## Supporting Information

### Appendix S1. Cohort Discovery Tool Rules

| Component | Component Elements |
|---|---|
| ICD Diagnoses for AD | 331.0 (ICD9), G30 (ICD10) |
| Medications for AD | Donepezil, Memantine, Galantamine, Rivastigmine |
| Psychological Testing | CPT codes from 96101 through 96127 |
| Complex Medical Decisions | CPT codes: 99214, 99215, 99354, 99355 |

**Table S1: The components considered for cohort the discovery rules.** Shown in the left column are the components we considered when formulating our rules, and shown in the right column are the specific elements we used to define the component. Each rule labeled patients experiencing one or more of these components as probable AD. Complexity in medical decisions was measured by the amount and variety of patient data examined by a physician, patient risk, and treatment options, as defined by the description of the CPT codes listed. Although the ICD10 code ‘F00’ is also used to denote an AD diagnosis, our EHR did not use this code, so we did not include it.

**Appendix S2. Population Adjusted PPV (Cohort Discovery)**

Since the proportion of probable AD individuals in the Michigan-ADRC dataset (~25%) is enriched compared to the general population 65 and older (10%)[1], unadjusted PPVs calculated directly from the Michigan-ADRC and RDW overlap are likely to be different from the general population. For this section, the general population of interest refers to those 65 and over.

All counts for true and false positives and negatives were taken from the overlap between the patients identified by the cohort discovery rule and Michigan-ADRC data, using the Michigan-ADRC diagnoses. The population adjusted PPV for each cohort discovery rule was calculated as

$$\frac{TP}{TP+FP_a(CN)+FP_a(MCI)}$$

where TP represents the number of true positives, $FP_a$(CN) represents the population adjusted number of cognitively normal false positives, and $FP_a$(MCI) represents the population adjusted number of MCI (mild cognitive impairment) false positives. $FP_a$(X), where X is CN or MCI, was calculated as

$$FP_a(X)=\frac{FP(X)}{n(X)}\times P\times C(X)$$

where FP(X) represents the unadjusted number of false positives identified with diagnosis X, n(X) is the number of patients with diagnosis X, P is the number of positive examples, and C(X) represents the proportion of the general population with diagnosis X divided by the proportion of the general population with AD.

The product of P and C(X) can be seen as the number of patients with diagnosis x that would have been found in a cohort representative of the general population. Multiplying that by the detection rate of diagnosis X within the cohort then gives the estimated number of false positives with diagnosis X. It is assumed that the detection rate of diagnosis x within the cohort holds for the general population. 10% was used as the percentage of the population with AD, and 15% was used as the percentage of the population with MCI[1].

**Appendix S3. Model Calibration (Risk Stratification)**

The following describes how the calibration curve was constructed. We split the held-out test set in half, using the first half (calibration set) to construct the calibration curve and to learn a recalibration function and the second half (calibration test set) to evaluate the recalibration function. On the calibration set, patients were split into five equally spaced bins according to their predicted probability of conversion. For each bin, we calculated the empirical conversion probability. The calibration curve was constructed by plotting the median values, with error bars representing empirical 95% confidence intervals. Using the calibration set, we recalibrated the predictions by fitting a cubic curve to these data. This learned relationship was then applied to the held-out calibration test data.

