

Supplementary Information

Table SI-I: ICD10 codes defining Alzheimer's Disease and Related Dementia (ADRD) cases

Code	Description
F00 F00.0 F00.1 F00.2 F00.9	Dementia in Alzheimer's disease
F01 F01.5 F01.50 F01.51	Vascular dementia
F03 F03.9 F03.90 F03.91	Unspecified dementia
G30 G30.0 G30.1 G30.8 G30.9	Alzheimer's disease
G31.0 G31.01	Pick's disease
G31.09	Other frontotemporal dementia
G31.83	Dementia with Lewy bodies

Table SI-II: **Target ADRD-related Prescription Drugs** Aducanumab (approved in 2021) was discontinued in 2024 "to reprioritize Biogen's resources in Alzheimer's disease", and not for reasons related to safety or efficacy[?]. We retained Aducanumab in the case definition as it was prescribed to patients during the period covered by the datasets we used in this study. Donanemab was approved in July 2024 as an anti-amyloid monoclonal antibody, but was not recoded in the used data and was not included into the case definition for this study.

Name	Year of FDA Approval	Disease Stage Targeted	Mechanism of Action	Market Status
Donepezil	1996 [?]	Mild to Severe Alzheimer's Disease	Acetylcholinesterase Inhibitor	Marketed
Galantamine	2001 [?]	Mild to Moderate Alzheimer's Disease	Acetylcholinesterase Inhibitor	Marketed
Rivastigmine	2000 [?]	Mild to Severe Alzheimer's Disease	Acetylcholinesterase and Butyrylcholinesterase Inhibitor	Marketed
Memantine	2003 [?]	Moderate to Severe Alzheimer's Disease	NMDA Receptor Antagonist	Marketed
Aducanumab	2021 [?]	Mild Cognitive Impairment or Mild Dementia due to AD	Monoclonal Antibody against Amyloid Beta	Discontinued (2024) [?]
Lecanemab	2023 [?]	Mild Cognitive Impairment or Mild Dementia due to AD	Monoclonal Antibody against Amyloid Beta	Marketed

Table SI-III: **Control subcohort exclusion conditions associated with or causally related to dementia.**

Condition	ICD Codes
Huntington's disease	G10
Parkinson's disease	G20
Secondary parkinsonism	G21 G21.0 G21.1 G21.11 G21.19 G21.2 G21.3 G21.4 G21.8 G21.9
Human immunodeficiency virus [HIV] disease	B20
Alcohol related disorders	F10 F10.1 F10.10 F10.11 F10.12 F10.120 F10.121 F10.129 F10.13 F10.130 F10.131 F10.132 F10.139 F10.14 F10.15 F10.150 F10.151 F10.159 F10.18 F10.180 F10.181 F10.182 F10.188 F10.19 F10.2 F10.20 F10.21 F10.22 F10.220 F10.221 F10.229 F10.23 F10.230 F10.231 F10.232 F10.239 F10.24 F10.25 F10.250 F10.251 F10.259 F10.26 F10.27 F10.28 F10.280 F10.281 F10.282 F10.288 F10.29

