

Supplementary material: Self-reported diagnoses as training labels for finding uncoded chronic fatigue, fibromyalgia, and endometriosis in the All of Us Research Program

Tables S1–S3 list the candidate features of the chronic fatigue, fibromyalgia, and endometriosis models. We generated them from the same feature registries we used to fit the models. Table S4 is the TRIPOD+AI checklist.

Table S1: Candidate features for the chronic fatigue model. Registry unm-chronicfatigue-1.0.0 (version 1), 73 candidate binary features. The label codes (G93.3, R53.82) are never features, and participants with a label code on or before day 0 were not scored. Symptom groups: autonomic, cognitive, fatigue, gastrointestinal, immune, mood, pain, sleep. All 73 features met the eligibility rule for the self-report label (main text, Models). Window: 5 y covers days 1,825 to 1 before day 0, and 1 y covers days 365 to 1 before day 0. We evaluated demographics at day 0. ICD-10-CM codes match by prefix on the source code in the condition table and, for the 35 features marked *, also in the observation table. A symptom-group count is 1 when diagnosis-code features from at least the stated number of symptom groups are recorded. CPT and HCPCS codes match source codes in the procedure table exactly (ranges inclusive). A laboratory feature is any recorded measurement of a standard LOINC measurement concept whose name contains the text shown (case-insensitive). A medication feature is any drug exposure record for an RxNorm or RxNorm Extension concept descending from the listed ingredients. Visit classes come from the visit type: inpatient, emergency, or outpatient (all other visit types except pharmacy and laboratory). A feature is 1 if a qualifying record exists in its window and 0 otherwise. A feature entered a model only if it met the eligibility rule in the main text (Models).

Feature ID	Label	Window	Codes by system
<i>Demographics</i>			
age_18_39	Age 18–39	Day 0	Age at day 0 (calendar year minus birth year), 18 to 39 years
age_40_49	Age 40–49	Day 0	Age at day 0 (calendar year minus birth year), 40 to 49 years
age_50_59	Age 50–59	Day 0	Age at day 0 (calendar year minus birth year), 50 to 59 years
age_60_69	Age 60–69	Day 0	Age at day 0 (calendar year minus birth year), 60 to 69 years
age_70p	Age 70 or older	Day 0	Age at day 0 (calendar year minus birth year), 70 years or older
sex_female	Female	Day 0	Sex at birth recorded as female
<i>Diagnosis codes</i>			
cx_fatigue_other*	Fatigue or malaise coded	5 y	ICD-10-CM: R53.0, R53.1, R53.81, R53.83 Symptom group: fatigue
cx_fatigue_recent*	Fatigue or malaise coded in the past year	1 y	ICD-10-CM: R53.0, R53.1, R53.81, R53.83

Table S1: (continued)

Feature ID	Label	Window	Codes by system
cx_sleep*	Sleep disturbance	5 y	ICD-10-CM: G47.0, G47.1, G47.8, G47.9, F51 Symptom group: sleep
cx_sleep_apnea*	Sleep apnea	5 y	ICD-10-CM: G47.33
cx_cognitive*	Cognitive symptoms (brain fog, memory)	5 y	ICD-10-CM: R41, G31.84 Symptom group: cognitive
cx_autonomic*	Palpitations, POTS, orthostatic hypotension, or dysautonomia	5 y	ICD-10-CM: R00.0, R00.2, I49.8, G90.A, I95.1, G90.9 Symptom group: autonomic
cx_dizzy_syncope*	Dizziness or fainting	5 y	ICD-10-CM: R42, R55 Symptom group: autonomic
cx_headache*	Headache	5 y	ICD-10-CM: R51, G44 Symptom group: pain
cx_migraine*	Migraine	5 y	ICD-10-CM: G43
cx_myalgia*	Muscle or joint pain	5 y	ICD-10-CM: M79.1, M25.5, M79.6
arthralgia*			Symptom group: pain
cx_fibromyalgia*	Fibromyalgia	5 y	ICD-10-CM: M79.7
cx_chronic_pain*	Chronic pain	5 y	ICD-10-CM: G89
cx_depression*	Depression	5 y	ICD-10-CM: F32, F33 Symptom group: mood
cx_anxiety*	Anxiety	5 y	ICD-10-CM: F41 Symptom group: mood
cx_ptsd*	Posttraumatic stress disorder	5 y	ICD-10-CM: F43.1
cx_ibs_gi*	Irritable bowel or gastrointestinal symptoms	5 y	ICD-10-CM: K58, R10, R11, R14, R19 Symptom group: gastrointestinal
cx_sore_throat_nodes*	Sore throat or enlarged lymph nodes	5 y	ICD-10-CM: J02, R59 Symptom group: immune
cx_fever*	Fever	5 y	ICD-10-CM: R50 Symptom group: immune
cx_ebv_mono*	Infectious mononucleosis (EBV)	5 y	ICD-10-CM: B27
cx_lyme*	Lyme disease	5 y	ICD-10-CM: A69.2
cx_covid*	COVID-19	5 y	ICD-10-CM: U07.1
cx_long_covid*	Long COVID	5 y	ICD-10-CM: U09.9
cx_eds_hypermobility*	Ehlers-Danlos syndrome or hypermobility	5 y	ICD-10-CM: Q79.6, M35.7
cx_mast_cell*	Mast cell disorder	5 y	ICD-10-CM: D89.4
cx_hypothyroid*	Hypothyroidism	5 y	ICD-10-CM: E03
cx_anemia*	Anemia	5 y	ICD-10-CM: D50, D51, D52, D53, D63, D64
cx_vitd_deficiency*	Vitamin D deficiency	5 y	ICD-10-CM: E55
cx_autoimmune*	Autoimmune or demyelinating disease	5 y	ICD-10-CM: M05, M06, M32, M35, L40, K50, K51, M45, G35
cx_obesity*	Obesity	5 y	ICD-10-CM: E66, Z68.3, Z68.4
cx_diabetes*	Diabetes	5 y	ICD-10-CM: E10, E11
cx_hypertension*	Hypertension	5 y	ICD-10-CM: I10
cx_heart_disease*	Coronary disease or heart failure	5 y	ICD-10-CM: I25, I50

