

Title: Predictors of Opioid Use in Individuals with Pain

Authors: Azin Zare^{1,2}, Matt Fillingim^{1,3}, Christophe Tanguay-Sabourin^{1,3,4}, Gianluca Guglietti^{1,3}, Jax Norman^{1,2}, Lindsay Neuert^{1,2}, Lise Hobeika^{1,5,6}, Marc O. Martel^{1,2,3}, Etienne Vachon-Presseau^{1,2,3}

Affiliations:

1. Alan Edwards Centre for Research on Pain, McGill University, Montreal, Quebec, Canada
2. Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Quebec, Canada
3. Department of Anaesthesia, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
4. Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada
5. Université Paris Cité, Institut Pasteur, AP-HP, INSERM, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, Paris, France.
6. Sorbonne Université, Institut du Cerveau—Paris Brain Institute—ICM, INSERM, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France.

Corresponding authors:

Azin Zare: azin.zare@mcgill.ca;

Etienne Vachon-Presseau: etienne.vachon-presseau@mcgill.ca

Supplementary methods

UK Biobank

The UK Biobank (UKBB) is a large-scale, prospective and ongoing study designed to investigate genetic factors and lifestyle determinants of a wide range of diseases in middle-aged and older adults. To reach its target sample size of ~500,000 participants, more than 9 million invitations were sent to UK National Health Service (NHS) registrants aged 40–69 years who lived within a reasonable distance of an assessment center. Baseline recruitment and data collection were conducted between 2006 and 2010, with 502,494 participants consenting to join the study across 22 assessment centers in Scotland, England, and Wales ¹.

Following the baseline phase, subsets of participants were invited for follow-up visits. This study uses datasets collected at the baseline in-person assessment visit (2006–2010) and the 9-year follow-up visit (2014–2020). During the baseline in-person assessment visit, data were collected from 503,317 participants through touchscreen questionnaires covering sociodemographic, lifestyle, psychosocial factors, and health history, as well as verbal interviews conducted by trained nurses on early life factors, employment, medical conditions, and medications. Physical measurements and biological samples were also obtained. During the follow-up visit around 9 years later a subset of 49,001 participants underwent brain scans and completed the same touchscreen questionnaires as in the baseline visit.

Participants with pain

Participants were asked whether they experienced any of the following pain conditions in the past month that interfered with their usual activities: "headache," "facial pain," "neck or shoulder pain," "back pain," "stomach or abdominal pain," "hip pain," "knee pain," "pain all over the body," or "prefer not to answer" (data-field: 6159). Selecting "pain all over the body" was exclusive, preventing the selection of other options. If participants reported experiencing pain, they were further asked whether the pain persisted for more than three months. Based on the International Association for the Study of Pain (IASP) criteria ², those reporting pain lasting more than three months at any site were classified as having "chronic pain." In contrast, participants who reported pain at one of the seven regional sites but indicated it lasted less than three months were classified as experiencing "acute pain."

Participants were asked whether a doctor had ever diagnosed them with cancer (data-field 2453). Those responding “yes” completed a verbal interview with a trained nurse to provide details on the number of self-reported cancer diagnoses (data-field 134). Participants with cancer were excluded from the analyses.

Opioid use

Medication use was self-reported by participants at both the baseline and follow-up visits, defined as the regular use of prescription drugs—taken on most days of the week over the past four weeks (data-field: 2492). Following this, participants were interviewed by a trained nurse, who recorded the names of the medications used (data-field: 20003), using either the generic or trade names provided. This assessment captured only regular, ongoing medications and did not include short-term or as-needed treatments.

For this study, the recorded medications were classified using the Anatomical Therapeutic Chemical (ATC) classification system developed by the World Health Organization (WHO)³. The coding approach followed the stratification used by Wu et al. (2019)⁴, with a focus on ATC code N02A (opioid analgesics). To ensure broader coverage of relevant opioid-containing compounds, we also included ATC code R05DA04 (codeine). Opioid use was identified at each time point using medications classified under either of these two ATC codes. A full list of included medications is presented in Supplementary Table 1.

Medical diagnoses

The UKBB provides curated data fields identifying the earliest recorded instance of diagnostic codes (Category ID: 1712) derived from self-reports, primary care, hospital inpatient records, and death registries. These diagnoses are mapped to three-digit ICD-10 codes and organized into 16 subcategories corresponding to ICD-10 chapters (Chapters I–XVII, excluding Chapter II: Neoplasms). To determine whether a diagnosis predated a study visit, we compared the date of first recorded diagnosis to the participant’s visit date at both baseline and 9-year follow-up. For this study, 15 of the 16 categories were included; the congenital and chromosomal abnormalities category was excluded due to limited sample size.

Opioid-related disorders

Opioid-related disorders were identified using ICD-10 code F11, corresponding to mental and behavioral disorders due to opioid use.

Exclusion and inclusion criteria

Responses coded as “prefer not to answer” or “do not know” were recoded as missing values. Participants with more than 20% missing data in selected features used to train our models at baseline, or missing data on pain location, were excluded to ensure accurate classification of pain sites (< 2.5% of participants). The study focused exclusively on individual with chronic non-cancer pain (CNCP). To enhance generalizability, no other exclusion criteria were applied, resulting in a final sample of 195,756 participants.

