Appendix 1: Technical Appendix [posted as supplied by author]

Multicenter Osteoarthritis Study

The MOST cohort includes a community-based sample of 3,026 subjects aged 50 to 79, drawn from the general population and selected for likelihood of either having pre-existing osteoarthritis or of being at high risk, as indicated by weight, knee symptoms, or a history of knee injuries or operations. Baseline exams began in 2003 and included an assessment of risk factors and disease characteristics, knee radiographs and knee MRIs, and a musculoskeletal examination to identify knee symptoms that do not emanate from the knee or hip joint. We selected 2,907 MOST participants aged 50 to 79 without evidence of knee replacement on baseline radiography and complete 30-month examination as validation cohort. 965 of these participants were classified as having knee osteoarthritis and the remaining 1,942 were at risk of developing knee osteoarthritis (see **Table A, appendix 2** for baseline characteristics of the two subgroups). In the MOST sample, medication information was limited to NSAIDs, Cox-2 inhibitors, and analgesics without specifying acetaminophen particularly.

Marginal structural modelling for estimating TKR effectiveness

Marginal structural models were defined as weighted generalized estimating equations (GEE), including a time-varying treatment variable, indicating status after TKR within each 12-month or 24-month time interval. For individuals who underwent TKR, the treatment variable was coded as "1" for all visits following the time that TKR was performed and "0" otherwise. To account for time-varying confounders, we used two logistic regression models to estimate the likelihood of status after TKR being present at the 12, 24, 36, 48, 72, and 96 month visit in a pooled dataset with visit as the unit of measurement ("long dataset format") limited to subjects without a previous TKR. The first pooled logistic regression model included a study visit indicator and baseline variables only, including: the baseline outcome variable, age, gender, race, income, education, knee injury in medical history, knee surgery in medical history, body-mass index, Charlson comorbidity score, use of osteoarthritis pain medication, physician's diagnosis of knee osteoarthritis, Kellgren-Lawrence radiographic grade (maximum of both knees), SF-12 PCS and MCS, total WOMAC score (maximum of both knees), and KOOS quality-of-life. The second logistic regression model additionally included longitudinal variables measured at the previous study visit for: the outcome variable, body-mass index, Charlson comorbidity score, use of osteoarthritis pain medication, doctor's diagnosis of knee osteoarthritis, Kellgren-Lawrence radiographic grade, SF-12 PCS, SF-12 MCS, total WOMAC score, and KOOS quality-of-life. We used the natural logarithm of the total

WOMAC score + 1 to improve model fit. In addition, we included 12-month changes in the outcome variable, use of osteoarthritis pain medication, Kellgren-Lawrence radiographic grade, SF-12 PCS, SF-12 MCS, total WOMAC score, and KOOS quality-of-life, all as observed at the visit preceding the visit of the outcome measurement. With these models we subsequently calculated the probability of having a TKR per visit, conditional on baseline confounders (first model) and baseline plus time-varying confounders (second model), in the 1,327 subjects with knee osteoarthritis. These probabilities were carried forward to later visits in subjects who had undergone TKR. The GEE models to estimate TKR effects included only the aforementioned baseline variables and an indicator for TKR, and were weighted with treatment propensity defined as the product of the probability estimated by the first logistic model and the inverse of the probability estimated by the second model ("stabilized weights").[1 2] Missing predictor and outcome values were imputed 20 times with a flexible additive model including status variables and the Nelson–Aalen estimator of the cumulative hazard for TKR and death. To estimate parameter uncertainty, imputation, pooled logistic regression and GEE models were re-fitted in 500 bootstrap datasets.