Table S1: (continued)

Feature ID	Label	Window	Codes by system
cx_cancer*	Cancer	5 y	ICD-10-CM: C
<i>Symptom-group counts</i>			
clusters_ge3*	Three or more symptom types	5 y	Number of symptom groups recorded: 3 or more
clusters_ge5*	Five or more symptom types	5 y	Number of symptom groups recorded: 5 or more
<i>Procedures</i>			
px_tilt	Tilt-table test	5 y	CPT or HCPCS: 93660
px_sleep_study	Sleep study	5 y	CPT or HCPCS: 95806–95811, G0398, G0399, G0400
px_brain_mri	Brain MRI	5 y	CPT or HCPCS: 70551–70553
px_echo	Echocardiogram	5 y	CPT or HCPCS: 93303–93308, 93312–93318, 93350–93352
px_cardiac_monitor	Holter or event monitor	5 y	CPT or HCPCS: 93224–93229, 93241–93248, 93268–93272
px_exercise_test	Exercise or cardiopulmonary exercise test	5 y	CPT or HCPCS: 94621, 93015–93018
px_neuropsych	Neuropsychological testing	5 y	CPT or HCPCS: 96132, 96133, 96136–96139
px_pt_eval	Physical therapy evaluation	5 y	CPT or HCPCS: 97161–97163
<i>Laboratory tests</i>			
lab_tsh	Thyroid (TSH) test	5 y	LOINC: measurement concepts whose name contains “Thyrotropin”
lab_ferritin	Ferritin test	5 y	LOINC: measurement concepts whose name contains “Ferritin”
lab_b12	Vitamin B12 test	5 y	LOINC: measurement concepts whose name contains “Cobalamin”
lab_vitamin_d	Vitamin D test	5 y	LOINC: measurement concepts whose name contains “25-hydroxyvitamin D”
lab_ana	Antinuclear antibody (ANA) test	5 y	LOINC: measurement concepts whose name contains “Nuclear Ab”
lab_crp	C-reactive protein test	5 y	LOINC: measurement concepts whose name contains “C reactive protein”
lab_esr	Sedimentation rate (ESR) test	5 y	LOINC: measurement concepts whose name contains “Erythrocyte sedimentation rate”
lab_ebv	Epstein-Barr virus test	5 y	LOINC: measurement concepts whose name contains “Epstein Barr virus”
lab_lyme	Lyme disease test	5 y	LOINC: measurement concepts whose name contains “Borrelia burgdorferi”
lab_cortisol	Cortisol test	5 y	LOINC: measurement concepts whose name contains “Cortisol”
lab_ck	Creatine kinase test	5 y	LOINC: measurement concepts whose name contains “Creatine kinase”
<i>Medications</i>			
rx_stimulant	Stimulant or wakefulness drug	5 y	RxNorm ingredients: modafinil, armodafinil, methylphenidate, amphetamine, dextroamphetamine, lisdexamfetamine

Table S1: (continued)

Feature ID	Label	Window	Codes by system
<code>rx_naltrexone</code>	Naltrexone	5 y	RxNorm ingredients: naltrexone
<code>rx_ssri_snri</code>	SSRI or SNRI antidepressant	5 y	RxNorm ingredients: sertraline, fluoxetine, citalopram, escitalopram, paroxetine, venlafaxine, duloxetine, desvenlafaxine
<code>rx_tricyclic_lowdose</code>	Amitriptyline, nortriptyline, or cyclobenzaprine	5 y	RxNorm ingredients: amitriptyline, nortriptyline, cyclobenzaprine
<code>rx_gabapentinoid</code>	Gabapentin or pregabalin	5 y	RxNorm ingredients: gabapentin, pregabalin
<code>rx_sleep_aid</code>	Sleep medication	5 y	RxNorm ingredients: zolpidem, trazodone, eszopiclone
<code>rx_beta_blocker</code>	Beta blocker	5 y	RxNorm ingredients: propranolol, metoprolol, atenolol, bisoprolol, nadolol, carvedilol
<code>rx_pots</code>	POTS medication	5 y	RxNorm ingredients: ivabradine, midodrine, fludrocortisone, pyridostigmine
<code>rx_opioid</code>	Opioid	5 y	RxNorm ingredients: oxycodone, hydrocodone, tramadol, morphine
<i>Health care use</i>			
<code>util_out_r_ge6</code>	6+ outpatient visits in the past year	1 y	Outpatient visits: 6 or more
<code>util_out_r_ge12</code>	12+ outpatient visits in the past year	1 y	Outpatient visits: 12 or more
<code>util_ed_r</code>	Emergency visit in the past year	1 y	Emergency visits: 1 or more
<code>util_ip_h</code>	Hospital stay in the past 5 years	5 y	Inpatient visits: 1 or more