Table SI-IV: ICD10 Codes for Most Common ADRD Risk Factors

Risk Factor	ICD Codes
Type 2 Diabetes Mellitus	E11 E11.0 E11.00 E11.01 E11.1 E11.10 E11.11 E11.2 E11.21 E11.22 E11.29 E11.3 E11.31 E11.311 E11.319 E11.32 E11.3211 E11.3212 E11.3213 E11.3219 E11.3291 E11.3292 E11.3293 E11.3299 E11.33 E11.3311 E11.3312 E11.3313 E11.3319 E11.3391 E11.3392 E11.3393 E11.3399 E11.34 E11.3411 E11.3412 E11.3413 E11.3419 E11.3491 E11.3492 E11.3493 E11.3499 E11.35 E11.3511 E11.3512 E11.3513 E11.3519 E11.3521 E11.3522 E11.3523 E11.3529 E11.3531 E11.3532 E11.3533 E11.3539 E11.3541 E11.3542 E11.3543 E11.3549 E11.3551 E11.3552 E11.3553 E11.3559 E11.3591 E11.3592 E11.3593 E11.3599 E11.36 E11.37 E11.37X1 E11.37X2 E11.37X3 E11.37X9 E11.39 E11.4 E11.40 E11.41 E11.42 E11.43 E11.44 E11.49 E11.5 E11.51 E11.52 E11.59 E11.6 E11.61 E11.610 E11.618 E11.62 E11.620 E11.621 E11.622 E11.628 E11.63 E11.630 E11.638 E11.64 E11.641 E11.649 E11.65 E11.69 E11.8 E11.9
Obesity	E66 E66.0 E66.01 E66.09 E66.1 E66.2 E66.3 E66.8 E66.9
Major Depressive Disorder	F32 F32.0 F32.1 F32.2 F32.3 F32.4 F32.5 F32.8 F32.81 F32.89 F32.9 F32.A F33 F33.0 F33.1 F33.2 F33.3 F33.4 F33.40 F33.41 F33.42 F33.8 F33.9
Hypoparathyroidism	E20 E20.0 E20.1 E20.8 E20.9
Hyperparathyroidism	E21 E21.0 E21.1 E21.2 E21.3 E21.4 E21.5
Hypertensive Heart and Chronic Kidney Disease	I13 I13.0 I13.1 I13.10 I13.11 I13.2
Hypertensive Heart Disease	I11 I11.0 I11.9
Hypertensive Chronic Kidney Disease	I12 I12.0 I12.9
Hyperlipidemia	E78.0 E78.00 E78.01 E78.1 E78.2 E78.4 E78.41 E78.49 E78.5
Essential Hypertension	I10
Atrial Fibrillation and Flutter	I48 I48.0 I48.1 I48.11 I48.19 I48.2 I48.20 I48.21 I48.3 I48.4 I48.9 I48.91 I48.92
Atherosclerosis	I25.1 I25.10 I25.11 I25.110 I25.111 I25.118 I25.119 I25.2 I25.7 I25.70 I25.700 I25.701 I25.708 I25.709 I25.71 I25.710 I25.711 I25.718 I25.719 I25.72 I25.720 I25.721 I25.728 I25.729 I25.73 I25.730 I25.731 I25.738 I25.739 I25.75 I25.750 I25.751 I25.758 I25.759 I25.76 I25.760 I25.761 I25.768 I25.769 I25.79 I25.790 I25.791 I25.798 I25.799 I25.81 I25.810 I25.811 I25.812
Alcohol Dependence/Abuse	F10 F10.1 F10.10 F10.11 F10.12 F10.120 F10.121 F10.129 F10.13 F10.130 F10.131 F10.132 F10.139 F10.14 F10.15 F10.150 F10.151 F10.159 F10.18 F10.180 F10.181 F10.182 F10.188 F10.19 F10.2 F10.20 F10.21 F10.22 F10.220 F10.221 F10.229 F10.23 F10.230 F10.231 F10.232 F10.239 F10.24 F10.25 F10.250 F10.251 F10.259 F10.26 F10.27 F10.28 F10.280 F10.281 F10.282 F10.288 F10.29 F10.9 F10.92 F10.920 F10.921 F10.929 F10.93 F10.930 F10.931 F10.932 F10.939 F10.94 F10.95 F10.950 F10.951 F10.959 F10.96 F10.97 F10.98 F10.980 F10.981 F10.982 F10.988 F10.99