All of Us Research Program

The All of Us Research Program (AoU) aims to enroll at least one million diverse participants across the United States, with a strong emphasis on including individuals from populations historically underrepresented in biomedical research. The data integrates participant-reported data from multiple surveys—including basic demographic questionnaires, the COVID-19 Participant Experience (COPE) survey, and the Overall Health Questionnaire, alongside physical measurements and electronic health record (EHR) data encompassing clinically diagnosed medical conditions. For this study, we used data from the Registered and Controlled Tier Dataset the available through the AoU Researcher Workbench. To be included in the out analysis, participants were required to meet the following criteria: 1. Completion of the Overall Health Questionnaire, and 2. Completion of the COVID-19 Participant Experience (COPE) survey.

Participants with pain

Pain-related information was collected via the Overall Health Questionnaire, which covers topics such as general health, daily activities, and women’s health. Participants were asked “In the past 7 days, how would you rate your pain on average?”, with responses recorded on a numerical scale from 0 (no pain) to 10 (worst pain imaginable). These variables were respectively used to train non-pain and pain models, as reported in **Figure 2**.

Opioid use

Opioid use was assessed using the COPE survey, which aimed to capture how participants' experiences and health were affected during the COVID-19 pandemic. Respondents were asked: "In the past month, have you used any of the following drugs? Select all that apply." Options included: "Prescription opioids (for example, fentanyl, oxycodone [OxyContin, Percocet], hydrocodone [Vicodin], methadone, buprenorphine, etc.)."

Opioid misuse

Following the opioid use item in the COPE survey, participants were asked: "Did you use prescription opioids in any way a doctor did not direct you to use it?" A "yes" response was classified as opioid misuse within the AoU cohort.

Opioid-related disorders

Participants with opioid use disorder were identified based on EHR-confirmed diagnoses. Diagnoses were determined according to SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms) codes related to opioid-related disorders.

Machine learning analyses

In the UK Biobank, the sample was divided into a training set of 178,715 participants and a testing set of 17,041 participants with longitudinal follow-up data. Models were trained on the training set, while the testing set was held out to evaluate performance in out-of-sample participants. The longitudinal data also enabled assessment of whether the models could predict opioid initiation and discontinuation at follow-up. The All of Us Research Program served as an external validation cohort, where models were trained using the cross-validation procedure described below. Unlike the UK Biobank, no participants were held out for longitudinal analyses. The pain and non-pain variables used to train the models are described in the **Supplementary Table 3** for the UK Biobank and **Supplementary Table 4** for the All of Us Research Program.

Statistical analyses were conducted using Python 3.8.8 and R 4.1.1. Predictive machine learning models, including cross-validated Lasso regression, were developed to predict opioid use. Lasso regression was used for variable selection and to prevent multicollinearity by shrinking coefficients towards zero, simplifying the model while minimizing prediction error. The shrinkage strength was controlled by the alpha parameter, excluding unnecessary variables. Features were

resampled using the Synthetic Minority Oversampling Technique (SMOTE) to address data imbalance. Each model underwent 5 iterations with unique random states (5-times repeated 5-fold nested cross-validation) to ensure reliability, and confidence intervals for model performance metrics were calculated.

Model evaluation involved calculating AUC scores, log-probabilities of predictions, and ROC curves for each cross-validation fold. Mean true positive rates (TPRs) were interpolated to a standard false positive rate (FPR) grid for visualization. Coefficients and AUC scores were aggregated across iterations to summarize overall model performance in left out participants. We used bootstrap resampling with 1,000 iterations to indicate the estimated error in the Cohen's d and ROC-AUC effect sizes.

Structure coefficients

To address potential instability and bias in feature weights from multivariate models, we used structure coefficients to identify features individually linked to opioid use. These coefficients reflect the individual association of features. For each participant, structure coefficients were calculated by correlating the predicted probabilities from the logistic regression model with the actual values.

Pseudo R-squared (R^2)

To evaluate the relative contribution of different feature domains in predicting opioid-related outcomes, we computed the pseudo R-squared (McFadden's R^2) for each domain-specific model. The 36 selected non-pain features were categorized into eight domains as: functional factors, socioeconomic, mood, substance use, sleep, demographic, occupational, neuroticism. Domain-specific models were fit separately using only the features corresponding to each domain.

Deviance explained

Deviance explained (D^2) was used to quantify the extent to which pain and non-pain predictors, individually and jointly, accounted for variance in opioid use as the primary outcome of interest. Predicted probabilities were derived from logistic regression models trained on three sets of predictors: pain, non-pain, and a combination of both. Each model's deviance was calculated by comparing its fit to that of a hypothetical model with perfect fit to the observed data.

The deviance explained was then computed by normalizing each model's deviance against that of a null model containing only an intercept, which served as a baseline representing no predictive value. This approach followed a modified version of the methodology described by Dinga et al. (2020)⁵.

To disentangle the contributions of each feature set, the unique deviance explained was computed as the additional variance explained by one set of predictors beyond what was accounted for by the other. The shared contribution reflected the overlapping explanatory power of pain and non-pain predictors in modeling opioid use.

Handling missing data and standardization of the variables

Responses coded as “prefer not to answer” or “do not know” were recoded as missing values. To address missing data and minimize bias, a Bayesian ridge regression model with a median prior was used for imputation, yielding results consistent with a median-only imputation approach. Features were then standardized by centering the mean to zero and scaling the variance to one, ensuring no feature dominated due to variance magnitude. The same process—exclusion, imputation, and standardization—was applied separately to the testing dataset to avoid data leakage.