Table S2: Candidate features for the fibromyalgia model. Registry unnm-fibromyalgia-1.0.0 (version 1), 83 candidate binary features. The label codes (M79.7) are never features, and participants with a label code on or before day 0 were not scored. Symptom groups: autonomic, cognitive, fatigue, gastrointestinal, immune, mood, pain, sleep. Of the 83 features, 80 met the eligibility rule for the self-report label (main text, Models). The exceptions were `cx_ebv_mono`, `cx_mast_cell`, and `px_tilt`. Windows, matching rules, and eligibility as in Table S1, except that all diagnosis-code and symptom-group features use both the condition and observation tables.

Feature ID	Label	Window	Codes by system
<i>Demographics</i>			
<code>age_18_39</code>	Age 18–39	Day 0	Age at day 0 (calendar year minus birth year), 18 to 39 years
<code>age_40_49</code>	Age 40–49	Day 0	Age at day 0 (calendar year minus birth year), 40 to 49 years
<code>age_50_59</code>	Age 50–59	Day 0	Age at day 0 (calendar year minus birth year), 50 to 59 years
<code>age_60_69</code>	Age 60–69	Day 0	Age at day 0 (calendar year minus birth year), 60 to 69 years
<code>age_70p</code>	Age 70 or older	Day 0	Age at day 0 (calendar year minus birth year), 70 years or older
<code>sex_female</code>	Female	Day 0	Sex at birth recorded as female
<i>Diagnosis codes</i>			
<code>cx_fatigue_other</code>	Fatigue or malaise coded	5 y	ICD-10-CM: R53.0, R53.1, R53.81, R53.83 Symptom group: fatigue
<code>cx_fatigue_recent</code>	Fatigue or malaise coded in the past year	1 y	ICD-10-CM: R53.0, R53.1, R53.81, R53.83
<code>cx_sleep</code>	Sleep disturbance	5 y	ICD-10-CM: G47.0, G47.1, G47.8, G47.9, F51 Symptom group: sleep
<code>cx_sleep_apnea</code>	Sleep apnea	5 y	ICD-10-CM: G47.33
<code>cx_cognitive</code>	Cognitive symptoms (brain fog, memory)	5 y	ICD-10-CM: R41, G31.84 Symptom group: cognitive
<code>cx_autonomic</code>	Palpitations, POTS, orthostatic hypotension, or dysautonomia	5 y	ICD-10-CM: R00.0, R00.2, I49.8, G90.A, I95.1, G90.9 Symptom group: autonomic
<code>cx_dizzy_syncope</code>	Dizziness or fainting	5 y	ICD-10-CM: R42, R55 Symptom group: autonomic
<code>cx_headache</code>	Headache	5 y	ICD-10-CM: R51, G44 Symptom group: pain
<code>cx_migraine</code>	Migraine	5 y	ICD-10-CM: G43
<code>cx_myalgia_arthralgia</code>	Muscle or joint pain	5 y	ICD-10-CM: M79.1, M25.5, M79.6 Symptom group: pain
<code>cx_chronic_pain</code>	Chronic pain	5 y	ICD-10-CM: G89
<code>cx_depression</code>	Depression	5 y	ICD-10-CM: F32, F33 Symptom group: mood
<code>cx_anxiety</code>	Anxiety	5 y	ICD-10-CM: F41 Symptom group: mood
<code>cx_ptsd</code>	Posttraumatic stress disorder	5 y	ICD-10-CM: F43.1
<code>cx_ibs_gi</code>	Irritable bowel or gastrointestinal symptoms	5 y	ICD-10-CM: K58, R10, R11, R14, R19 Symptom group: gastrointestinal
<code>cx_sore_throat_nodes</code>	Sore throat or enlarged lymph nodes	5 y	ICD-10-CM: J02, R59 Symptom group: immune

Table S2: (continued)

