

Supplementary table 1. Confusion matrix–derived performance metrics of the LASSO model evaluated using claims data at T-12M, T-9M and T-6M

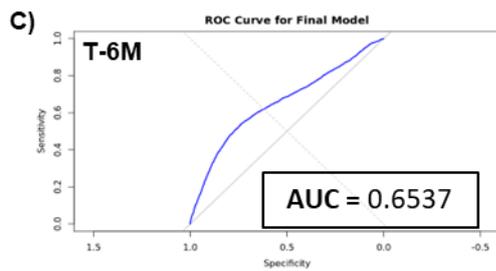
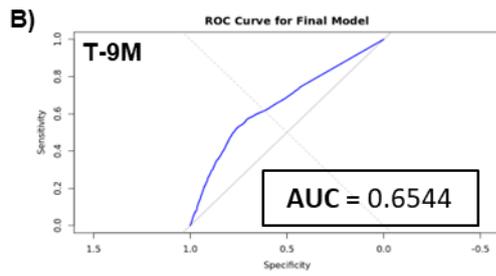
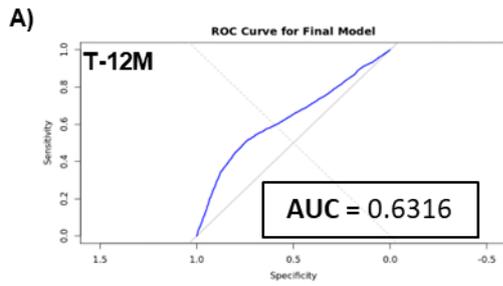
Time period prior to the index date	True positives (n)	False negatives (n)	True negatives (n)	False positives (n)	Sensitivity (%)	Specificity (%)	Accuracy
T-12M	2127	2606	3743	990	44.9	79.1	62.00
T-9M	2339	2394	3723	1010	49.4	78.7	64.00
T-6M	2808	1925	3276	1457	59.3	69.2	64.30

Supplementary table 3. LASSO coefficients for codes across T-9M associated with higher or lower MM diagnosis risk. Positive coefficients are labelled in green. Negative coefficients are labelled in red

Variable Name	Claim code or code group description	Coefficient
GNM_G2	Flu vaccination	0.385640834
DXG_1	Anemias	0.163557452
GNM_G1	COVID vaccination	0.122031183
PROC_G1	M-protein / immunoglobulin quantification	0.099514777
DXG_9	COVID or viral exposure	0.087946864
DX_Z23	Encounter for immunization	0.084490743
DXG_2	Neutropenia	0.079666392
DXG_5	Aortic aneurysm	0.071635085
GNM_G3	Pneumococcal vaccination	0.05250873
DX_K210	GERD with esophagitis	0.045338417
DXG_4	Cardiovascular disorders	0.044692935
PROC_G13	Observation / short-term hospital care	0.042698701
HCP_00030	Hematology & oncology visit	0.041553532
DX_I472	Ventricular tachycardia	0.038138744
DXG_11	Other infection concern	0.036153447
DXG_6	Musculoskeletal pain	0.031377837
PROC_7025F	Next-mammogram alert flag	0.029774555
DXG_8	Shortness of breath	0.028473718
PROC_92012	Interim ophthalmic exam, established pt	0.023837805
DX_D696	Thrombocytopenia (unspecified)	0.023624812
PROC_G8	Vaccination administration	0.022874899
PROC_83883	Nephelometry assay (not specified)	0.02276766
DX_D61818	Other pancytopenia	0.021192948
PROC_G2	Anemia workup	0.017766265
PROC_88313	Special stains (group 2)	0.015054476
DXG_3	Chronic kidney or metabolic disease	0.012940402
PROC_G12	Routine follow-up office visit	0.010302767
HCP_00058	Pathology visit	0.009374816
PROC_88342	Immunochem/cytochem first antibody	0.003543287
PROC_G14	Nursing facility / long-term care visit	-0.038230504
PROC_G15	Routine monitoring	-0.043885999
GNM_00798	Nirmatrelvir/ritonavir	-0.044857031
HCP_00077	Registered nurse practitioner visit	-0.047454626
DX_F32A	Depression (unspecified)	-0.072382333
DX_R4182	Altered mental status (unspecified)	-0.080074096
PROC_1159F	Medication list documented	-0.082655529
HCP_00055	Other non-physician provider visit	-0.085490702
dementia	Dementia	-0.117926849