Open code on Github

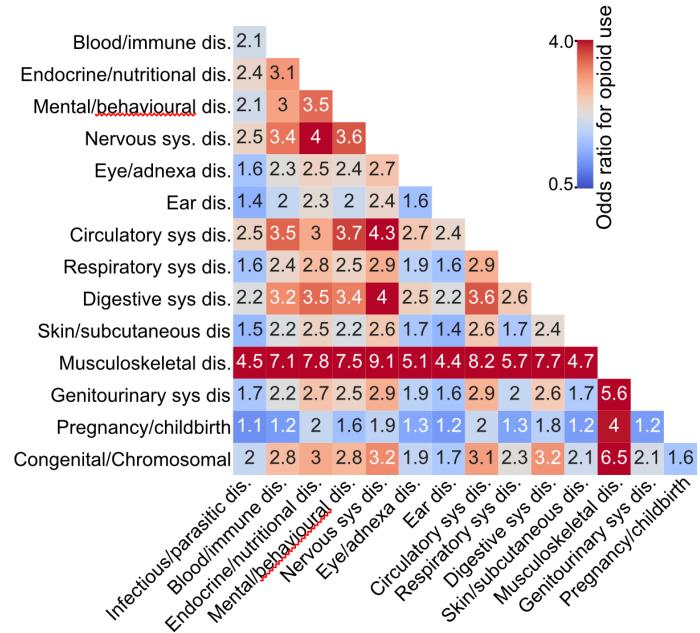
For the complete code used to perform the analyses in the main text of this paper please refer to our GitHub repository at: <https://github.com/EVPlab>

Supplementary tables and figures

	Prescription opioids	Number	Percentage (%)
Weak opioids	Combined	13,904	63.75
	Codeine	2,020	9.26
	Dihydrocodeine	1,444	6.62
	Dextropropoxyphene	314	1.43
	Meptazinol	61	0.28
	Pethidine	21	0.09
Strong opioids	Tramadol	4,646	21.3
	Morphine	360	1.65
	Buprenorphine	306	1.40
	Fentanyl	220	1
	Oxycodone	166	0.76

Supplementary Table 1 Regular prescription opioids reported by the chronic non-cancer pain patients. Combined opioids comprise opioids in combination with a non-opioid analgesic.

Health Conditions' Co-occurrences and Opioid Use in the UKBB

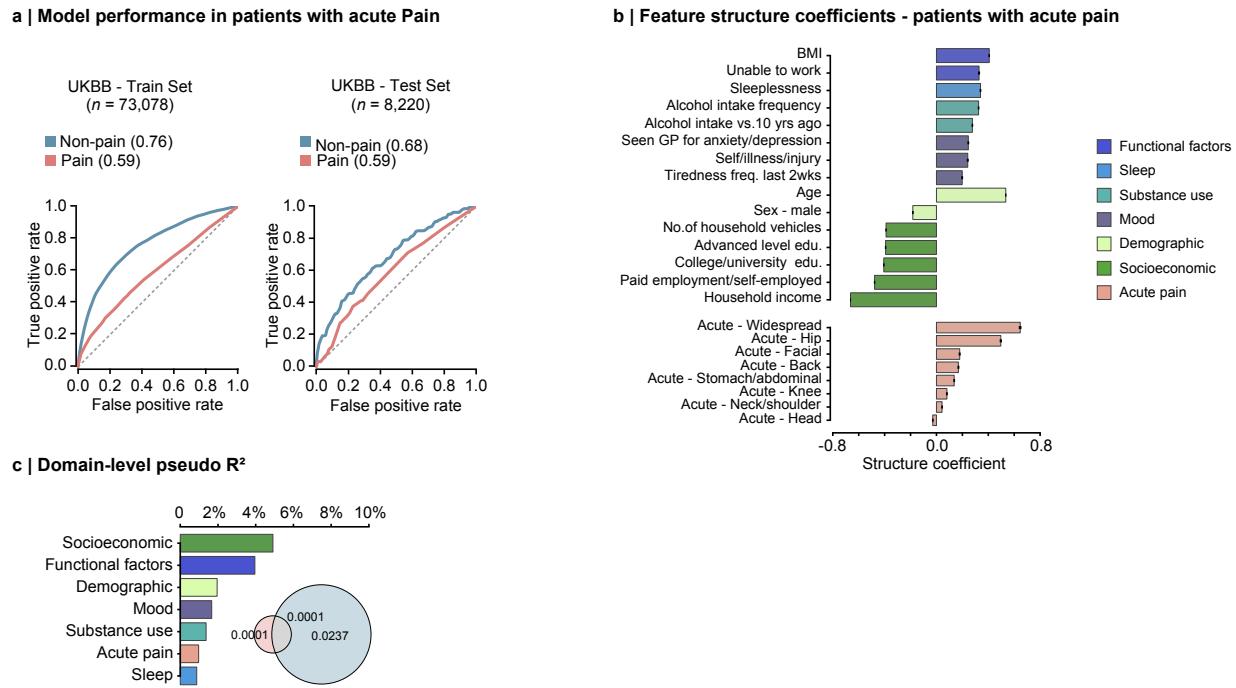


Supplementary Figure 1. Heatmap shows the co-occurrence of diagnoses for opioid use. ORs are calculated for opioid use at the co-occurrence of two major categories compared to opioid use at other categories. The odds of opioid use were highest among participants with musculoskeletal disorders, but were also elevated for other conditions, including nervous system and digestive system disorders.