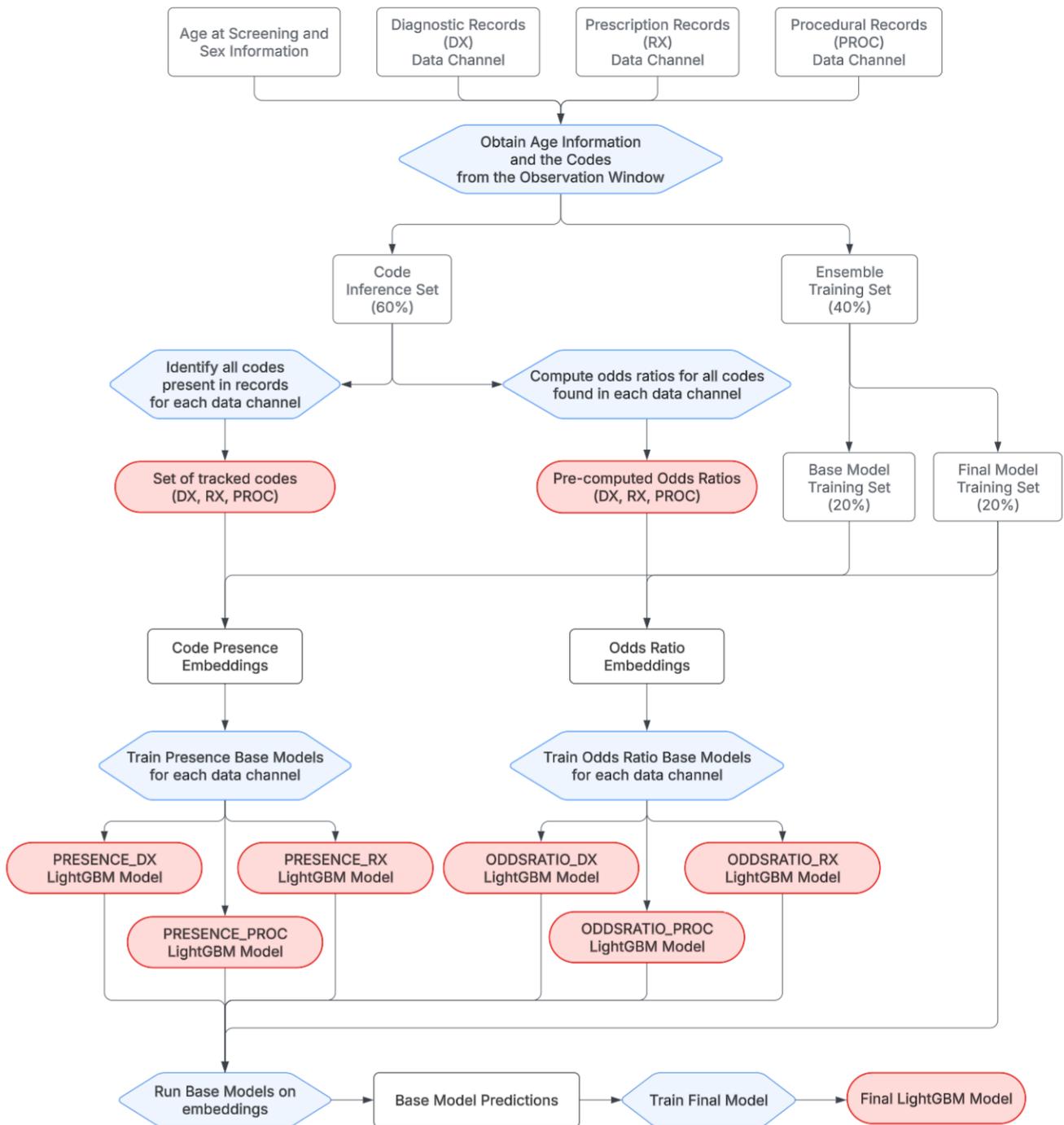


Figure SI-1: ZeBRA Training Diagram. We split the training dataset into the Code Inference Set to define the list of DX, RX, and PROC codes to track for code presence and odds ratio embeddings, and Ensemble Training Set that is embedded using inferred embeddings and used for training of base models and the final model that uses the predictions of base models as input to produce the final prediction.