Supplementary table 4. LASSO coefficients for codes across T-6M associated with higher or lower MM diagnosis risk. Positive coefficients are labelled in green. Negative coefficients are labelled in red.

Variable Name	Claim code or code group description	Coefficient
DXG_1	Anemias	0.17684081
PROC_G1	M-protein / immunoglobulin quantification	0.15987831
DXG_9	COVID or viral exposure	0.15654916
GNM_G2	Flu vaccination	0.15263565
PROC_G8	Vaccination administration	0.15235427
DXG_5	Aortic aneurysm	0.10891396
DXG_2	Neutropenia	0.09646837
DXG_6	Musculoskeletal pain	0.07732585
DX_R0781	Pleurodynia	0.06300981
PROC_G13	Observation / short-term hospital care	0.05985073
HCP_00030	Hematology & oncology visit	0.0594154
DXG_4	Cardiovascular disorders	0.0579952
DXG_11	Other infection concern	0.05118473
PROC_7025F	Next-mammogram alert flag	0.04613101
DX_K210	GERD with esophagitis	0.03708037
DX_D61818	Other pancytopenia	0.03035508
DX_G5602	Carpal tunnel syndrome	0.02876568
DX_R161	Splenomegaly	0.02531103
PROC_92012	Interim ophthalmic exam, established pt	0.02499462
DX_R51	Headache	0.02401146
PROC_G5	Autoimmune and inflammatory markers	0.02391142
PROC_83883	Nephelometry assay (not specified)	0.02389999
DXG_3	Chronic kidney and metabolic disease	0.01420622
PROC_88342	Immunochem/cytochem first antibody	0.01377713
PROC_G11	New patient office visit	0.00800802
DX_D696	Thrombocytopenia (unspecified)	0.00767641
DXG_8	Shortness of breath	0.00563316
PROC_G3	Bone marrow and cytogenetics	0.00486413
DX_Z23	Encounter for immunization	0.00295338
PROC_G2	Anemia workup	0.002483
PROC_88313	Special stains – group 2	0.00246132
GNM_00980	Semaglutide	-0.00161042
PROC_1126F	Amount of pain assessed – none present	-0.00261368
PROC_G15	Routine monitoring	-0.04438367
OP	Outpatient visit	-0.04737767
PROC_G14	Nursing facility or long-term care visit	-0.05201291
GNM_00798	Nirmatrelvir/ritonavir	-0.05328354
HCP_00077	Registered nurse practitioner visit	-0.06187192
PROC_0241U	Respiratory viral RNA panel (4-target)	-0.06332053
HCP_00055	Other non-physician provider visit	-0.09007853



Supplementary figure 1. Receiver operating characteristic (ROC) curve performance of the Random Forest model, evaluated using claims data at T-12M(A), T-9M(B), and T-6M(C). The area under the curve (AUC) for each model is shown.