	UKBB	
	No opioid use (n = 79,764)	Opioid Users (n = 1,534)
Sex		
Female	40,110 (50.3%)	893 (58.2%)
Male	39,654 (49.17)	641(41.8%)
Age		
Mean (SD)	55.87(8.30)	59.1 (7.95)
Median [Min, Max]	56.3 [38.8, 71.2]	60.9 [40.5, 70.3]
Ethnicity		
White	73,893 (92.6%)	1,432 (93.4%)
Asian	2,391 (3.0%)	29 (1.9%)
Black	1,752 (2.2%)	42 (2.7%)
Other	944 (1.2%)	16 (1.0%)
Mixed	546 (0.7%)	7 (0.5%)
Missing	238 (0.3%)	8 (0.5%)
Education		
^a Higher education	31,162 (39.1%)	362 (23.6%)
^b Vocational/Post-secondary	5,364 (6.7%)	122 (8.0%)
^c Secondary	30,083 (37.7%)	506 (33.0%)
None	11,776 (14.8%)	501 (32.7%)
Missing	1,379 (1.7%)	43 (2.8%)
Employment		
Employed	51,408 (64.5%)	619 (40.4%)
Retired	22,166 (27.8%)	666 (43.4%)
Unable to work-Sickness/Disability	1,204 (1.5%)	154 (10.0%)
^d Unemployed	4,249 (5.3%)	77 (5.0%)
None of above	465 (0.6%)	9 (0.6%)
Missing	272 (0.3%)	9 (0.6%)
^eHousehold income		
Low	30,205 (37.9%)	824 (53.7%)
Middle	34,111 (42.8%)	357 (23.3%)
High	4,058 (5.1%)	25 (1.6%)
Prefer not to answer	7,711 (9.7%)	190 (12.4%)
Missing	3,679(4.6%)	138 (9.0%)
Acute pain type		
Back pain	27,257 (34.2%)	543 (35.4%)
Knee pain	15,375 (19.3%)	310 (20.2%)
Neck/shoulder pain	21,735 (27.2%)	416 (27.1%)
Headaches	27,984 (35.1%)	516 (33.6%)
Hip pain	6,281 (7.9%)	193 (12.6%)
Stomach/abdominal pain	9,816 (12.3%)	206 (13.4%)
Facial pain	1,917 (2.4%)	48 (3.1%)
Widespread pain	1,478 (1.9%)	120 (7.8%)

^aHigher education refers to college or university degree and other professional qualifications in the UKBB. ^bVocational/post-secondary education includes NVQs, HNDs, HNCs, or their equivalents in the UKBB. ^cSecondary education includes A levels/ AS levels, O levels/ GCSEs or equivalent, and CSEs or equivalent in the UKBB. ^dUnemployed individuals include unemployed, taking care of home and/or family, unpaid or voluntary work, or full- or part-time students. ^eIncome levels are classified as follows: low including less than £18,000 or between £18,000 and £30,999, middle including £31,000 to £51,999 and £52,000 to £100,000; and high referring to more than £100,000 per year in the UKBB.

Supplementary Table 2. Characteristics of participants in the UK Biobank (UKBB) stratified by opioid use in participants with acute pain. Mean, standard deviation (SD), and median are reported for the continuous variables (Age and BMI), and percentages are reported for categorical variables. Missing data are noted where applicable.



Supplementary Figure 2. Non-pain and pain models predicting opioid use in patients with acute pain. **a)** Receiver Operating Characteristic (ROC) curves display the classification performance of two LASSO logistic regression models: a pain model and a non-pain model, evaluated on the UKBB training set (left), test set (right). The models were trained to predict opioid use based on domain-specific features, with feature selection determined by non-zero LASSO coefficients following regularization. **B)** structure coefficients show the univariate associations between each feature and the model prediction. **c)** Domain-level variance explained ($pseudo R^2$) is shown for each feature category in the non-pain model (eight domains) and the corresponding pain-related domains (chronic and acute pain). Adjacent Venn diagrams illustrate the proportion of deviance explained (D^2) uniquely by the non-pain (blue, right circle) and pain (pink, left circle) feature sets, as well as the shared variance (overlap).

Feature	Associated Question	UK Biobank data field
Age	Age when attended assessment center	21003
Sex	Sex of participant	31
Body Mass Index (BMI)	Body Mass Index (BMI)	21001
Ethnicity White	Ethnic background	21000
Ethnicity Mixed	Ethnic background	21000
Ethnicity Asian	Ethnic background	21000
Ethnicity Black	Ethnic background	21000
Ethnicity Other	Ethnic background	21000
Number In Household	Including yourself, how many people are living together in your household? (Include those who usually live in the house such as students living away from home during term, partners in the armed forces or professions such as pilots)	709
Live With Partner	How are the other people who live with you related to you? (You can select more than one answer) (1= Husband, wife or partner , 2=Son and/or daughter, 3=Brother and/or sister, 4=Mother and/or father, 5=Grandparent, 6=Grandchild, 7=Other related, 8=Other unrelated)	6141
Live With Children	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2= son and/or daughter , 3=Brother and/or sister, 4=mother and/or father, 5=Grandparent, 6=Grandchild, 7=Other related, 8=Other unrelated)	6141
Live With Sibling	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2= son and/or daughter, 3=Brother and/or sister , 4=mother and/or father, 5=Grandparent, 6=Grandchild, 7=Other related, 8=Other unrelated)	6141
Live With Parent	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2= son and/or daughter, 3=Brother and/or sister, 4=mother and/or father , 5=Grandparent, 6=Grandchild, 7=Other related, 8=Other unrelated)	6141
Live With Grandparent	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2= son and/or daughter, 3=Brother and/or sister, 4=mother and/or father, 5=Grandparent , 6=Grandchild, 7=Other related, 8=Other unrelated)	6141
Live With Grandchild	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2= son and/or daughter, 3=Brother and/or sister, 4=mother and/or father, 5=Grandparent, 6=Grandchild , 7=Other related, 8=Other unrelated)	6141
Live With Other Related	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2 son and/or daughter,	6141

