOS	LEAS	NPRS
Peri-acetabular osteotomies	Veterans Rand 12	HOOS JR	UCLA	VAS, Likert
Proximal femoral osteotomies	EQ-5D	Oxford Hip Score	Modified Tegner	
		Harris Hip Score		
		WOMAC		
		PROMIS Physical Function		
		iHOT-33		
		iHOT-12		

[A] Patient Reported Outcomes Measurement Information System.
[B] European Quality of Life - 5 Dimensions.
[C] Short Form – 12.
[D] Short Form – 36.
[E] Hip dysfunction and Osteoarthritis Outcome Score.
[F] Western Ontario and McMaster Universities Arthritis Index.
[G] Short Musculoskeletal Functional Assessment.
[H] International Hip Outcome Tool.
[I] Lower Extremity Activity Scale.
[J] Numeric Pain Rating Scale.

6.1.4.2 The European Quality of Life (EQ-5D) is used by the British National Health Service and National Joint Registry of England, Wales, and Northern Ireland to assess the HRQL change after THA (**37**). SF-36 and SF-12 (VR-36 and VR-12) are frequently used as HRQL outcome instruments. However, the quality-of-life summary measure (SF-6D) is a calculated summary score and does not allow patient preference weighting for calculation of change in HRQL. Both the EQ-5D and SF-6D (VR-6D) can be used to calculate QALYs for cost-effectiveness or comparative-effectiveness analyses.

6.1.4.3 EQ-5D appears to be the favored HRQL measure for hip fractures and has been used for femoral neck fractures (**38**), intertrochanteric fractures (**39**), and subtrochanteric fractures (**40**).

6.1.5 *Activity Level Scales* (Table 1)—Activity scales can be used to subdivide patient populations into categories based on their level of activity to allow assessment of patients based on greater or lesser activity level. The Lower Extremity Activity Scale (LEAS) (**41**) is a valid activity scale for hip osteoarthritis and revision hip replacement surgery. The LEAS received positive ratings for reliability and construct validity for hip arthritis (**42**). The UCLA Activity Scale is also used to assess activity level in patients with hip osteoarthritis (**42**). A cross-walk has been developed between LEAS and UCLA (**43**).

6.1.5.1 LEAS and the UCLA Activity Scale have ceiling effects for young, active athletes. Historically, the modified Tegner Activity Scale (**44**) has been used in higher-functioning athletic populations. The Hip Sports Activity Scale (HSAS) can also be used in athletic populations (**45**).

6.1.6 *Pain Relief* (Table 1)—Pain level can be measured with a Numeric Pain Rating Scale (NPRS), Visual Analog Scale (VAS), or Likert scale.

6.2 *Safety:*

6.2.1 Adverse event rates are a measure of safety and should be defined by the study protocol. All adverse events shall be recorded. Adverse events directly related to the HRS or otherwise required by regulatory guidance shall be reported. Time windows for adverse event reporting should be based on regulatory guidance.[7] Adverse event reporting may be reported and analyzed according to both: (*1*) regulatory requirements and (*2*) time windows included in this guide in order to capture all adverse events and determine if different time windows affect adverse event rates.

6.2.2 Adverse event collection, analyses, and reporting protocols should include provision for a priori categorization (including standardized definitions for each type of adverse event) and grading of adverse events (including grading for severity and relatedness to the device and/or surgical procedure). In addition, provision for identification and reporting of serious adverse events should be established in the protocol. An independent Data Safety Review Board should be considered when appropriate. A Clinical Events Committee should be considered when appropriate.