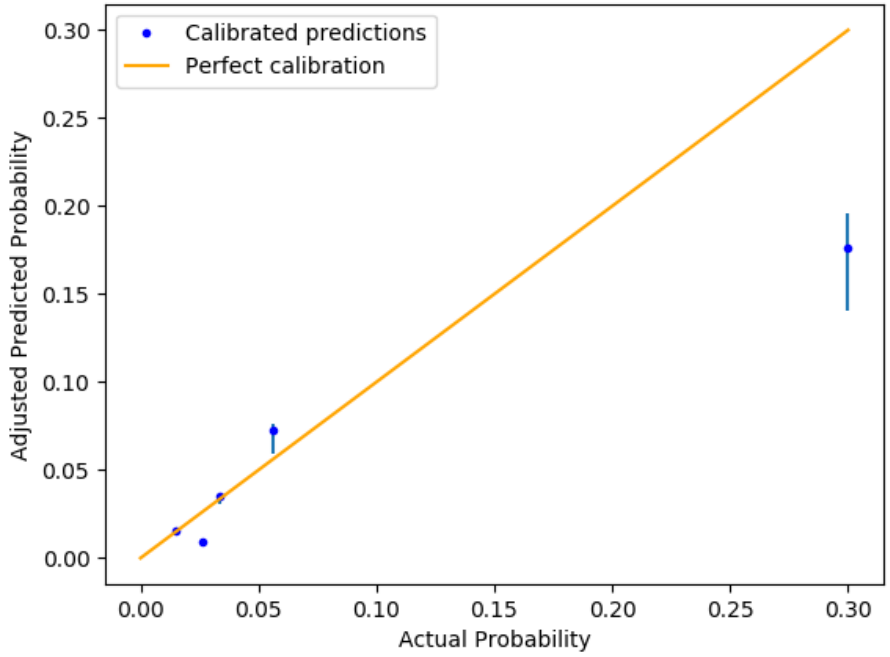


**Figure S1: The calibration curve.** Predicted probability refers to the probability of converting to probable AD within 10 years as given by the model. Actual probability refers to the proportion of patients who actually convert to probable AD within the next 10 years. Results shown are for a held-out test set. Points represent the median over 1,000 bootstrapped samples, and error bars represent 95% confidence intervals.

**Appendix S4. Cohort Discovery – Cohort Characteristics**

| Description | Study Population (N=624) |
|---|---|
| Percent < 73 | 50.0% |
| Percent 73-79 | 25.0% |
| Percent > 79 | 25.0% |
| Percent female | 60.6% |
| Percent probable AD | 24.8% |
| Years covered | 2005-2019 |

**Table S2: Study population characteristics.** This corresponds to the Michigan-ADRC and RDW overlap used in evaluating the cohort discovery tool.

**Appendix S5. Encounter Level Results (Cohort Discovery)**

| | Predicted Not Probable AD | Predicted Probable AD |
|---|---|---|
| **Actually Not Probable AD** | 29,185.50 (27,027.83-31,757.25) | 4,064.00 (2,324.18-6,076.40) |
| **Actually Probable AD** | 1,224.00 (901.90-1,641.13) | 5,890.00 (4,107.90-7,825.03) |

**Table S3: Encounter Level Confusion Matrix.** This measured the cohort discovery tool's ability to identify probable AD at the encounter level or the cohort discovery tool with the highest F1 score (identification by ICD diagnosis code only). The median over 1,000 bootstrapped samples is given, with a 95% confidence interval in parentheses.

**Appendix S6. Feature Breakdown (Risk Stratification)**

| Feature Category | Time Invariant | Number of Features |
|---|---|---|
| Demographics | Yes | 268 |
| Diagnoses | No | 1,304 |
| Procedures | | 1,232 |
| Laboratory test results | | 1,111 |
| Medications | | 250 |
| Vital signs | | 45 |
| Healthcare utilization | | 21 |

**Table S4: Features for Risk Stratification.** The overall breakdown for the features used in the predictive model is shown in Table S2. Time invariant features occurred once in a patient feature vector. Time variant features occurred four times in a patient feature vector, once for each 250 day time window between 1,000 days prior to alignment and alignment.