	3=Brother and/or sister, 4=mother and/or father, 5=Grandparent, 6=Grandchild, 7=Other related , 8=Other unrelated)	
Live With Other Unrelated	How are the other people who live with you related to you? (You can select more than one answer) (1=Husband, wife or partner, 2 son and/or daughter, 3=Brother and/or sister, 4=mother and/or father, 5=Grandparent, 6=Grandchild, 7=Other related , 8=Other unrelated)	6141
Number Vehicles Household	How are the other people who live with you related to you? (You can select more than one answer)	6141
Income Household Pre Tax	What is the average total income before tax received by your HOUSEHOLD? (1=Less than 18,000, 2=18,000-30,999, 3=31,000-51,999, 4=52,000-100,000, 5=Greater than 100,000)	100295
Current employment status: Paid Employment or Self Employed	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed , 2=Retired, 3=Looking after home and/or family, 4=Unable to work because of sickness or disability, 5=Unemployed, 6=Doing unpaid or voluntary work, 7=Full or part-time student, -7=None of the above)	100295
Current employment status: Retired	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed , 2=Retired , 3=Looking after home and/or family, 4=Unable to work because of sickness or disability, 5=unemployed, 6=Doing unpaid or voluntary work, 7=Full or part-time student, -7=None of the above)	100295
Current employment status: Look After Home Family	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed , 2=Retired, 3=Looking after home and/or family , 4=Unable to work because of sickness or disability, 5=unemployed, 6=Doing unpaid or voluntary work, 7=Full or part-time student, -7=None of the above)	100295
Current employment status: Unable Work Due Sickness Disability	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed , 2=Retired, 3=Looking after home and/or family, 4=Unable to work because of sickness or disability , 5=unemployed, 6=Doing unpaid or voluntary work, 7=Full or part-time student, -7=None of the above)	100295
Current employment status: Unemployed	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed , 2=Retired, 3=Looking after home and/or family, 4=Unable to work because of sickness or disability, 5=unemployed , 6=Doing unpaid or voluntary work, 7=Full or part-time student, -7=None of the above)	100295
Current employment status: Voluntary Work	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed , 2=Retired, 3=Looking after home and/or family, 4=Unable to work because of	100295

	sickness or disability, 5=unemployed, 6=Doing unpaid or voluntary work ,7=Full or part-time student, -7=None of the above)	
Current employment status: Full Part Time Student	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed, 2=Retired,3=Looking after home and/or family, 4=Unable to work because of sickness or disability, 5=unemployed, 6=Doing unpaid or voluntary work, 7=Full or part-time student , -7=None of the above)	100295
Current employment status: None Of Employment Proposed	Which of the following describes your current situation? (You can select more than one answer) (1=In paid employment or self-employed, 2=Retired , 3=Looking after home and/or family, 4=Unable to work because of sickness or disability, 5=unemployed, 6=Doing unpaid or voluntary work, 7=Full or part-time student, -7=None of the above)	100295
Education: College or University	Which of the following qualifications do you have? (You can select more than one) (1=College or University degree , 2=A levels/AS level or equivalent, 3=O Levels/GCSEs or equivalent, 4 = Other Professional Qualifications)	100305
Education: Advanced Level	Which of the following qualifications do you have? (You can select more than one) (1=College or University degree, 2=A levels/AS level or equivalent , 3=O Levels/GCSEs or equivalent, 4=Cses or equivalent, 5=Nvq or HND or HNC or equivalent, 6=Other Professional Qualifications)	100305
Education: Ordinary Level	Which of the following qualifications do you have? (You can select more than one) (1=College or University degree, 2=A levels/AS level or equivalent, 3=O Levels/GCSEs or equivalent , 4=Cses or equivalent, 5=Nvq or HND or HNC or equivalent, 6=Other Professional Qualifications)	100305
Education: Certificate Secondary Education	Which of the following qualifications do you have? (You can select more than one) (1=College or University degree, 2=A levels/AS level or equivalent, 3=O Levels/GCSEs or equivalent, 4=Cses or equivalent , 5=Nvq or HND or HNC or equivalent, 6=Other Professional Qualifications)	100305
Education: Practical Career Diploma	Which of the following qualifications do you have? (You can select more than one) (1=College or University degree, 2=A levels/AS level or equivalent, 3=O Levels/GCSEs or equivalent, 4=Cses or equivalent, 5=Nvq or HND or HNC or equivalent , 6=Other Professional Qualifications)	100305
Education: Other Professional Qualification	Which of the following qualifications do you have? (You can select more than one) (1=College or University degree, 2=A levels/AS level or equivalent, 3=O Levels/GCSEs or equivalent, 4=Cses or equivalent 5=Nvq or HND or HNC or equivalent, 6=Other Professional Qualifications)	100305
Sleep Duration	About how many hours sleep do you get in every 24 hours? (please include naps)	1160

