

## Multimedia Appendix 1: The osteoarthritis initiative (OAI) dataset

The OAI dataset consist of historical clinical evaluation data, radiological images and a biospecimen repository from 4796 subjects. The subjects were enrolled between February 2004 and May 2006 and consisted of roughly equal number of men and women, aged between 45 – 79 years. Of these 2,712 were invited to participate in a physical activity study in the fourth year of the study and 2127 consented. 2001 subjects contributed activity data from one or more valid monitoring days.

**Progression subcohort:** Subjects with at least one symptomatic tibiofemoral knee (pain, aching or stiffness on most days for at least a month and a Kellgren & Lawrence score  $\geq 2$ ) at baseline were eligible for the Progression subcohort.

**Incidence subcohort:** Incident symptomatic knee OA was defined as absence of the symptoms as defined for the progression cohort above, “frequent” knee symptoms and definite tibiofemoral osteophytes in the symptomatic knee.

Table 1 gives key demographic attributes of the subjects that participated in the physical activity study.

Table 1: Key Demographic attributes

Cohort size	2001
Incidence subcohort	1490
Progression subcohort	505
Control subcohort	6
Elective knee replacement surgery before date of study	63
Age (mean, sd)	(65.08, 9.09)
Gender (% male)	44.53
BMI (mean, sd)	(28.52, 4.87)
Mean Comorbidity Index (Modified Charleston Deyo)	0.52
Median physical activity days	7

**Baseline physical capacity measurements.** Shown below are the baseline distributions of the 400m walk test, the 20m walk test (pace) and the 5 sit-stand test measurements on the subjects in the training partition.