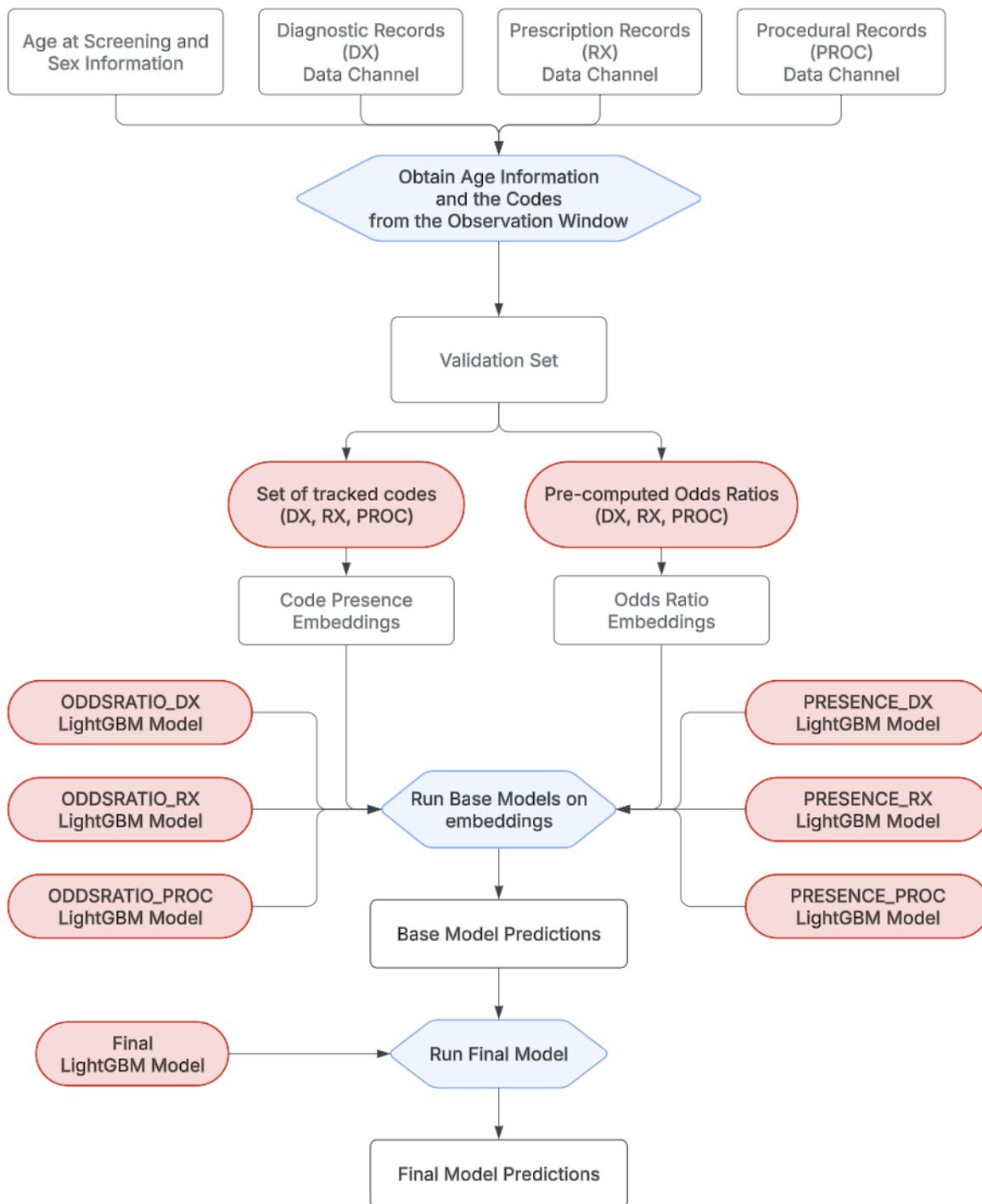


Figure SI-2: **ZeBRA Retrospective Validation Diagram.** We embed the validation dataset using sets of tracked DX, RX, and PROC codes and their precomputed odds ratios inferred during training to produce the input for base models. Predictions of the base models are used by final model of ZeBRA to produce the final prediction.