Feature ID	Label	Window	Codes by system
cx_fever	Fever	5 y	ICD-10-CM: R50 Symptom group: immune
cx_ebv_mono	Infectious mononucleosis (EBV)	5 y	ICD-10-CM: B27
cx_lyme	Lyme disease	5 y	ICD-10-CM: A69.2
cx_covid	COVID-19	5 y	ICD-10-CM: U07.1
cx_long_covid	Long COVID	5 y	ICD-10-CM: U09.9
cx_eds_hypermobility	Ehlers-Danlos syndrome or hypermobility	5 y	ICD-10-CM: Q79.6, M35.7
cx_mast_cell	Mast cell disorder	5 y	ICD-10-CM: D89.4
cx_hypothyroid	Hypothyroidism	5 y	ICD-10-CM: E03
cx_anemia	Anemia	5 y	ICD-10-CM: D50, D51, D52, D53, D63, D64
cx_vitd_deficiency	Vitamin D deficiency	5 y	ICD-10-CM: E55
cx_autoimmune	Autoimmune or demyelinating disease	5 y	ICD-10-CM: M05, M06, M32, M35, L40, K50, K51, M45, G35
cx_obesity	Obesity	5 y	ICD-10-CM: E66, Z68.3, Z68.4
cx_diabetes	Diabetes	5 y	ICD-10-CM: E10, E11
cx_hypertension	Hypertension	5 y	ICD-10-CM: I10
cx_heart_disease	Coronary disease or heart failure	5 y	ICD-10-CM: I25, I50
cx_cancer	Cancer	5 y	ICD-10-CM: C
cx_cfs	Chronic fatigue or ME/CFS	5 y	ICD-10-CM: G93.3, R53.82 Symptom group: fatigue
cx_tmj	Temporomandibular joint disorder	5 y	ICD-10-CM: M26.6 Symptom group: pain
cx_rls	Restless legs syndrome	5 y	ICD-10-CM: G25.81 Symptom group: sleep
cx_interstitial_cystitis	Interstitial cystitis or bladder pain	5 y	ICD-10-CM: N30.1
cx_osteoarthritis	Osteoarthritis	5 y	ICD-10-CM: M15, M16, M17, M18, M19
cx_neck_back_pain	Neck or low back pain	5 y	ICD-10-CM: M54.2, M54.5 Symptom group: pain
<i>Symptom-group counts</i>			
clusters_ge3	Three or more symptom types	5 y	Number of symptom groups recorded: 3 or more
clusters_ge5	Five or more symptom types	5 y	Number of symptom groups recorded: 5 or more
<i>Procedures</i>			
px_tilt	Tilt-table test	5 y	CPT or HCPCS: 93660
px_sleep_study	Sleep study	5 y	CPT or HCPCS: 95806–95811, G0398, G0399, G0400
px_brain_mri	Brain MRI	5 y	CPT or HCPCS: 70551–70553
px_echo	Echocardiogram	5 y	CPT or HCPCS: 93303–93308, 93312–93318, 93350–93352
px_cardiac_monitor	Holter or event monitor	5 y	CPT or HCPCS: 93224–93229, 93241–93248, 93268–93272
px_exercise_test	Exercise or cardiopulmonary exercise test	5 y	CPT or HCPCS: 94621, 93015–93018

Table S2: (continued)

Feature ID	Label	Window	Codes by system
px_neuropsych	Neuropsychological testing	5 y	CPT or HCPCS: 96132, 96133, 96136–96139
px_pt_eval	Physical therapy evaluation	5 y	CPT or HCPCS: 97161–97163
<i>Laboratory tests</i>			
lab_tsh	Thyroid (TSH) test	5 y	LOINC: measurement concepts whose name contains “Thyrotropin”
lab_ferritin	Ferritin test	5 y	LOINC: measurement concepts whose name contains “Ferritin”
lab_b12	Vitamin B12 test	5 y	LOINC: measurement concepts whose name contains “Cobalamin”
lab_vitamin_d	Vitamin D test	5 y	LOINC: measurement concepts whose name contains “25-hydroxyvitamin D”
lab_ana	Antinuclear antibody (ANA) test	5 y	LOINC: measurement concepts whose name contains “Nuclear Ab”
lab_crp	C-reactive protein test	5 y	LOINC: measurement concepts whose name contains “C reactive protein”
lab_esr	Sedimentation rate (ESR) test	5 y	LOINC: measurement concepts whose name contains “Erythrocyte sedimentation rate”
lab_ebv	Epstein-Barr virus test	5 y	LOINC: measurement concepts whose name contains “Epstein Barr virus”
lab_lyme	Lyme disease test	5 y	LOINC: measurement concepts whose name contains “Borrelia burgdorferi”
lab_cortisol	Cortisol test	5 y	LOINC: measurement concepts whose name contains “Cortisol”
lab_ck	Creatine kinase test	5 y	LOINC: measurement concepts whose name contains “Creatine kinase”
lab_rf	Rheumatoid factor test	5 y	LOINC: measurement concepts whose name contains “Rheumatoid factor”
lab_ccp	Anti-CCP antibody test	5 y	LOINC: measurement concepts whose name contains “Cyclic citrullinated peptide”
<i>Medications</i>			
rx_stimulant	Stimulant or wakefulness drug	5 y	RxNorm ingredients: modafinil, armodafinil, methylphenidate, amphetamine, dextroamphetamine, lisdexamfetamine
rx_naltrexone	Naltrexone	5 y	RxNorm ingredients: naltrexone
rx_ssri_snri	SSRI or SNRI antidepressant	5 y	RxNorm ingredients: sertraline, fluoxetine, citalopram, escitalopram, paroxetine, venlafaxine, duloxetine, desvenlafaxine
rx_tricyclic_lowdose	Amitriptyline, nortriptyline, or cyclobenzaprine	5 y	RxNorm ingredients: amitriptyline, nortriptyline, cyclobenzaprine
rx_gabapentinoid	Gabapentin or pregabalin	5 y	RxNorm ingredients: gabapentin, pregabalin
rx_sleep_aid	Sleep medication	5 y	RxNorm ingredients: zolpidem, trazodone, eszopiclone
rx_beta_blocker	Beta blocker	5 y	RxNorm ingredients: propranolol, metoprolol, atenolol, bisoprolol, nadolol, carvedilol

Table S2: (continued)