Development of the KOSMOS (Knee Osteoarthritis Microsimulation) model

The KOSMOS (Knee OSteoarthritis MicrOSimulation) model was developed using TreeAge Pro 2015 (TreeAge Software. 2015. TreeAge Pro 2015. Williamstown, MA: TreeAge Software, Inc.) and includes four health states: 1) *Alive without TKR*; 2) *After TKR*; 3) *After revision TKR*; and 4) *Dead*. A one-year cycle length was implemented. One-year rates for transitions to the *After TKR* and to the *Dead* states were based on survival analysis of 9-year follow-up data of the total study population of 4,498 OAI participants. Cumulative hazards from age 45 until age 85 were individualized using two multivariable cause-specific Cox regression models with chronological age as the time scale and baseline hazard functions estimated with the Nelson-Aalen method (using the *R coxph* function within the survival package in R). We considered the following candidate predictors for TKR risk: gender, race, education, body-mass index, history of knee injury, history of knee surgery, natural logarithm of WOMAC pain score + 1, and Kellgren-Lawrence radiographic grades. Furthermore we adjusted for status of knee osteoarthritis vs. high-risk for knee osteoarthritis at baseline in order to predict outcomes for those with knee osteoarthritis at baseline. Continuous predictors were "winsorized" using the 1st and 99th percentiles to limit influence of outliers. We used a backward selection approach using the Akaike Information Criterion (AIC) as selection criterion. For modelling death, we predefined gender, race, education, body-mass index, and body-mass index squared as the predictors.

Baseline cumulative hazard functions were smoothed and extrapolated beyond age 85 by restricted cubic spline functions using four knots with the 5th, 35th, 65th and 95th percentiles as locations. To take into account parameter uncertainty, 500 different TKR effects, Cox models, and smoothed baseline cumulative hazard functions were estimated in bootstrap datasets stratified for status of knee osteoarthritis at baseline. Single imputation with a multivariable algorithm as described in the manuscript within each bootstrap was performed to handle missing predictors.

Rates for the transition from *After TKR* to *After revision TKR* were based on long-term age-specific TKR revision data from the United Kingdom's National Joint Registry.[3] We estimated age-specific cumulative hazards at the log scale as a function of log time in years since TKR for prediction and extrapolation using simple linear regression.

Because only one transition is possible within every cycle, we took into account competing risks by assuming constant 1-year hazard rates H for primary and revision TKRs and competing death within each model cycle. The one-year probability of TKR was calculated using the following formula:

$$P\,TKR_{Age_j} = \frac{H\,TKR_{Age_j}}{H\,TKR_{Age_j} + H\,death_{Age_j}} \times \left(1 - e^{-\left(H\,TKR_{Age_j} + H\,death_{Age_j}\right)}\right)$$

with individualizing 1-yr cumulative hazard rates H for each Age_j using the smoothed baseline hazard H_0 and the linear predictor $\beta[X]$ for primary TKR:

$$H_{Age_j} = H_{0\,Age_j} \times e^{(\beta[X])}$$

For revision TKR, we tracked the age at and time since primary TKR and subsequently calculated the applicable one-year rates using the previously mentioned linear regression equation that calculated the log cumulative hazard rate as a function of log time since and of age at primary TKR. The likelihood of a bilateral primary TKR was defined as a second TKR performed in the contralateral knee within 1 year of the first procedure and estimated using OAI data.

Validation of the KOSMOS model

We compared the predicted total life expectancy for OAI and MOST study populations with the total life expectancy estimated by age and gender specific U.S. 2011 Life Tables through age 100 and recalibrated the baseline mortality hazard to match average life expectancy. Subsequently, predictions of TKR by the KOSMOS model were acquired for both OAI and MOST individuals. Statistical significance of any difference in hazards ratios across OAI and MOST development cohorts were calculated by using the linear predictor based on OAI data as offset variable while additionally including predictor variables.[4] We assessed the validity of 9-year and 2-year prediction of any TKR by KOSMOS in individuals with knee osteoarthritis at baseline in OAI (N=1327) and MOST (N=965) respectively by overall calibration and the Harrell's concordance statistic (C-statistic) and. To calculate Harrell's C-statistics, predictions of TKR were acquired by simulation of the 1327 OAI and 965 MOST patients with aggregating TKR results at the individual level. Parameter uncertainty was addressed by simulation of patients for each set of bootstrap equations.