Nap During Day	Do you have a nap during the day? (1=Never/rarely, 2=Sometimes, 3=Usually)	1190
Difficulty Waking Up	On an average day, how easy do you find getting up in the morning? (1=Not at all easy, 2=Not very easy, 3=fairly easy, 4=Very easy)	1170
Late Chronotype	Do you consider yourself to be? (1=Definitely a 'morning' person, 2=More of a 'morning' than 'evening' person, 3=More an 'evening' than a 'morning' person, 4=Definitely an 'evening' person)	1180
Sleeplessness	Do you have trouble falling asleep at night or do you wake up in the middle of the night? (1=Never/rarely, 2=Sometimes, 3=Usually)	1200
Narcolepsy	"How likely are you to doze off or fall asleep during the daytime when you don't mean to? (e.g. when working, reading or driving)" (0=Never/rarely, 1=Sometimes, 2=Often, 3=All of the time)	1220
Ever Smoked	"Current tobacco smoking" (Field 1239) and "Past tobacco smoking" (Field 1249). Individual classed as Ever smoker if Current tobacco smoking=Most days (1) or occasionally (2) OR Past tobacco smoking=Most days (1) or occasionally (2) or tried once or twice (3).	20160
Daily Smoker	Do you smoke tobacco now? (1=Yes, on most or all days , 2=Only occasionally, 0=No)	1239
Occasional Smoker	Do you smoke tobacco now? (1=Yes, on most or all days, 2=Only occasionally , 0=No)	1239
Previous Smoker	Smoking status (0=Never, 1=Previous, 2=Current)	20116
Past Tobacco Smoking	In the past, how often have you smoked tobacco? (1=Smoked on most days, 2=Smoked occasionally, 3=Just tried once or twice, 4=I have never smoked)	1249
Smoking Household	Does anyone in your household smoke? (1=Yes, one household member smokes, 2=Yes, more than one household member smokes, 0=No)	1259
Exposure Tabacco Home	At home, about how many hours per WEEK are you exposed to other people's tobacco smoke?	1269
Alcohol Intake Frequency	About how often do you drink alcohol? (1=Daily or almost daily, 2=Three or four times a week, 3=Once or twice a week, 4=One to three times a month, 5=Special occasions only, 6=Never)	1558
Alcohol Intake Versus 10 Yrs	Compared to 10 years ago, do you drink? (1=More nowadays, 2=About the same, 3=Less nowadays_	1628
Previous Drinker	Alcohol drinker status (0=Never, 1=Previous , 2=Current)	20117
Never Drank	Alcohol drinker status (0=Never , 1=Previous, 2=Current)	20117
Total Neuroticism	Derived from the 12 neurotic behaviors as a sum score of the total "yes" answers to these questions	20127
Mood Swings	Does your mood often go up and down? (Yes/No)	1920

Miserableness	Do you ever feel 'just miserable' for no reason? (Yes/No)	1930
Irritability	Are you an irritable person? (Yes/No)	1940
Sensitivity Hurt Feelings	Are your feelings easily hurt? (Yes/No)	1950
Fed Up Feelings	Do you often feel 'fed-up'? (Yes/No)	1960
Nervous Feelings	Would you call yourself a nervous person? (Yes/No)	1970
Worry Anxious Feelings	Do you worry too long after an embarrassing experience? (Yes/No)	1980
Tense Strung	Would you call yourself tense or 'highly strung'? (Yes/No)	1990
Worry Long After Embarrassment	Do you worry too long after an embarrassing experience? (Yes/No)	2000
Suffer From Nerves	Do you suffer from 'nerves'? (Yes/No)	2010
Loneliness Isolation	Do you often feel lonely? (Yes/No)	2020
Guilty Feelings	Are you often troubled by feelings of guilt? (Yes/No)	2030
Risk Taking	Would you describe yourself as someone who takes risks? (Yes/No)	2040
Frequency Depressed Mood 2 Weeks	Over the past two weeks, how often have you felt down, depressed or hopeless? (1=Not at all, 2=Several days, 3=More than half the days, 4=Nearly every day)	2050
Frequency Disinterest Last 2 Weeks	Over the past two weeks, how often have you had little interest or pleasure in doing things? (1=Not at all, 2=Several days, 3=More than half the days, 4=Nearly every day)	2060
Frequency Restlessness Last 2 Weeks	Over the past two weeks, how often have you felt tense, fidgety or restless? (1=Not at all, 2=Several days, 3=More than half the days, 4=Nearly every day)	2070
Frequency Tiredness Last 2 Weeks	Over the past two weeks, how often have you felt tired or had little energy? (1=Not at all, 2=Several days, 3=More than half the days, 4=Nearly every day)	2080
Seen GP Anxiety Depression	Have you ever seen a general practitioner (GP) for nerves, anxiety, tension, or depression? (Yes/No)	2090
Seen Psychiatrist Anxiety Depression	Have you ever seen a psychiatrist for nerves, anxiety, tension, or depression? (Yes/No)	2100
Illness Injury Assault Other	In the last 2 years have you experienced any of the following? (You can select more than one answer) (1=Serious illness, injury or assault to yourself, 2=Serious illness, injury or assault of a close relative , 3=Death of a close relative, 4=Death of a spouse or partner, 5=Marital separation/divorce, 6=Financial difficulties	6145
Illness Injury Assault Self	In the last 2 years have you experienced any of the following? (You can select more than one answer) (1=Serious illness, injury or assault to yourself ,	6145