Supplementary table 5. Confusion matrix–derived performance metrics of the Random Forest model evaluated using claims data at T-12M, T-9M and T-6M

Time period prior to index date	True positives (n)	False negatives (n)	True negatives (n)	False positives (n)	Sensitivity (%)	Specificity (%)	Accuracy (%)
T-12M	2,165	2,568	3,744	989	45.7	79.1	62.4
T-9M	2,418	2,315	3,636	1,097	51.0	76.8	64.0
T-6M	2,372	2,361	3,655	1,078	50.1	77.2	63.7

Supplementary table 6. Importance scores influencing the Random Forest model derived from variable obtained across T-12M

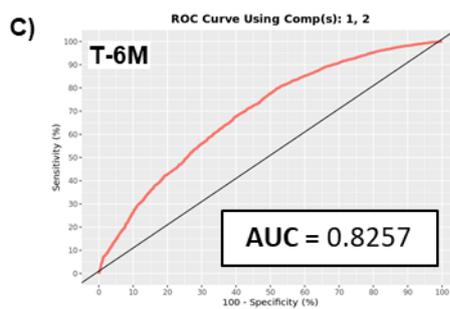
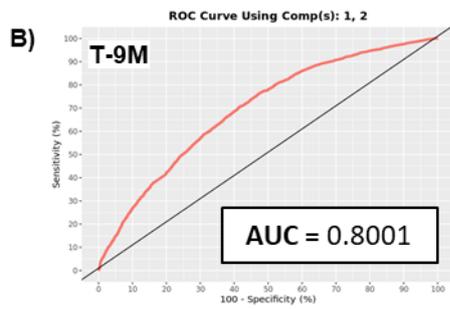
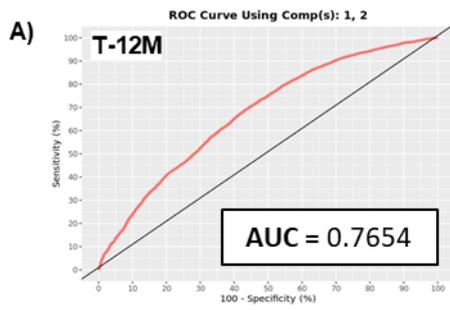
Variable Name	Claim code or code group description	Importance
GNM_G2	Flu vaccination	86.42833886
DXG_1	Anemias	32.3644001
PROC_G10	COVID-19 related diagnostics	31.23478121
GNM_G3	Pneumococcal vaccination	24.17851449
PROC_G13	Observation / short-term hospital care	18.23174641
PROC_G12	Established patient office visit	16.47987593
HCP_00030	Hematology & oncology visit	16.31060487
PROC_G15	Routine monitoring	16.13226792
HCP_00055	Other non-physician provider visit	15.71165173
DX_K210	GERD with esophagitis	14.59649721
PROC_G1	M-protein / immunoglobulin quantification	14.0654595
dementia	Dementia	13.67799212
PROC_G14	Nursing facility / long-term care visit	12.871378
PROC_G2	Anemia workup	11.49121089
DXG_11	Other infection concern	10.41527828
OP	Outpatient visit	10.20762067

Supplementary table 7. Importance scores influencing the Random Forest model derived from variable obtained across T-9M

