FDA-Arthritis Foundation OA Workshop

Assessment of Long-term Benefit

Concept Endpoints Informing Design Considerations for Confirmatory Clinical Trials in Osteoarthritis

Yura Kim, PhD
Division of Biometrics III, Office of Biostatistics
Office of Translational Sciences, CDER
U.S. Food and Drug Administration
June 22, 2021

Outline

- Background
- Study Population
- Selected Candidate Outcome Measures
  - Total knee replacement
  - WOMAC scores
- Concept Clinical Endpoints
- Evaluation of Feasibility and Application in Clinical Trials
- Discussion

BACKGROUND

Currently Available OA Treatments

- Currently available approved drugs treat only short-term symptoms of osteoarthritis (OA), primarily pain and function, and do not target the underlying causes or long-term progression of the disease

- There is an unmet need for therapies that target the underlying pathophysiology of osteoarthritis (OA)

Long-term OA Trials

- Utilize an assessment of structural changes in the joint based on imaging as a primary endpoint
  - E.g., x-ray and/or MRI measures of cartilage thickness/catabolism/anabolism, pathological remodeling of subchondral bone, or synovial inflammation

- The ability of treatment effects on common measures of structural progression to reliably predict treatment effects on direct measures of how patients function and feel, has not been established

- A clinically relevant approach to development of such drugs is to utilize a long-term clinical endpoint, i.e., a direct measure of how patients' function, feel, or survive, as the primary endpoint in clinical trials

Total Knee Replacement (TKR)

- Total replacement of the joint, including TKR, has been previously proposed as a primary outcome in OA randomized trials
  - A direct measure of the survival of the joint
  - Involves risk and considerable patient recovery time and effort

- Considerations on using TKR as a primary endpoint
  - Feasibility due to low incidence rate
  - Factors beyond pain and function (e.g., race, gender, socioeconomic status, access to care, surgeon preference, and health care systems)

Project Overview

- Objectives:
  - Identify candidate endpoints for a long-term OA trial that would directly measure how a patient feels, functions, or survives
  - Evaluate the feasibility of such endpoints using data from the Osteoarthritis Initiative

- Outcome: Defined potential endpoints based on total knee replacement (TKR) and composite endpoints defined by TKR and conservative thresholds of patient-reported outcomes (PROs) of pain and function
  - While a single PRO instrument (WOMAC Likert version 3.1) considered here, other appropriate PROs (including other versions of the WOMAC) may be considered using a similar approach. Thus, we refer to these as “concept endpoints”

STUDY POPULATION

Osteoarthritis Initiative (OAI)

- Multi-center, longitudinal, observational study
- Population: 4,796 women and men, ages 45-79
  - Progression Cohort (N=1,390): subjects that had frequent knee symptoms and radiographic evidence of tibiofemoral knee OA
  - Incidence Cohort (N=3,284): subjects that had eligibility risk factors of knee OA, which included knee symptoms, frequent use of medications, being overweight, knee injury/surgery, and family history
  - Healthy Controls (N=122)

Osteoarthritis Initiative (OAI)

- Multi-center, longitudinal, observational study
- Population: 4,796 women and men, ages 45-79
  - Progression Cohort (N=1,390): subjects that had frequent knee symptoms and radiographic evidence of tibiofemoral knee OA
  - Incidence Cohort (N=3,284): subjects that had eligibility risk factors of knee OA, which included knee symptoms, frequent use of medications, being overweight, knee injury/ surgery, and family history
  - Healthy Controls (N=122)
- In the study, we focused on the progression cohort, because this is most representative of subjects who would be expected to be included in clinical trials to evaluate potentially disease-modifying OA drugs

Follow Up Period

- For the analyses, we focused on each subject's initial five years of follow-up data
- Approximately 80% of progression cohort participants had follow-up of at least 60 months

CANDIDATE OUTCOME MEASURES

TKR in the OAI

- In the OAI, the date of joint replacement surgery and the type of surgery (partial or total) were provided

- TKR status was available for each knee

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

- Developed for hip and knee OA
- Three subscales: pain, stiffness, and functional disability
- WOMAC pain and disability have been used as both primary and secondary endpoints in previous OA trials
- Questionnaire items scored on a scale of 0-4: None (0), mild (1), moderate (2), severe (3), and extreme (4)
- Higher scores indicate worse pain, stiffness, and functional limitations