**Appendix S7. Discriminative Performance (Risk Stratification)**

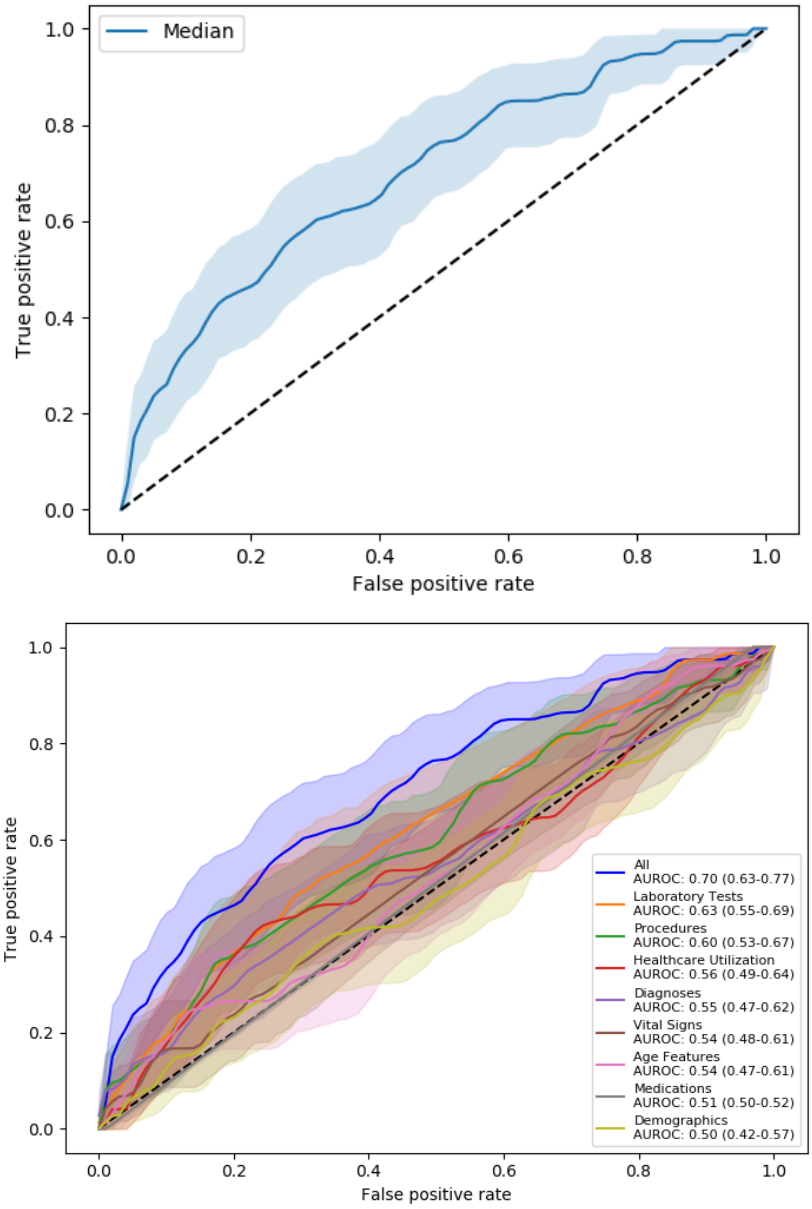


**Figure S2. (a) The median ROC curve.** The ROC for the model plotted alongside the ROC curve when predictions are made at random. Points to the left of the random curve are considered to be an improvement over random predictions. The shaded area around the median ROC curve represents the 95% confidence interval.

**(b) ROC curves for different EHR data components.** The ROC curves and 95% confidence intervals for each EHR component is shown. The dashed black line represent random predictions.

| | Predicted Not Probable AD | Predicted Probable AD |
|---|---|---|
| **Actually Not Probable AD** | 1075 (1063-1084) | 550 (538-563) |
| **Actually Probable AD** | 27 (18-38) | 44 (31-57) |

**Table S5: Confusion matrix for test set performance.** This measured the predictive model's ability to identify probable AD among patients in the test set. The median over 1,000 bootstrapped samples is given, with a 95% confidence interval in parentheses.