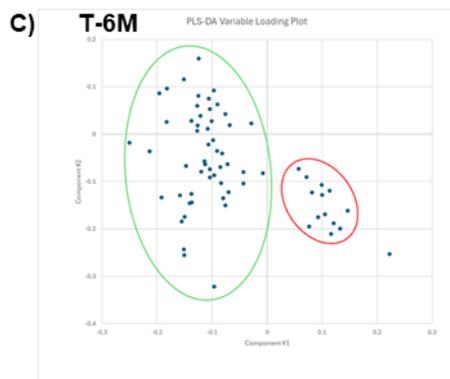
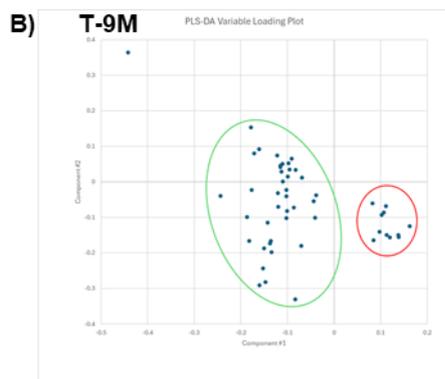
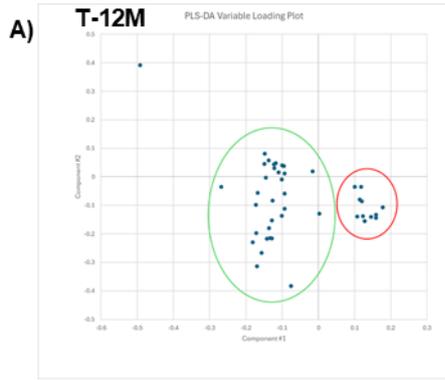
Variable Name	Claim code or code group description	Importance
GNM_G2	Flu vaccination	82.25674875
PROC_G10	COVID-19 related diagnostics	30.60178597
GNM_G1	COVID vaccination	28.26651902
DXG_1	Anemias	25.75405419
PROC_G13	Observation / short-term hospital care	19.65446147
HCP_00030	Hematology & oncology visit	19.33729139
dementia	Dementia	16.46607852
DXG_11	Other infection concern	16.09684195
PROC_G14	Nursing facility / long-term care visit	13.83380816
HCP_00055	Other non-physician provider visit	13.28737525
GNM_G3	Pneumococcal vaccination	12.71087186
PROC_G1	M-protein / immunoglobulin quantification	12.29070368
PROC_G2	Anemia workup	11.58102392
DXG_2	Neutropenia	11.5244387
DX_R4182	Altered mental status unspecified	11.5187495
DX_Z23	Encounter for immunization	11.38566872
DXG_9	COVID or viral exposure	11.18101576
DX_K210	GERD with esophagitis	10.95947294
PROC_83550	Iron-binding test	10.16376392

Supplementary table 8. Importance scores influencing the Random Forest model derived from variable obtained across T-6M

Variable Name	Claim code or code group description	Importance
GNM_G2	Flu vaccination	41.49333483
GNM_G1	COVID vaccination	31.86712438
PROC_G10	COVID-19 related diagnostics	31.46300774
DXG_1	Anemias	26.98852274
PROC_G8	Vaccination administration	21.58694684
HCP_00030	Hematology & oncology visit	20.23724264
PROC_G1	M-protein / immunoglobulin quantification	19.20668796
dementia	Dementia	16.95323903
DX_R0781	Pleurodynia	13.82870927
PROC_G13	Observation / short-term hospital care	13.72996895
DXG_2	Neutropenia	12.72185917
PROC_G12	Established patient office visit	11.37011114
PROC_G14	Nursing facility / long-term care visit	11.20628663
DXG_5	Aortic aneurysm	11.08541207
PROC_G2	Anemia workup	10.6572917
DX_F32A	Depression unspecified	10.47106861
PROC_G4	Peripheral blood analysis	10.33675145
DX_R4182	Altered mental status unspecified	10.20136063
HCP_00055	Other non-physician provider visit	10.19785569



Supplementary Figure 2. Receiver operating characteristic (ROC) curve performance of the Partial Least Squares Discriminant Analysis (PLS-DA) model evaluated using claims data at T-12M(A), T-9M(B), and T-6M(C). The area under the curve (AUC) for each model is shown.



Supplementary figure 3. Partial Least Squares Discriminant Analysis loading plots derived from claims data at T-12M (A), T-9M (B) and T-6M (C). Green and red circles highlight clusters of individual or groups of claim codes that have shared covariance patterns.

Supplementary table 9. Confusion matrix–derived performance metrics of the Sequential Neural Network model evaluated using claims data at T-12M, T-9M and T-6M

Time period prior to index date	True positives (n)	False negatives (n)	True negatives (n)	False positives (n)	Sensitivity (%)	Specificity (%)	Accuracy (%)
T-12M	2,943	1,790	3,541	1,192	62.1	74.8	68.3
T-9M	3,581	1,152	3,116	1,617	75.6	65.1	70.7
T-6M	3,734	999	3,137	1,596	78.9	66.3	72.8