After recalibration of the KOSMOS model's baseline mortality rates, total life expectancy of OAI patients equalled estimations by U.S. 2011 Life Tables: 22.39 (95% UI 21.13 to 23.85) year by KOSMOS vs. 22.46 year by U.S. 2011 Life Tables. For MOST patients, total life expectancy predicted by KOSMOS was 20.94 (95% UI 19.47 to 22.87) vs. 20.92 year by U.S. 2011 Life Tables. Hazard ratios (HRs) for the included predictors were very similar in OAI and MOST, with no statistically significant differences (**table B in appendix 2**). Predictions of TKR by KOSMOS calibrated reasonably well with observed 9-year cumulative TKR incidence in OAI individuals with knee osteoarthritis (N=1327): 19.5% (95% UI 16.7 to 22.4) vs. 21.5% (95% CI 19.1 to 23.9), but underestimated 2-year cumulative TKR incidence in MOST individuals (N=965) with knee osteoarthritis: 7.5% (95% UI 5.6 to 9.2) predicted vs. 11.1 (95% CI 9.2 to 13.2). Therefore we recalibrated the baseline cumulative hazard function for MOST by a hazard ratio of 1.55, and obtained a 2-yr predicted TKR risk of 11.1 (95% UI 8.9 to 13.4). KOSMOS predictions of any TKR in the knee osteoarthritis population demonstrated C-statistics of 0.77 (95% CI 0.74 to 0.80) in 9-year OAI and 0.68 (95% CI 0.64 to 0.73) in 2-year MOST follow-up data.

Model inputs for annual costs

For modelling pharmacological treatment we used annual drug prices as reported by Losina et al.[5] The annual cost of prescription non-steroidal anti-inflammatory drugs (NSAIDs) was calculated using annual cost estimates of five different drug types weighted for their reported prevalence of use: diclofenac

(11.7%); ibuprofen (9.4%); meloxicam (30.5%); nabumetone (4.7%); and naproxen (13.3%). The annual cost of non-prescription NSAIDs was estimated based on the use of over-the-counter ibuprofen and naproxen with a fifty-fifty ratio. Annual costs of celecoxib and acetominophen use were modelled separately. Subsequently, these costs were applied to simulated OAI participants if they were modelled to be using osteoarthritis pain medication assuming 18.6% would be using prescription NSAIDs, 55.4% non-prescription NSAIDs, 22.2% celecoxib, and 31.2% acetominophen. These percentages were based on the reported use among those OAI participants with knee osteoarthritis at baseline. For non-pharmacological treatment, we considered acupuncture, chiropractic, massage, and other types. The cost per visit for each option was estimated from Gore et al.,[6] by dividing the annual reported costs by the reported average number of visits annually. As cost input for overall use of non-pharmacological treatment, we used the sum of the products of cost per visit, proportion of use, and number of annual visits per treatment option within OAI participants with knee osteoarthritis and non-pharmacological treatment use at baseline. Proportions of use and number of annual visits were directly estimated with OAI data. Other annual costs associated with the care for knee osteoarthritis consisted of one physician office visit and a knee X-ray. Cost inputs for these services were adopted from Losina et al.[5]

Model inputs for costs associated with TKR procedures

Costs for hospital stays with TKR and revision TKR as primary reason, were estimated from charges using hospital-level cost-to-charge ratios as reported by the HCUP database. Because these costs do not include physician fees, we included fees for a surgeon and anaesthesiologist as reported by Losina et al. Costs associated with post-surgical rehabilitation were assumed to be similar for primary and revision TKR and were based on the calculations as provided by Losina et al.[5]

Sensitivity analyses

For estimation of the effect of TKR on productivity loss, we performed marginal structural modelling using information on self-reported missed work days (paid or unpaid) for a subset of 819 OAI participants with knee osteoarthritis who reported to be employed. In these models, the outcome was the absence and number of work days missed in the last 3 months based on questionnaires assessed at 12, 24, 36, 48, 72, and 96 months. We calculated the adjusted marginal difference in work days missed by the recycled prediction method.[7] We repeated the analysis with a marginal structural model omitting data reported the first visit after TKR to evaluate the impact of rehabilitation on missed work days.