Table SI-V: Early runs - All Metrics at 95% Specificity

Prediction Window (years)	Dataset	TPR	PPV	NPV	Accuracy	LR+	LR-	AUC
0	National	0.789 ± 0.00	0.659 ± 0.00	0.973 ± 0.00	0.932 ± 0.00	15.737 ± 0.04	0.223 ± 0.00	0.957 ± 0.00
	Chicago	0.680 ± 0.04	0.624 ± 0.02	0.960 ± 0.01	0.921 ± 0.01	13.718 ± 0.88	0.337 ± 0.05	0.916 ± 0.02
	All Of Us	0.460 ± 0.05	0.528 ± 0.03	0.935 ± 0.01	0.896 ± 0.01	9.060 ± 1.05	0.569 ± 0.06	0.877 ± 0.02
1	National	0.689 ± 0.00	0.628 ± 0.00	0.961 ± 0.00	0.921 ± 0.00	13.637 ± 0.05	0.328 ± 0.00	0.933 ± 0.00
	Chicago	0.468 ± 0.02	0.534 ± 0.01	0.936 ± 0.00	0.897 ± 0.00	9.241 ± 0.36	0.560 ± 0.02	0.874 ± 0.01
	All Of Us	0.419 ± 0.05	0.504 ± 0.03	0.930 ± 0.01	0.892 ± 0.01	8.418 ± 1.03	0.611 ± 0.05	0.856 ± 0.02
2	National	0.647 ± 0.00	0.613 ± 0.00	0.956 ± 0.00	0.916 ± 0.00	12.771 ± 0.05	0.372 ± 0.00	0.925 ± 0.00
	Chicago	0.435 ± 0.02	0.516 ± 0.01	0.932 ± 0.00	0.893 ± 0.00	8.920 ± 0.38	0.594 ± 0.02	0.867 ± 0.01
	All Of Us	0.407 ± 0.05	0.509 ± 0.02	0.929 ± 0.01	0.891 ± 0.01	8.203 ± 1.02	0.624 ± 0.05	0.845 ± 0.02
3	National	0.590 ± 0.00	0.591 ± 0.00	0.950 ± 0.00	0.910 ± 0.00	11.690 ± 0.06	0.432 ± 0.00	0.914 ± 0.00
	Chicago	0.410 ± 0.02	0.500 ± 0.01	0.929 ± 0.00	0.891 ± 0.00	8.287 ± 0.40	0.621 ± 0.02	0.858 ± 0.01
	All Of Us	0.386 ± 0.05	0.484 ± 0.03	0.927 ± 0.01	0.888 ± 0.01	7.698 ± 1.00	0.646 ± 0.05	0.831 ± 0.03
4	National	0.545 ± 0.00	0.571 ± 0.00	0.945 ± 0.00	0.906 ± 0.00	10.838 ± 0.07	0.479 ± 0.00	0.904 ± 0.00
	Chicago	0.383 ± 0.02	0.483 ± 0.01	0.926 ± 0.00	0.888 ± 0.00	7.789 ± 0.41	0.649 ± 0.02	0.849 ± 0.01
	All Of Us	0.364 ± 0.05	0.469 ± 0.03	0.924 ± 0.01	0.886 ± 0.01	7.375 ± 0.98	0.669 ± 0.05	0.822 ± 0.03
5	National	0.516 ± 0.00	0.558 ± 0.00	0.941 ± 0.00	0.902 ± 0.00	10.264 ± 0.07	0.510 ± 0.00	0.899 ± 0.00
	Chicago	0.355 ± 0.02	0.464 ± 0.02	0.923 ± 0.00	0.885 ± 0.00	7.484 ± 0.42	0.679 ± 0.02	0.837 ± 0.01
	All Of Us	0.341 ± 0.05	0.452 ± 0.04	0.922 ± 0.01	0.883 ± 0.01	6.780 ± 0.95	0.694 ± 0.05	0.808 ± 0.03
6	National	0.480 ± 0.00	0.540 ± 0.00	0.937 ± 0.00	0.898 ± 0.00	9.503 ± 0.08	0.548 ± 0.00	0.888 ± 0.00
	Chicago	0.338 ± 0.02	0.452 ± 0.02	0.921 ± 0.00	0.884 ± 0.00	6.878 ± 0.44	0.697 ± 0.02	0.837 ± 0.01
	All Of Us	0.325 ± 0.05	0.441 ± 0.04	0.920 ± 0.01	0.882 ± 0.01	6.788 ± 0.94	0.710 ± 0.05	0.799 ± 0.03
7	National	0.442 ± 0.00	0.519 ± 0.00	0.933 ± 0.00	0.894 ± 0.00	8.775 ± 0.09	0.588 ± 0.01	0.878 ± 0.00
	Chicago	0.321 ± 0.02	0.439 ± 0.02	0.920 ± 0.00	0.881 ± 0.00	6.487 ± 0.46	0.715 ± 0.02	0.832 ± 0.01
	All Of Us	0.325 ± 0.05	0.441 ± 0.04	0.920 ± 0.01	0.882 ± 0.01	6.531 ± 0.94	0.711 ± 0.05	0.787 ± 0.03
8	National	0.408 ± 0.01	0.500 ± 0.00	0.929 ± 0.00	0.890 ± 0.00	8.042 ± 0.10	0.623 ± 0.01	0.865 ± 0.00
	Chicago	0.304 ± 0.03	0.426 ± 0.02	0.918 ± 0.00	0.879 ± 0.00	6.150 ± 0.50	0.732 ± 0.03	0.821 ± 0.02
	All Of Us	0.307 ± 0.05	0.427 ± 0.04	0.918 ± 0.01	0.879 ± 0.01	6.113 ± 0.92	0.730 ± 0.05	0.774 ± 0.03
9	National	0.373 ± 0.01	0.477 ± 0.00	0.925 ± 0.00	0.886 ± 0.00	7.342 ± 0.11	0.661 ± 0.01	0.850 ± 0.00
	Chicago	0.300 ± 0.03	0.422 ± 0.03	0.917 ± 0.00	0.879 ± 0.00	6.238 ± 0.61	0.737 ± 0.03	0.820 ± 0.02
	All Of Us	0.283 ± 0.05	0.407 ± 0.04	0.916 ± 0.01	0.877 ± 0.01	5.695 ± 0.90	0.754 ± 0.05	0.756 ± 0.03
10	National	0.333 ± 0.01	0.449 ± 0.00	0.921 ± 0.00	0.882 ± 0.00	6.575 ± 0.11	0.703 ± 0.01	0.831 ± 0.00
	Chicago	0.311 ± 0.04	0.430 ± 0.03	0.919 ± 0.00	0.881 ± 0.00	6.267 ± 0.76	0.725 ± 0.04	0.824 ± 0.02
	All Of Us	0.257 ± 0.04	0.383 ± 0.04	0.913 ± 0.01	0.874 ± 0.01	5.154 ± 0.86	0.783 ± 0.05	0.737 ± 0.03