	2=Serious illness, injury or assault of a close relative, 3=Death of a close relative, 4=Death of a spouse or partner, 5=Marital separation/divorce, 6=Financial difficulties	
Death Spouse	In the last 2 years have you experienced any of the following? (You can select more than one answer) (1=Serious illness, injury or assault to yourself, 2=Serious illness, injury or assault of a close relative, 3=Death of a close relative, 4=Death of a spouse or partner , 5=Marital separation/divorce, 6=Financial difficulties	6145
Death Other	In the last 2 years have you experienced any of the following? (You can select more than one answer) (1=Serious illness, injury or assault to yourself, 2=Serious illness, injury or assault of a close relative, 3=Death of a close relative , 4=Death of a spouse or partner, 5=Marital separation/divorce, 6=Financial difficulties	6145
Separation Divorce	In the last 2 years have you experienced any of the following? (You can select more than one answer) (1=Serious illness, injury or assault to yourself, 2=Serious illness, injury or assault of a close relative, 3=Death of a close relative, 4=Death of a spouse or partner, 5=Marital separation/divorce , 6=Financial difficulties)	6145
Financial Difficulties	In the last 2 years have you experienced any of the following? (You can select more than one answer) (1=Serious illness, injury or assault to yourself, 2=Serious illness, injury or assault of a close relative, 3=Death of a close relative, 4=Death of a spouse or partner, 5=Marital separation/divorce, 6=Financial difficulties)	6145
Chronic Widespread Pain	Have you had pains all over the body for more than 3 months? (Yes/No)	2956
Chronic Neck Shoulder Pain	Have you had neck or shoulder pains for more than 3 months? (Yes/No)	3404
Chronic Hip Pain	Have you had hip pains for more than 3 months? (Yes/No)	3414
Chronic Back Pain	Have you had back pains for more than 3 months? (Yes/No)	3571
Chronic Stomach Abdominal Pain	Have you had stomach or abdominal pains for more than 3 months? (Yes/No)	3741
Chronic Knee Pain	Have you had knee pains for more than 3 months? (Yes/No)	3773
Chronic Headaches	Have you had headaches for more than 3 months? (Yes/No)	3799
Chronic Facial Pain	Have you had facial pains for more than 3 months? (Yes/No)	4067
Acute Neck Shoulder Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2=Facial pain, 3=Neck or shoulder pain , 4=Back pain, 5=Stomach or abdominal pain, 6=Hip pain, 7=Knee pain, 8=Pain all over the above)	6159

Acute Hip Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2=Facial pain, 3=Neck or shoulder pain, 4=Back pain, 5=Stomach or abdominal pain, 6=Hip pain , 7=Knee pain, 8=Pain all over the above)	6159
Acute Back Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2=Facial pain, 3=Neck or shoulder pain, 4=Back pain , 5=Stomach or abdominal pain, 6=Hip pain, 7=Knee pain, 8=Pain all over the above)	6159
Acute Stomach Abdominal Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2=Facial pain, 3=Neck or shoulder pain, 4=Back pain, 5=Stomach or abdominal pain , 6=Hip pain, 7=Knee pain, 8=Pain all over the above)	6159
Acute Knee Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2=Facial pain, 3=Neck or shoulder pain, 4=Back pain, 5=Stomach or abdominal pain, 6=Hip pain, 7=Knee pain , 8=Pain all over the above)	6159
Acute Facial Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2= facial pain , 3=Neck or shoulder pain, 4=Back pain, 5=Stomach or abdominal pain, 6=Hip pain, 7=Knee pain, 8=Pain all over the above)	6159
Acute Headaches	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache , 2=Facial pain, 3=Neck or shoulder pain, 4=Back pain, 5=Stomach or abdominal pain, 6=Hip pain, 7=Knee pain, 8=Pain all over the above)	6159
Acute Widespread Pain	In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer) (1=Headache, 2=Facial pain, 3=Neck or shoulder pain, 4=Back pain, 5=Stomach or abdominal pain, 6=Hip pain, 7=Knee pain, 8=Pain all over the above)	6159

Supplementary Table 3. Features used in UK Biobank model. Question column presents question presented to participants and data coding. Each data-field refers to a fundamental block of data showing the results of a single question or a measurement in the UK Biobank repository. Feature domains: **Chronic Pain**, **Functional factors**, **Socioeconomic**, **Sleep**, **Mood**, **Substance use**, **Demographic**, **Neuroticism**, **Occupational**, **Acute pain**.