Feature ID	Label	Window	Codes by system
<code>rx_pots</code>	POTS medication	5 y	RxNorm ingredients: ivabradine, midodrine, fludrocortisone, pyridostigmine
<code>rx_opioid</code>	Opioid	5 y	RxNorm ingredients: oxycodone, hydrocodone, tramadol, morphine
<code>rx_milnacipran</code>	Milnacipran	5 y	RxNorm ingredients: milnacipran
<code>rx_nsaid</code>	NSAID pain reliever	5 y	RxNorm ingredients: naproxen, ibuprofen, meloxicam, diclofenac, ketorolac
<code>rx_muscle_relaxant</code>	Muscle relaxant	5 y	RxNorm ingredients: tizanidine, methocarbamol, baclofen
<i>Health care use</i>			
<code>util_out_r_ge6</code>	6+ outpatient visits in the past year	1 y	Outpatient visits: 6 or more
<code>util_out_r_ge12</code>	12+ outpatient visits in the past year	1 y	Outpatient visits: 12 or more
<code>util_ed_r</code>	Emergency visit in the past year	1 y	Emergency visits: 1 or more
<code>util_ip_h</code>	Hospital stay in the past 5 years	5 y	Inpatient visits: 1 or more

Table S3: Candidate features for the endometriosis model. Registry unnm-endometriosis-1.0.0 (version 1), 94 candidate binary features. The label codes (N80) are never features, and participants with a label code on or before day 0 were not scored. Symptom groups: autonomic, cognitive, fatigue, gastrointestinal, immune, menstrual, mood, pain, sleep, urinary. Of the 94 features, 90 met the eligibility rule for the self-report label (main text, Models). The exceptions were `sex_female`, `cx_ebv_mono`, `px_diag_laparoscopy`, and `px_lap_excision`. Windows, matching rules, and eligibility as in Table S1, except that all diagnosis-code and symptom-group features use both the condition and observation tables.

Feature ID	Label	Window	Codes by system
<i>Demographics</i>			
<code>age_18_39</code>	Age 18–39	Day 0	Age at day 0 (calendar year minus birth year), 18 to 39 years
<code>age_40_49</code>	Age 40–49	Day 0	Age at day 0 (calendar year minus birth year), 40 to 49 years
<code>age_50_59</code>	Age 50–59	Day 0	Age at day 0 (calendar year minus birth year), 50 to 59 years
<code>age_60_69</code>	Age 60–69	Day 0	Age at day 0 (calendar year minus birth year), 60 to 69 years
<code>age_70p</code>	Age 70 or older	Day 0	Age at day 0 (calendar year minus birth year), 70 years or older
<code>sex_female</code>	Female	Day 0	Sex at birth recorded as female
<i>Diagnosis codes</i>			
<code>cx_fatigue_other</code>	Fatigue or malaise coded	5 y	ICD-10-CM: R53.0, R53.1, R53.81, R53.83 Symptom group: fatigue
<code>cx_fatigue_recent</code>	Fatigue or malaise coded in the past year	1 y	ICD-10-CM: R53.0, R53.1, R53.81, R53.83
<code>cx_sleep</code>	Sleep disturbance	5 y	ICD-10-CM: G47.0, G47.1, G47.8, G47.9, F51 Symptom group: sleep
<code>cx_sleep_apnea</code>	Sleep apnea	5 y	ICD-10-CM: G47.33
<code>cx_cognitive</code>	Cognitive symptoms (brain fog, memory)	5 y	ICD-10-CM: R41, G31.84 Symptom group: cognitive
<code>cx_autonomic</code>	Palpitations, POTS, orthostatic hypotension, or dysautonomia	5 y	ICD-10-CM: R00.0, R00.2, I49.8, G90.A, I95.1, G90.9 Symptom group: autonomic
<code>cx_dizzy_syncope</code>	Dizziness or fainting	5 y	ICD-10-CM: R42, R55 Symptom group: autonomic
<code>cx_headache</code>	Headache	5 y	ICD-10-CM: R51, G44 Symptom group: pain
<code>cx_migraine</code>	Migraine	5 y	ICD-10-CM: G43
<code>cx_myalgia_arthralgia</code>	Muscle or joint pain	5 y	ICD-10-CM: M79.1, M25.5, M79.6 Symptom group: pain
<code>cx_fibromyalgia</code>	Fibromyalgia	5 y	ICD-10-CM: M79.7
<code>cx_chronic_pain</code>	Chronic pain	5 y	ICD-10-CM: G89
<code>cx_depression</code>	Depression	5 y	ICD-10-CM: F32, F33 Symptom group: mood
<code>cx_anxiety</code>	Anxiety	5 y	ICD-10-CM: F41 Symptom group: mood
<code>cx_ptsd</code>	Posttraumatic stress disorder	5 y	ICD-10-CM: F43.1
<code>cx_ibs_gi</code>	Irritable bowel or gastrointestinal symptoms	5 y	ICD-10-CM: K58, R10, R11, R14, R19 Symptom group: gastrointestinal

Table S3: (continued)