For the additional cost-effectiveness analysis, we used the results of the first marginal structural model and assumed the estimated difference in work days missed resembles a short-term period of additional absenteeism due to rehabilitation. We furthermore assumed that the costs associated with this short-term period would be equal to foregone earnings and that these foregone earnings are not well captured by a decrease in health-related utility. Therefore, when TKR was performed in simulated employed patients younger than 65, we applied a one-time increase in the accrued costs. This cost penalty was equal to the estimated additional amount of work days missed across four years, the average self-reported number of hours worked per day (as available in the OAI) and the 2013 national average hourly wage of $20.47.[8] Four years was chosen based on the effect duration within the marginal structural modelling, i.e. the difference in the maximum follow-up time and median time to TKR in employed OAI participants who had a TKR.

The number of days missed from work within the last 3 months increased with TKR in patients with knee osteoarthritis who reported to be employed in the OAI cohort: 2.92 (95% UI 0.75 to 7.25) work days missed in the last 3 months, although this result was sensitive to omitting the first questionnaire after TKR (**see table D in appendix 2**). By evaluating the trends in work days missed up to the time of TKR in the OAI participants who received TKR, we could however not rule out a co-existing beneficial effect of TKR on work days missed beyond the study's follow-up period. We estimated this beneficial effect to comprise a decrease of up to 2 fewer work days missed per 3 months. We therefore additionally modelled a long-term TKR benefit ranging from 0 to 8 fewer work days missed per year starting from four years after TKR. We used a uniform distribution for modelling uncertainty of the number of days. When a revision TKR was performed, only the cost penalty for short-term productivity loss was applied, and any long-term cost savings due to the primary TKR were set to zero. For the costs of informal caregiving, we assumed these would comprise 52% (95% CI 49 to 55) of the total costs due to productivity loss of the patient,[9] and we used a beta distribution for the parameter uncertainty of this proportion.

To evaluate the influence of number of bootstraps on the stability of our cost-effectiveness results, we performed cost-effectiveness analyses with sampling from fewer bootstrap datasets: N=100, N=200, N=300, and N=400. The results from these analyses demonstrated that incremental cost-effectiveness outcomes and rankings were reasonably stable across the different sets of bootstraps (**table J in appendix 2**).

Appendix References

1. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;**11**(5):561-70.
2. Hernan MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. Stat Med 2002;**21**(12):1689-709.
3. 11th Annual Report: National Joint Registry for England, Wales, and Northern Ireland 2014.
4. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;**13**:33.
5. Losina E, Paltiel AD, Weinstein AM, et al. Lifetime medical costs of knee osteoarthritis management in the United States: impact of extending indications for total knee arthroplasty. Arthritis care & research 2015;**67**(2):203-15.
6. Gore M, Tai KS, Sadosky A, et al. Use and costs of prescription medications and alternative treatments in patients with osteoarthritis and chronic low back pain in community-based settings. Pain practice : the official journal of World Institute of Pain 2012;**12**(7):550-60.
7. Graubard BI, Korn EL. Predictive margins with survey data. Biometrics 1999;**55**(2):652-9.
8. Employer costs for employee compensation – June 2013: Bureau of Labor Statistics U.S. Department of Labor, 2013.
9. Gupta S, Hawker GA, Laporte A, et al. The economic burden of disabling hip and knee osteoarthritis (OA) from the perspective of individuals living with this condition. Rheumatology 2005;**44**(12):1531-7.