Alzheimer’s disease is a neurodegenerative illness which follows a devastating path over several decades, eventually leading to a loss of large volumes of cortex with commensurate deterioration of the ability to function independently and communicate. Since early phases of the disease are to a large extent asymptomatic, it is difficult to monitor early disease processes when perhaps the pathophysiology could be altered. It is also difficult to predict later milestones in progression. However, since so many processes (neurological and otherwise) are altered, due both to the underlying pathophysiology and the impact of disease progression, one might expect that many biochemical and neuroanatomical changes could be detected and used as biomarkers. Unfortunately, no single biomarker seems to signal the likelihood of Alzheimer’s to progress from asymptomatic (or lightly symptomatic) to sever.

Due to the presence in the Alzheimer’s Disease Neuroimaging Inititative (ADNI) database of a comprehensive list of biomarker data, with many hundreds of plasma, CSF, MRI, PET, and cognitive performance values, we tested the plausibility of incorporating multi modal data from hundreds of patients to assess relative risk of progression across a well populated milestone (progression from MCI to AD diagnosis). Because we suspected the relationships might be highly non-linear, we used machine learning techniques, including categorical and regression tree and random forest approaches. We have identified a particular variant of the analysis which gave good results, improving as we incorporated additional modalities. Surprisingly, our approach worked well even when focused only on inexpensive and non-invasive data like plasma biomarkers and cognitive test results, thus suggesting the possibility of population-level testing.