Feature ID	Label	Window	Codes by system
cx_sore_throat_nodes	Sore throat or enlarged lymph nodes	5 y	ICD-10-CM: J02, R59 Symptom group: immune
cx_fever	Fever	5 y	ICD-10-CM: R50 Symptom group: immune
cx_ebv_mono	Infectious mononucleosis (EBV)	5 y	ICD-10-CM: B27
cx_lyme	Lyme disease	5 y	ICD-10-CM: A69.2
cx_covid	COVID-19	5 y	ICD-10-CM: U07.1
cx_long_covid	Long COVID	5 y	ICD-10-CM: U09.9
cx_eds_hypermobility	Ehlers-Danlos syndrome or hypermobility	5 y	ICD-10-CM: Q79.6, M35.7
cx_mast_cell	Mast cell disorder	5 y	ICD-10-CM: D89.4
cx_hypothyroid	Hypothyroidism	5 y	ICD-10-CM: E03
cx_anemia	Anemia	5 y	ICD-10-CM: D50, D51, D52, D53, D63, D64
cx_vitd_deficiency	Vitamin D deficiency	5 y	ICD-10-CM: E55
cx_autoimmune	Autoimmune or demyelinating disease	5 y	ICD-10-CM: M05, M06, M32, M35, L40, K50, K51, M45, G35
cx_obesity	Obesity	5 y	ICD-10-CM: E66, Z68.3, Z68.4
cx_diabetes	Diabetes	5 y	ICD-10-CM: E10, E11
cx_hypertension	Hypertension	5 y	ICD-10-CM: I10
cx_heart_disease	Coronary disease or heart failure	5 y	ICD-10-CM: I25, I50
cx_cancer	Cancer	5 y	ICD-10-CM: C
cx_pelvic_pain	Pelvic pain	5 y	ICD-10-CM: R10.2 Symptom group: pain
cx_dysmenorrhea	Painful periods (dysmenorrhea)	5 y	ICD-10-CM: N94.4, N94.5, N94.6 Symptom group: menstrual
cx_dyspareunia	Pain with intercourse	5 y	ICD-10-CM: N94.1 Symptom group: pain
cx_menstrual	Heavy or irregular menstrual bleeding	5 y	ICD-10-CM: N92 Symptom group: menstrual
cx_infertility	Female infertility	5 y	ICD-10-CM: N97
cx_ovarian_cyst	Ovarian cyst	5 y	ICD-10-CM: N83.2
cx_fibroids	Uterine fibroids	5 y	ICD-10-CM: D25
cx_pcos	Polycystic ovary syndrome	5 y	ICD-10-CM: E28.2
cx_interstitial_cystitis	Interstitial cystitis or bladder pain	5 y	ICD-10-CM: N30.1 Symptom group: urinary
cx_back_pain	Low back pain	5 y	ICD-10-CM: M54.5 Symptom group: pain
cx_constipation	Constipation	5 y	ICD-10-CM: K59.0 Symptom group: gastrointestinal
<i>Symptom-group counts</i>			
clusters_ge3	Three or more symptom types	5 y	Number of symptom groups recorded: 3 or more
clusters_ge5	Five or more symptom types	5 y	Number of symptom groups recorded: 5 or more
<i>Procedures</i>			
px_tilt	Tilt-table test	5 y	CPT or HCPCS: 93660

Table S3: (continued)

Feature ID	Label	Window	Codes by system
px_sleep_study	Sleep study	5 y	CPT or HCPCS: 95806–95811, G0398, G0399, G0400
px_brain_mri	Brain MRI	5 y	CPT or HCPCS: 70551–70553
px_echo	Echocardiogram	5 y	CPT or HCPCS: 93303–93308, 93312–93318, 93350–93352
px_cardiac_monitor	Holter or event monitor	5 y	CPT or HCPCS: 93224–93229, 93241–93248, 93268–93272
px_exercise_test	Exercise or cardiopulmonary exercise test	5 y	CPT or HCPCS: 94621, 93015–93018
px_neuropsych	Neuropsychological testing	5 y	CPT or HCPCS: 96132, 96133, 96136–96139
px_pt_eval	Physical therapy evaluation	5 y	CPT or HCPCS: 97161–97163
px_pelvic_us	Pelvic or transvaginal ultrasound	5 y	CPT or HCPCS: 76830, 76856, 76857
px_diag_laparoscopy	Diagnostic laparoscopy	5 y	CPT or HCPCS: 49320
px_lap_excision	Laparoscopic excision or fulguration of pelvic lesions	5 y	CPT or HCPCS: 58662
px_hysterectomy	Hysterectomy	5 y	CPT or HCPCS: 58150, 58570–58573
px_pelvic_mri	Pelvic MRI	5 y	CPT or HCPCS: 72195–72197
<i>Laboratory tests</i>			
lab_tsh	Thyroid (TSH) test	5 y	LOINC: measurement concepts whose name contains “Thyrotropin”
lab_ferritin	Ferritin test	5 y	LOINC: measurement concepts whose name contains “Ferritin”
lab_b12	Vitamin B12 test	5 y	LOINC: measurement concepts whose name contains “Cobalamin”
lab_vitamin_d	Vitamin D test	5 y	LOINC: measurement concepts whose name contains “25-hydroxyvitamin D”
lab_ana	Antinuclear antibody (ANA) test	5 y	LOINC: measurement concepts whose name contains “Nuclear Ab”
lab_crp	C-reactive protein test	5 y	LOINC: measurement concepts whose name contains “C reactive protein”
lab_esr	Sedimentation rate (ESR) test	5 y	LOINC: measurement concepts whose name contains “Erythrocyte sedimentation rate”
lab_ebv	Epstein-Barr virus test	5 y	LOINC: measurement concepts whose name contains “Epstein Barr virus”
lab_lyme	Lyme disease test	5 y	LOINC: measurement concepts whose name contains “Borrelia burgdorferi”
lab_cortisol	Cortisol test	5 y	LOINC: measurement concepts whose name contains “Cortisol”
lab_ck	Creatine kinase test	5 y	LOINC: measurement concepts whose name contains “Creatine kinase”
lab_ca125	CA-125 test	5 y	LOINC: measurement concepts whose name contains “Cancer Ag 125”
<i>Medications</i>			