6.2.3 The following clinically expected events should be reported separately as hospital and/or surgeon quality measures:

6.2.3.1 Any infection requiring re-operation within one year of surgery;

6.2.3.2 Symptomatic deep vein thrombosis or pulmonary emboli or both within 90 days of surgery;

6.2.3.3 All-cause non-elective 30-day hospital admission/readmission and cause for admission/readmission (**46**);

6.2.3.4 Intraoperative or postoperative acetabular or femoral fracture occurring within one year of surgery. Time from index surgery and mechanism of injury should be included;

6.2.3.5 *Hip Reoperation/Revision Surgery (no time limit)*—A hip reoperation is defined as any procedure that is performed on the index hip subsequent to the index procedure and the reason(s) for the reoperation should be recorded. A revision surgery is defined as a procedure performed to remove, replace, or reposition any implant inserted at the index procedure, or to repair an intraoperative fracture. Center for Devices and Radiologic Health (CDRH) definitions distinguish reoperations, revisions, removals, and supplemental fixations (**14**).

6.2.3.6 *All Serious Adverse Events (SAEs)*—An event is serious and should be reported to the FDA when the patient outcome is: death, life-threatening, disability or permanent damage, congenital anomaly/birth defect, hospital admission or prolongation of hospitalization, or required intervention to prevent permanent impairment or damage.[8]

6.3 *Imaging Outcomes:*

6.3.1 Radiographic analysis should be conducted. Measurements made on radiographs to determine implant position/migration (**47**) or osteotomy healing (**48**) are standardized in the literature. Radiographs to assess development of hip osteoarthritis after HRS should be evaluated and graded using a classification system such as the Kellgren-Lawrence grading scale (**49**) or other appropriate classification system. Some hip surgeries may not conform to standard radiographic measurement techniques. In such situations, alternative measurement/assessment techniques should be proposed. In either case, "radiographic failure" should be defined and the number of failures should be reported.

6.3.1.1 Radiographs and other imaging data should be independently reviewed according to a pre-specified radiographic analysis plan which includes detailed definitions for relevant imaging parameters, preferentially using a blinded two-reader system, with a third reviewer available to serve as an adjudicator.

6.3.2 Magnetic resonance imaging (MRI) should be used, when appropriate, to evaluate articular cartilage, labrum, capsule, bone lesions, soft-tissue lesions, and pseudotumors.

6.4 *Wear and Other Radiographic Measures*—Radiostereometric analysis (RSA) and/or other radiographic methods may be used for measuring wear and implant stability/migration relative to bone in hip replacement surgery.

6.5 *Retrieval Analysis*—Retrieval analyses should be conducted in compliance with Practice F561 and Guide F2979 and ISO 12891-1 and 12891-2.

6.6 *Data Collection Time Course:*

6.6.1 The study protocol for clinical trials should specify which followup periods will be included in the protocol. The protocol should also specify which clinical outcomes and/or PROs will be collected in each time period. For non-registry study designs, the time windows around the time intervals should be distinct and as small as possible (**50**). Patient-reported outcome measures may be collected electronically and do not require a patient clinic visit.

6.6.2 *Preoperative (within Three Months before Surgery to Two Weeks after Surgery)*—Patient demographics, primary diagnosis, and comorbidities. The two weeks after surgery time extension for pre-operative assessment is for hip or acetabular fracture subjects where pre-operative data cannot be collected prospectively prior to the fracture or treatment. The PROs are collected post-operatively for the patient's pre-operative state.

6.6.3 *Hospitalization/Surgery Center*—Intraoperative data, intraoperative adverse events, perioperative adverse events, and length of stay.

6.6.4 *Post-Operative*—Defined by study protocol.

6.6.4.1 For registries, the time periods should be defined so that all data collected may be analyzed in standardized time period analyses. Table 2 provides an example of follow-up time periods and associated standardized time windows.

6.7 *Number of Subjects:*

6.7.1 Statistical power-based sample size calculations for clinical trials should be based on the MCID of the primary outcome(s) for the population of interest. Adequate statistical power for subgroups is not required. However, adequate power for subgroup(s) is recommended if a definitive conclusion on effectiveness for a specific subgroup will be needed for coverage decisions or other reasons. The MCID may be

[7] Clinical Data Presentation for Orthopedic Device Applications, Food and Drug Administration, December 2, 2004, https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm072283.pdf.

[8] https://www.fda.gov/safety/medwatch/howtoreport/ucm053087.htm.