WOMAC Disability

- Consists of 17 questionnaire items, with a possible total score range of 0-68
- Thus, a threshold of 51 (= a score of 3 representing “severe difficulty” × 17 items), represents a subject with severe disability in all categories or extreme disability in some categories

WOMAC Pain

- Consists of 5 questionnaire items, with a possible total score range of 0-20
- Thus, a threshold of 15 (= a score of 3 representing “severe pain” × 5 items), represents a subject with severe pain in all categories or extreme pain in some categories

CONCEPT CLINICAL ENDPOINTS

Concept Clinical Endpoints

1) Time to TKR
2) Time to TKR or severe disability
3) Time to TKR or severe pain
4) Time to TKR or (severe disability in two consecutive visits)
5) Time to TKR or (severe pain in two consecutive visits)
6) Time to TKR or severe disability or severe pain
7) Time to TKR or (severe disability AND severe pain)

While a single PRO instrument (WOMAC Likert version 3.1) considered in our analyses, other appropriate PROs (including other versions of the WOMAC) may be considered using a similar approach. Thus, we refer to these as “concept endpoints”

Choice of Thresholds

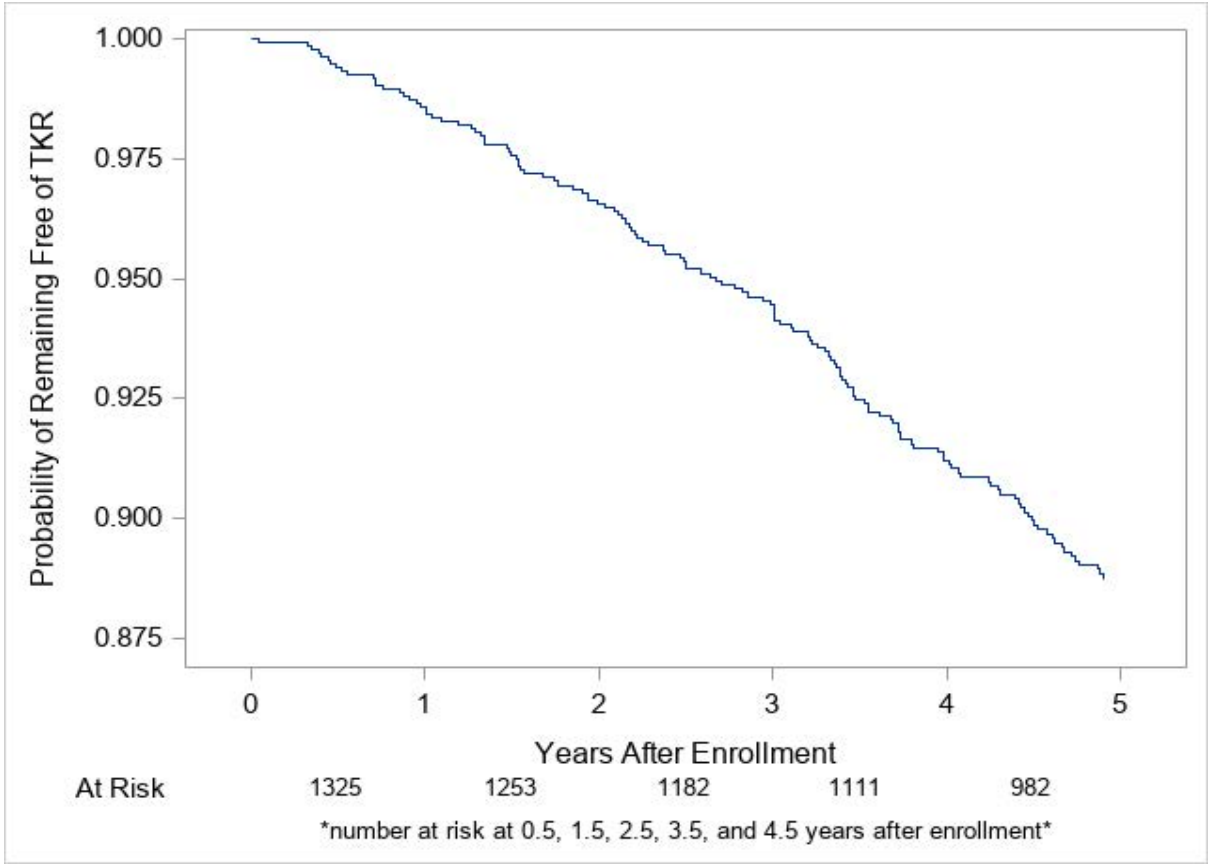
- Thresholds selected to identify severe levels of pain and functional disability, not to predict future knee replacement surgeries
- The thresholds of 51 for WOMAC disability and 15 for WOMAC pain correspond to severe disability and severe pain, respectively
- Endpoint (7) allows for requiring BOTH disability and pain. Because of this combined requirement, we considered a lower threshold for each subscale. The thresholds correspond to half severe and half moderate responses