Table S3: (continued)

Feature ID	Label	Window	Codes by system
<code>rx_stimulant</code>	Stimulant or wakefulness drug	5 y	RxNorm ingredients: modafinil, armodafinil, methylphenidate, amphetamine, dextroamphetamine, lisdexamfetamine
<code>rx_naltrexone</code>	Naltrexone	5 y	RxNorm ingredients: naltrexone
<code>rx_ssri_snri</code>	SSRI or SNRI antidepressant	5 y	RxNorm ingredients: sertraline, fluoxetine, citalopram, escitalopram, paroxetine, venlafaxine, duloxetine, desvenlafaxine
<code>rx_tricyclic_lowdose</code>	Amitriptyline, nortriptyline, or cyclobenzaprine	5 y	RxNorm ingredients: amitriptyline, nortriptyline, cyclobenzaprine
<code>rx_gabapentinoid</code>	Gabapentin or pregabalin	5 y	RxNorm ingredients: gabapentin, pregabalin
<code>rx_sleep_aid</code>	Sleep medication	5 y	RxNorm ingredients: zolpidem, trazodone, eszopiclone
<code>rx_beta_blocker</code>	Beta blocker	5 y	RxNorm ingredients: propranolol, metoprolol, atenolol, bisoprolol, nadolol, carvedilol
<code>rx_pots</code>	POTS medication	5 y	RxNorm ingredients: ivabradine, midodrine, fludrocortisone, pyridostigmine
<code>rx_opioid</code>	Opioid	5 y	RxNorm ingredients: oxycodone, hydrocodone, tramadol, morphine
<code>rx_hormonal</code>	Hormonal contraceptive or progestin	5 y	RxNorm ingredients: norethindrone, levonorgestrel, drospirenone, medroxyprogesterone, etonogestrel, desogestrel, norgestimate
<code>rx_gnrh</code>	GnRH agonist or antagonist	5 y	RxNorm ingredients: leuprolide, elagolix, relugolix
<code>rx_nsaid</code>	NSAID pain reliever	5 y	RxNorm ingredients: naproxen, ibuprofen, meloxicam, diclofenac, ketorolac
<code>rx_aromatase</code>	Aromatase inhibitor	5 y	RxNorm ingredients: letrozole, anastrozole
<i>Health care use</i>			
<code>util_out_r_ge6</code>	6+ outpatient visits in the past year	1 y	Outpatient visits: 6 or more
<code>util_out_r_ge12</code>	12+ outpatient visits in the past year	1 y	Outpatient visits: 12 or more
<code>util_ed_r</code>	Emergency visit in the past year	1 y	Emergency visits: 1 or more
<code>util_ip_h</code>	Hospital stay in the past 5 years	5 y	Inpatient visits: 1 or more

Table S4: TRIPOD+AI checklist. Items of the TRIPOD+AI reporting guideline (Collins et al., BMJ 2024;385:e078378), paraphrased, with the place in the manuscript where we report each. Methods refers to the Materials and Methods section, and section and subsection names are those of the main text. Not reported: the manuscript does not address the item. Notes in parentheses identify partial reporting.

Item	Topic	Checklist item (paraphrased)	Where reported
<i>Title and abstract</i>			
1	Title	Identify the study as developing or evaluating a multivariable prediction model, with the target population and the outcome	Title (names the conditions, training labels, and cohort but does not use the term prediction model)
2	Abstract	Follow the TRIPOD+AI for Abstracts checklist	Abstract (calibration and registration not reported)
<i>Introduction</i>			
3a	Background	Health care context (diagnostic or prognostic), rationale, and existing models	Introduction, paragraphs 1–3
3b	Background	Target population, intended purpose in the care pathway, and intended users	Introduction (final paragraph), Statement of significance, and Discussion (paragraph 4)
3c	Background	Known health inequalities between sociodemographic groups	Not reported
4	Objectives	Objectives, including whether the study develops or evaluates a model, or both	Introduction, final paragraph
<i>Methods</i>			
5a	Data	Data sources for development and evaluation, rationale, and representativeness	Methods: Data source and population
5b	Data	Dates of data collection, start and end of accrual, and end of follow-up	Methods: Data source and population (EHR visit periods and data cutoff), Labels and evaluation targets (follow-up to the data cutoff)
6a	Participants	Study setting, and number and location of centers	Methods: Data source and population (number of contributing health systems not stated)
6b	Participants	Eligibility criteria	Methods: Data source and population
6c	Participants	Treatments received and how they were handled	Methods: Features (medication exposures are predictors)
7	Data preparation	Preprocessing and quality checks, and whether they were similar across sociodemographic groups	Methods: Features and Models (rounding adjustment), with checks by sociodemographic group not reported
8a	Outcome	Outcome and time horizon, how and when assessed, rationale, and consistency across sociodemographic groups	Methods: Labels and evaluation targets, Limitations (no fixed horizon, survey instance of the answer not checked)
8b	Outcome	Qualifications and demographics of outcome assessors, if assessment is subjective	Not reported (labels are survey answers and recorded codes)
8c	Outcome	Blinding of outcome assessment	Not reported