Feature	Associated Question	All of Us Questionnaire
Age	Age at the start of study participation	The Basics
Sex	What was your biological sex assigned at birth? (Female, Male, Intersex)	The Basics
What Race Ethnicity: Asian	Which categories describe you? Select all that apply. Note, you may select more than one group	The Basics
What Race Ethnicity: Black	Which categories describe you? Select all that apply. Note, you may select more than one group	The Basics
What Race Ethnicity: Hispanic	Which categories describe you? Select all that apply. Note, you may select more than one group	The Basics
What Race Ethnicity: Race Ethnicity None of These	Which categories describe you? Select all that apply. Note, you may select more than one group	The Basics
What Race Ethnicity: White	Which categories describe you? Select all that apply. Note, you may select more than one group (White, Black, Hispanic, more than one, Asian, Other)	The Basics
Another single population	Which categories describe you? Select all that apply. Note, you may select more than one group	The Basics
More than one population	Which categories describe you? Select all that apply. Note, you may select more than one group	The Basics
Employed for wages or self-employed	What is your current employment status ? Please select 1 or more of these categories. (Employed for Wages , Homemaker, Out Of Work Less Than One, Out of Work One Or More, Retired, Self Employed , Student, Unable to Work, Student)	The Basics
Not currently employed for wages	What is your current employment status ? Please select 1 or more of these categories. (Employed for Wages, Homemaker, Out of Work Less Than One , Out of Work One or More , Retired, Self Employed, Student, Unable to Work)	The Basics
Highest Grade: College graduate or advanced degree	What is the highest grade or year of school you completed? (College One to Three, College Graduate, Advance Degree , Twelve or GED, Nine Through Eleven, Five Through Eight), One Through Four, Never Attended)	The Basics
Highest Grade: College One to Three	What is the highest grade or year of school you completed? (College One to Three , College Graduate, Advance Degree, Twelve or GED, Nine Through Eleven, Five Through Eight), One Through Four, Never Attended)	The Basics
Highest Grade: Twelve Or GED	What is the highest grade or year of school you completed? (College One to Three, College Graduate, Advance Degree, Twelve or GED , Nine Through Eleven, Five Through Eight), One Through Four, Never Attended)	The Basics

Highest Grade: Less than a high school degree or equivalent	What is the highest grade or year of school you completed? (College One to Three, College Graduate, Advance Degree, Twelve or GED, Nine Through Eleven, Five Through Eight, One Through Four, Never Attended)	The Basics
Alcohol: Average Daily Drink Count	On a typical day when you drink, how many drinks do you have? (1 or 2, 3 or 4, 5 or 6, 7 to 9, 10 or more)	Lifestyle
Alcohol: Drink Frequency Past Year	How often did you have a drink containing alcohol in the past year? (Never, Monthly or less, Two to four times a month, Two to three times a week, Four or more times a week)	Lifestyle
Alcohol: 6 or More Drinks Occurrence	How often did you have six or more drinks on one occasion in the past year? (Less than month, Monthly, Weekly, Daily or almost daily, Never in the last year)	Lifestyle
Smokeless Tobacco: Smokeless Tobacco Participant	Have you ever used smokeless tobacco products, even one or two times? (Smokeless tobacco products include snus pouches, Skoal Bandits, loose snus, moist snuff, dip, spit, and chewing tobacco.) (Yes/No)	Lifestyle
Overall Health: General Quality	In general, would you say your health is: (Poor, Fair, Good, Very Good, Excellent)	Overall Health
Overall Health: Average Fatigue 7 Days	In the past 7 days, how would you rate your fatigue? (None, Mild, Moderate, Severe, Very Severe)	Overall Health
Overall Health: Emotional Problem 7 Days	In the past 7 days, how often have you been bothered by emotional problems such as feeling anxious, depressed or irritable? (Sometimes, Rarely, Never, Often, Always)	Overall Health
Overall Health: General Physical Health	In general, how would you rate your physical health? (Poor, Fair, Good, Very Good, Excellent)	Overall Health
Overall Health: Everyday Activities	To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair? (Completely, Mostly, Moderately, A Little, Not At All)	Overall Health
Body Mass Index (BMI)	Lab Measurement	Labs & Measurements
Overall Health: Social Satisfaction	In general, how would you rate your satisfaction with your social activities and relationships? (Poor, Fair, Good, Very Good, Excellent)	Overall Health
Income: Annual Income	What is your annual household income from all sources? <10k, 10k- 25k, 25k-35k, 35k-50k, 50k-75k, 75k-100k, 100k-150k, 150k-200k, >200k	The Basics
Tobacco at Home	In the past month: Did someone in your home smoke tobacco/nicotine (including cigarettes, cigar, cigarillos, pipes, hookah) every day, some days, or not at all?	

Feeling Nervous	In the past 2 weeks, how often have you been bothered by the following problem? Feeling nervous, anxious, or on edge	
Pain	In the past 7 days, how would you rate your pain on average? (0-10)	Overall Health

Supplementary Table 4. Features used in the AoU model. Feature domains: **Pain Intensity**, **Functional factors**, **Socioeconomic**, **Mood**, **Substance use**, **Demographic**, **Occupational**

1. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-9.
2. Treede R-D, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). *pain*. 2019;160(1):19-27.
3. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. 2017;16(1):19-34.
4. Wu Y, Byrne E, Zheng Z, Kemper K, Yengo L, Mallett A. Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nat Commun*. 2019; 10 (1): 1891. Epub 2019/04/25. <https://doi.org/10.1038/s41467-019-09572-5> PMID: 31015401; 2019.
5. Dinga R, Schmaal L, Penninx BW, Veltman DJ, Marquand AF. Controlling for effects of confounding variables on machine learning predictions. *BioRxiv*. 2020:2020.08. 17.255034.