EVALUATION OF FEASIBILITY

Feasibility Assessment

- Incidence rate of each endpoint-defined event based on five-year follow-up
- Calculated the sample size to detect different magnitudes of treatment effects with 80% power and a 5% two-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group
- A constant hazard rate over time for each group
- Equal average follow-up time for the control group and treatment group

Time to TKR in OAI



Estimated incidence rate: 23.71 cases per 1,000 person-years

Sample Size Calculations (TKR)

Number of Subjects Needed to Evaluate a Treatment Effect on TKR

HR	IR On treatment	Events Required	Sample Size Required (3-year study)	Sample Size Required (4-year study)	Sample Size Required (5-year study)
0.85	20.16	1190	18083	13563	10886
0.75	17.79	380	6105	4579	3675
0.67	15.89	196	3300	2475	1987

Total sample size with 80% power and a 5% two-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group

Abbreviations: HR=hazard ratio, IR=incidence rate

Sample Size Calculation: TKR

HR	IR On treatment	Events Required	Sample Size Required (3-year study)	Sample Size Required (4-year study)	Sample Size Required (5-year study)
0.85	20.16	1190	18083	13563	10886
0.75	17.79	380	6105	4579	3675
0.67	15.89	196	3300	2475	1987

Total sample size with 80% power and a 5% two-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group

Abbreviations: HR=hazard ratio, IR=incidence rate

Sample Size Calculation: TKR

HR	IR On treatment	Events Required	Sample Size Required (3-year study)	Sample Size Required (4-year study)	Sample Size Required (5-year study)
0.85	20.16	1190	18083	13563	10886
0.75	17.79	380	6105	4579	3675
0.67	15.89	196	3300	2475	1987

Total sample size with 80% power and a 5% two-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group

Abbreviations: HR=hazard ratio, IR=incidence rate

Sample Size Calculation: TKR

HR	IR On treatment	Events Required	Sample Size Required (3-year study)	Sample Size Required (4-year study)	Sample Size Required (5-year study)
0.85	20.16	1190	18083	13563	10886
0.75	17.79	380	6105	4579	3675
0.67	15.89	196	3300	2475	1987

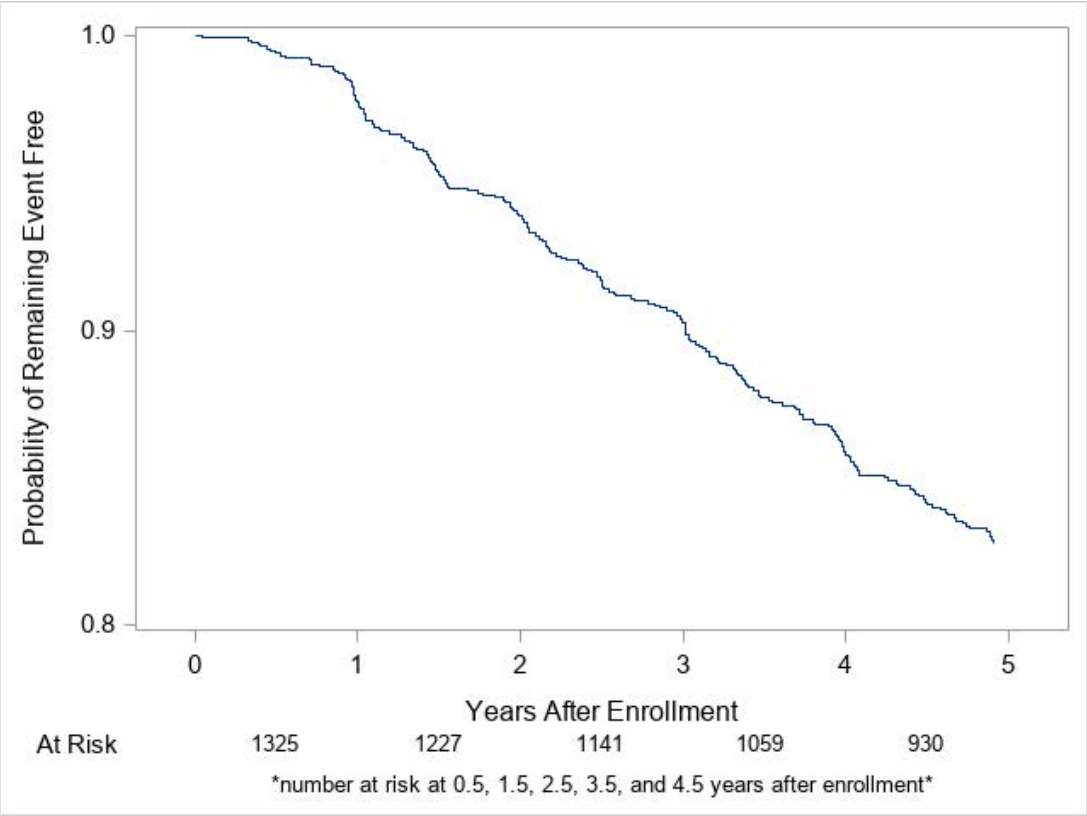
Total sample size with 80% power and a 5% two-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group