Table S4: (continued)

Item	Topic	Checklist item (paraphrased)	Where reported
9a	Predictors	Choice of initial predictors and any preselection before model building	Methods: Features, Models (eligibility rule and number of eligible features), and Tables S1–S3
9b	Predictors	Definition of all predictors, including how and when they were measured	Methods: Features and Tables S1–S3
9c	Predictors	Qualifications and demographics of predictor assessors, if measurement is subjective	Not reported (predictors are recorded EHR events)
10	Sample size	How the study size was arrived at, and whether it was sufficient	Not reported
11	Missing data	Handling of missing data and reasons for omitting data	Methods: Labels and evaluation targets (skipped answer counted as a report and no answer treated as not reporting), Features (absent record coded 0), and Limitations
12a	Analytical methods	Use and partitioning of the data for development and evaluation	Methods: Models (80% training split), Evaluation (20% validation split), and Limitations (reproducibility of the split)
12b	Analytical methods	Handling of predictors (functional form, rescaling, transformation, standardization)	Methods: Features (binary indicators) and Models (standardized features for the elastic net)
12c	Analytical methods	Model type and rationale, model-building steps, hyperparameter tuning, and internal validation	Methods: Models and Evaluation
12d	Analytical methods	Handling and quantification of heterogeneity across clusters	Not reported for the 3 conditions. Methods: Secondary analysis (held-out sites, defined by the EHR source of the index event, Long COVID cohort only)
12e	Analytical methods	Performance measures and plots, with rationale, and comparison of models	Methods: Evaluation and Limitations (no comparison on a common feature set and no simple baselines)
12f	Analytical methods	Model updating arising from the evaluation	Not reported
12g	Analytical methods	How predictions were calculated for the evaluation	Methods: Models (score formula)
13	Class imbalance	Class imbalance methods, rationale, and any recalibration	Methods: Models (no class weighting)
14	Fairness	Approaches used to address model fairness, and rationale	Not reported
15	Model output	Model output, and how any classification thresholds were chosen	Methods: Models (mapping to probabilities) and Evaluation (top 5% and 10% of scores)
16	Training vs evaluation	Differences between development and evaluation data in setting, eligibility, outcome, and predictors	Methods: Models, Evaluation (random split of one cohort), and Limitations
17	Ethical approval	Ethics committee, and informed consent or its waiver	Methods: Disclosure control and ethics (program-level consent and no project-specific review)

Table S4: (continued)

Item	Topic	Checklist item (paraphrased)	Where reported
<i>Open science</i>			
18a	Funding	Source of funding and role of the funders	Funding (no specific funding)
18b	Conflicts of interest	Conflicts of interest and financial disclosures for all authors	Declaration of competing interest
18c	Protocol	Where the protocol can be accessed, or that none was prepared	Not reported
18d	Registration	Registration details, or that the study was not registered	Not reported
18e	Data sharing	Availability of the study data	Data and code availability
18f	Code sharing	Availability of the analysis code	Data and code availability (model files, specification, and synthetic test patients for the chronic fatigue and fibromyalgia models, plus analysis code to be released in a public repository, URL at posting)
<i>Patient and public involvement</i>			
19	Patient and public involvement	Involvement in design, conduct, reporting, interpretation, or dissemination, or a statement of none	Not reported
<i>Results</i>			
20a	Participants	Flow of participants, numbers with and without the outcome, and follow-up time	Results: Cohorts (numbers with and without each label) and Limitations (exclusions at each eligibility step and follow-up time not reported)
20b	Participants	Characteristics overall and by setting, including key predictors, outcome events, and missing data	Results: Cohorts and Features associated with self-report (selected characteristics, no characteristics table)
20c	Participants	Comparison of predictor and outcome distributions between development and evaluation data	Not reported
21	Model development	Numbers of participants and outcome events in each analysis	Results: Cohorts and Table 1
22	Model specification	Full model, to allow predictions in new individuals and third-party evaluation, with any access restrictions	Methods: Models (score formula), Data and code availability (exported model files), with weights not tabulated
23a	Model performance	Performance with confidence intervals, including key subgroups	Results: Discrimination by training label (AUROC for self-report by sex, without intervals, and no other subgroups) and Table 1
23b	Model performance	Heterogeneity of performance across clusters, if examined	Results: Binary features in Long COVID (held-out sites, defined by the EHR source of the index event, Long COVID cohort only)
24	Model updating	Results of any model updating	Not reported
<i>Discussion</i>			

Table S4: (continued)

Item	Topic	Checklist item (paraphrased)	Where reported
25	Interpretation	Overall interpretation, including fairness, in light of the objectives and previous studies	Discussion, paragraphs 1–3 (fairness not discussed)
26	Limitations	Limitations and their effects on bias, uncertainty, and generalizability	Discussion: Limitations
27a	Usability	How poor-quality or unavailable input data should be assessed and handled in implementation	Methods: Features (absent records coded 0) and Limitations (mapping to FHIR not validated)
27b	Usability	Whether users interact with the input data or the model, and the expertise required	Discussion, paragraph 4 (expertise not stated)
27c	Usability	Next steps for research on applicability and generalizability	Discussion, paragraphs 3–4, and Conclusion