**Appendix S8. Prediction Sensitivity (Risk Stratification)**

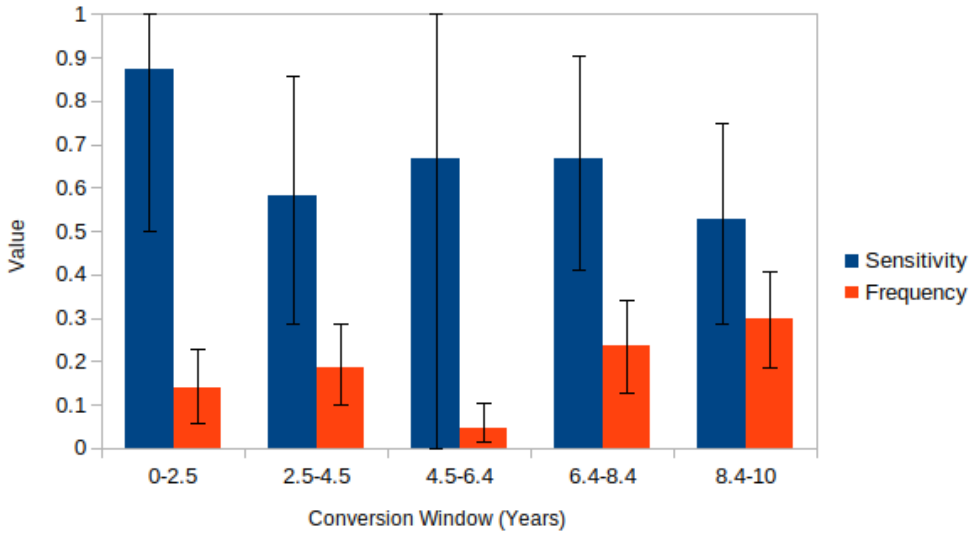


**Figure S3: Analysis of AD Converters.** Among those identified to convert to probable AD, we analyzed the model's ability to predict conversion relative to amount of time it took to convert starting from alignment. Our model was able to predict conversion on large time windows as well as small ones. Error bars represent 95% confidence intervals over 1,000 bootstrapped samples.

**Appendix S9. Discriminative Performance Revisited (Risk Stratification)**

In this section, we examined the model's ability to distinguish between those from the test set who had memory impairments and develop AD within 10 years of alignment (N=2) and those from the test set with memory impairments, but do not convert to AD within 10 years (N=18). We considered those with an ICD diagnosis code of 780.93, 310.89, 784.69, 315.31, 300.12, 437.7, 331.83, and R41.2 as memory impaired. R41.2 is an ICD10 code while the rest are ICD9 codes. When classifying patients who convert to AD among patients with memory impairments (N=20), we achieved an AUROC of 0.78 (95% CI=0.44-1.00) (Figure A.4, Table A.6).

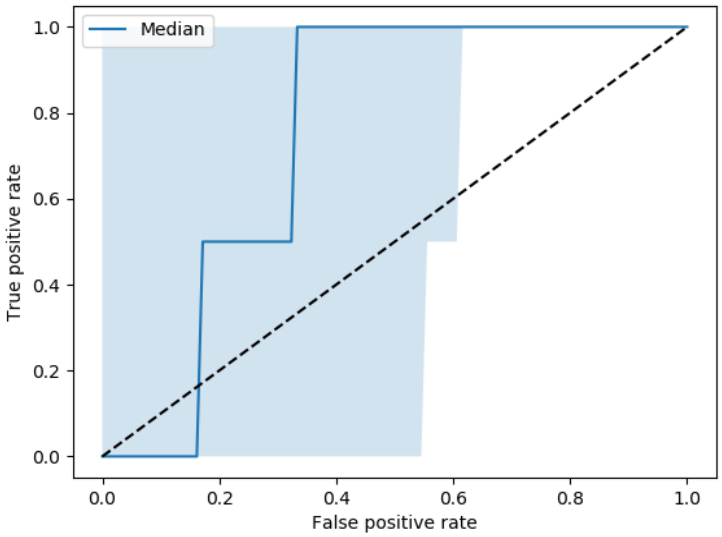


**Figure S4: The median ROC curve.** The ROC for the model plotted alongside the ROC curve when predictions are made at random (dotted black line).

| | Predicted Not Probable AD | Predicted Probable AD |
|---|---|---|
| **Actually Not Probable AD** | 12 (10-13) | 6 (5-8) |
| **Actually Probable AD** | 1 (0-2) | 1 (0-2) |

**Table S6: Confusion Matrix.** This measured the predictive model's ability to identify probable AD among patients in the test set who had memory complaints. The median over 1,000 bootstrapped samples is given, with a 95% confidence interval in parentheses.