Abbreviations: HR=hazard ratio, IR=incidence rate

Observed Incidence Rates in OAI

	Endpoint	Incidence Rates (cases per 1,000 person-years)
1	Time to TKR	23.71
2	Time to TKR or WOMAC disability of ≥ 51	32.59
3	Time to TKR or WOMAC pain of ≥ 15	37.07
4	Time to TKR or WOMAC disability of ≥ 51 in two consecutive visits	25.68
5	Time to TKR or WOMAC pain of ≥ 15 in two consecutive visits	27.34
6	Time to TKR or WOMAC disability of ≥ 51 or WOMAC pain of ≥ 15	38.73
7	Time to TKR or (WOMAC disability of ≥ 43 AND WOMAC pain of ≥ 13)	37.82

Time to TKR or severe disability or severe pain in OAI



Estimated incidence rate: 38.73 cases per 1,000 person-years

Sample Size Calculation: Endpoint (6)

Time to TKR or severe disability or severe pain

HR	IR On treatment	Events Required	Sample Size Required (3-year study)	Sample Size Required (4-year study)	Sample Size Required (5-year study)
0.85	32.92	1190	11073	8305	6666
0.75	29.05	380	3738	2804	2251
0.67	25.95	196	2021	1516	1217

Total sample size with 80% power and a 5% two-sided type 1 error probability, assuming a parallel-group design with 1:1 randomization to the treatment group and control group

Abbreviations: HR=hazard ratio, IR=incidence rate

DISCUSSION

Summary

- A composite clinical endpoint (time to total knee replacement or surpassing a meaningful threshold on patient-reported outcomes of pain or disability) can improve feasibility of clinical trials of products intended to alter the underlying pathophysiology of the disease by increasing the background incidence rate
- This approach directly incorporates pain and function so the patients who did not get a TKR surgery due to factors beyond pain and function (e.g., race, gender, socioeconomic status, access to care, surgeon preference, and health care systems) but have severe pain or disability also can be captured

Discussion

- Analyses based on the simple method used to calculate the sample sizes based on a constant hazard rate. Did not consider an accrual period
- The feasibility of a trial could be further improved by employing enrichment strategies, innovative trial designs, or use of models of accelerated OA, such as post-traumatic OA
- Other appropriate PROs (including other versions of the WOMAC) may be considered using a similar approach
- The choice of thresholds for PROs would require further discussion to ensure the definitions capture chronic, substantial pain and/or disability, but our examples can be considered as proof of concept that such composite endpoints can improve feasibility and capture additional relevant patient outcomes

Acknowledgement

The OAI is a public-private partnership comprised of five contracts funded by the National Institutes of Health. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc.


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Affected Knee at Baseline

- Among the progression cohort participants
  - 474 had OA in the right knee but not the left knee at baseline ('right-knee-affected')
  - 427 had OA in the left knee but not the right knee at baseline ('left-knee-affected')
  - 489 had OA in both knees at baseline ('both-knees-affected')

Follow Up Based on the Knee Affected at Baseline

- For the both-knees-affected participants, we considered the worst outcome for Patient Reported Outcomes (PROs) and the time to first knee (either knee) total knee replacement (TKR)
- For the right-knee-affected/left-knee-affected participants, only measurements on right/left knee were considered
- Depending on the trial design, different definitions of target knees may be more appropriate