APPENDIX B: Missing Data

We addressed the challenge of missing data by using a multiple imputation methodology and by restricting the number of variables examined. For the multiple imputation, we used the SAS procedures PROC MI to do the imputation and PROC MIANALYSE to combine results from analyses run on the different imputed datasets.[41] For each matched dataset (OAI, MOST, NEBH, TMC) we created ten imputed datasets. We imputed main effects only and later calculated interactions from the imputed data for the component main effects. We did discuss the alternative of creating interactions first, then imputing data for the missing interactions. We opted against that approach since we were concerned we could be generating main effects and interactions at a subject level that did not correspond to each other. We used available data, including 1-year outcome data, for each dataset when creating the imputed data. This means that the variables used for the imputation for study differed based on availability, but maximized the information we had for each study.