## Appendix D. The Process for Creating Predictive Models for the Pain and Function Outcomes

# APPENDIX D: The process for creating predictive models for the pain and function outcomes

*Model Development* Initial variable selection was done (development and updating phases) using a stepwise selection process in one dataset with one observation per knee created from averaging 10 multiply-imputed copies of the OAI database (and later the pooled OAI and MOST databases). The candidate variables and interactions included in the selection process were chosen from those available in the OAI and MOST database and that stakeholders and the project team considered important and plausible. The selected variables then were used to make a separate model for each of the individual imputed datasets, and we combined the results from these 10 models to get the parameter estimates and associated standard errors that accounted for both variation in the data and the amount of missingness. If variables in the model were no longer significant at the $p < 0.10$ level, then they were removed using a backward selection process.

*Model Validation* Performance of the linear regression models derived on the OAI matched database was tested on the MOST dataset. We looked at scatter plots of predicted outcomes (based on the equation from the OAI model) versus true 1-year outcomes in the MOST database, and also *r*-square values. We did this for the pooled data and stratified by treatment status.

*Model Updating* The MOST data and OAI databases were then pooled together (10 imputed copies of each) and the beta coefficients from the validated model were re-estimated on the larger pooled database. We then re-explored adding additional interactions of variables with the treatment indicator variable and removing variables based on significance, clinical, or pragmatic reasons. These final models were called the P1 and F1 models for pain and function.

*Re-derivation* A project objective was to make patient-specific predictions accounting for their individual characteristics. For a statistical model, this implies interactions of variables representing patient characteristics with those representing the treatment. To better screen for

interactions, we used a larger dataset that included OAI and MOST data, and also the NEBH and TMC matched datasets. One tradeoff with using these latter data sources is that they had fewer candidate variables to draw upon. The model development process we used for the pooled data from the four data sources was similar to that used for the model development. The final models created from the pooled data from the four data sources were called the P2 and F2 models.