Table SI-VI: Number of codes and code prefixes tracked used in Code Presence feature vectors.

DX		RX		PROC	
Prefix	No. Tracked	Prefix	No. Tracked	Prefix	No. Tracked
1	36	3	30	1	28
2	289	4	174	2	188
3	2,014	5	239	3	1,180
5	5,713	8	1,713	4	5,963
6	13,018			5	17,261
TOTAL	21,070	TOTAL	2,156	TOTAL	24,620

Table SI-VII: Odds Ratios Aggregations

Aggregation Name	Formula
mean_or	$\text{mean}(r), \forall r \in x_j^{\text{OR}}$
max_or	$\text{max}(r), \forall r \in x_j^{\text{OR}}$
sum_or	$\text{sum}(r), \forall r \in x_j^{\text{OR}}$
std_or	$\text{standard_deviation}(r), \forall r \in x_j^{\text{OR}}$
range_or	$\text{max}(r) - \text{min}(r), \forall r \in x_j^{\text{OR}}$
geomean_or	$\text{geometric_mean}([r \text{ if } r \neq 0 \text{ else } 10^{-7}], \forall r \in x_j^{\text{OR}})$
variation_or	$\text{variation}([r \text{ if } r \neq 0 \text{ else } 10], \forall r \in x_j^{\text{OR}})$
meanstd_or	$\text{mean_or} \times \text{std_or}$
meanrng_or	$\text{mean_or} \times \text{range_or}$
meanstd_sq_or	$\text{mean_or}^{\text{std_or}}$
geomeanrng_or	$\text{geomean_or} \times \text{range_or}$
geomean_sq_sum	$\sqrt{\text{geomean_or}^{\text{sum_or}}}$