**Table 1.** ICEES feature variables: name, description, and binning strategy.\*

|  |  |
| --- | --- |
| **Feature Variable** | **Description and binning strategy** |
| CohortID | *Cohort ID# & input variables used to define cohort* |
| PatientID | *Randomly assigned patient identifier* |
| StudyPeriod | *Years 2010, 2011, 2012, 2013, 2015* |
| AgeStudyStart | *Age at ‘study’ start date: calculated from birth date and binned as 0...2, 3...17, 18...34, 35...50, 51...69, 70–89 years*† |
| Sex | *Male, Female, Unknown, Other* |
| Sex2 | *Male, Female* |
| Race | *Caucasian, African American, Asian, Native Hawaiian/Pacific Islander, American/Alaskan Native, Other* |
| Ethnicity | *Hispanic (1=Yes, 0=No)* |
| ObesityDx | *ICD code†† for obesity anytime over 'study' period (1=Yes, 0=No)* |
| ObesityBMI | *BMI ≥ 30 anytime over 'study' period (1=Yes, 0=No)* |
| Diagnoses:  AsthmaDx  CroupDx  ReactiveAirwayDx  CoughDx  PneumoniaDx  UterineCancerDx  CervicalCancerDx  OvarianCancerDx  ProstateCancerDx  TesticularCancerDx  KidneyCancerDx  PregnancyDx  MenopauseDX  EndometriosisDx  OvarianDysfunctionDx  TesticularDysfunction  DiabetesDx  AlopeciaDx  FibromyalgiaDx  AlcoholDependenceDx  DrugDependenceDx  DepressionDx  AnxietyDx  AutismDx | *One or more diagnoses for select conditions (see list in left column) (defined by high-level ICD categories) over 'study' period; 1=Yes, 0=No)* |
| EstResidentialDensity | *US Census Bureau ACS 2012-2016 estimated total population [block group], binned according to US Census Bureau definitions (1=rural [0,2500), 2=urban cluster [2500,50000), 3=urbanized area [50000,inf))* |
| UrbanRural | *US Census Bureau ACS 2012-2016 estimated total population [block group], binned according to US Census Bureau definitions (R=rural [0,2500), U=urban cluster [2500,50000] or urbanized area [50000,inf))* |
| EstResidentialDensity25Plus | *US Census Bureau ACS 2012-2016 estimated total population aged 25 years or older [block group], binned as quintiles (1, 2, 3, 4, 5)* |
| EstProbabilityNonHispWhite | *US Census Bureau ACS 2012-2016 estimated proportion of persons who are non-Hispanic white [block group], binned as quartiles (1, 2, 3, 4)* |
| EstProbabilityHouseholdNonHispWhite | *US Census Bureau ACS 2012-2016 estimated proportion of households that are non-Hispanic white [block group], binned as quartiles (1, 2, 3, 4)* |
| EstProbabilityHighSchoolMaxEducation | *US Census Bureau ACS 2012-2016 estimated proportion persons aged 25 y or older with a HS diploma or less at their highest level of schooling [block group], binned as quartiles (1, 2, 3, 4)* |
| EstProbabilityNoAuto | *US Census Bureau ACS 2012-2016 estimated proportion of households without an automobile [block group], binned as quartiles (1, 2, 3, 4)* |
| EstProbabilityNoHealthIns | *US Census Bureau ACS 2012-2016 estimated proportion of persons without health insurance [block group], binned as quartiles (1, 2, 3, 4)* |
| EstProbabilityESL | *US Census Bureau ACS 2012-2016 estimated proportion of persons 5 y or older who sometimes speak a language other than English at home [block group], binned as quartiles (1, 2, 3, 4)* |
| EstHouseholdIncome | *US Census Bureau ACS 2012-2016 estimated median household income [block group], binned as quintiles (1, 2, 3, 4, 5)* |
| MajorRoadwayHighwayExposure | *US Census TIGERline distance in meters from household to nearest major road/highway (1 = 0-49, 2 = 50-99, 3 = 100-199, 4 = 200-299, 5 = 300-499, 6 = >=500 meters)* |
| RoadwayDistanceExposure | *US DOT distance in meters from household to nearest roadway (1 = 0-49, 2 = 50-99, 3 = 100-199, 4 = 200-299, 5 = 300-499, 6 = >=500 meters)* |
| RoadwayType | *UNC DOT roadway classification (e.g., major highway)* |
| RoadwayAADT | *US DOT Annual average daily traffic estimate* |
| RoadwaySpeedLimit | *US DOT Roadway speed limit* |
| RoadwayLanes | *UNC DOT Roadway number of lanes* |
| UNC IE CMAQ exposure estimates: PM2.5, ozone§  Variable nomenclature format:  AvgDaily**XX**Exposure\_StudyAvg  MaxDaily**XX**Exposure\_StudyAvg  AvgDaily**XX**Exposure\_StudyMax  MaxDaily**XX**Exposure\_StudyMax | *UNC IE average of estimated average daily PM2.5 exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)*  *UNC IE average of estimated maximum daily exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)*  *UNC IE maximum of estimated average daily PM2.5 exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)*  *UNC IE maximum of estimated maximum daily exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)* |
| UNC IE CMAQ exposure estimates: PM2.5, ozone§  Variable nomenclature format:  AvgDaily**XX**Exposure\_StudyAvg\_qcut  MaxDaily**XX**Exposure\_StudyAvg\_qcut  AvgDaily**XX**Exposure\_StudyMax\_qcut  MaxDaily**XX**Exposure\_StudyMax\_qcut | *UNC IE average of estimated average daily PM2.5 exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)*  *UNC IE average of estimated maximum daily exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)*  *UNC IE maximum of estimated average daily PM2.5 exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)*  *UNC IE maximum of estimated maximum daily exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)* |
| US EPA conUS CMAQ exposure estimates: PM2.5, ozone, CO, NO, NO2, NOx, SO2, acetaldehyde, formaldehyde, benzene\*\*  Variable nomenclature format:  AvgDaily**XX**Exposure\_2  MaxDaily**XX**Exposure\_2 | *US EPA conUS CMAQ daily exposure estimates for PM2.5, CO, NO, NO2, NOx, SO2, acetaldehyde, formaldehyde, or benzene, averaged over ‘study’ period, binned by data values (pandas cut) ((1,2,3,4,5)*  *US EPA conUS CMAQ 8-hour PM2.5 exposure estimate, averaged over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)* |
| US EPA conUS CMAQ exposure estimates: PM2.5, ozone, CO, NO, NO2, NOx, SO2, acetaldehyde, formaldehyde, benzene\*\*  Variable nomenclature format:  AvgDaily**XX**Exposure\_2\_qcut  MaxDaily**XX**Exposure\_2\_qcut | *US EPA conUS CMAQ daily exposure estimates for PM2.5, CO, NO, NO2, NOx, SO2, acetaldehyde, formaldehyde, or benzene, averaged over ‘study’ period, binned by quintiles (pandas.qcut) (1,2,3,4,5)*  *US EPA conUS CMAQ 8-hour PM2.5 exposure estimate, averaged over 'study' period, binned by quintiles (pandas.qcut) (1,2,3,4,5)* |
| Medications Prescribed or  Administered:  Prednisone  Fluticasone  Mometasone  Budesonide  Beclomethasone  Ciclesonide  Flunisolide  Albuterol  Metaproterenol  Diphenhydramine  Fexofenadine  Cetirizine  Ipratropium  Salmeterol  Arformoterol  Formoterol  Indacaterol  Theophylline  Omalizumab  Mepolizumab | *One or more prescriptions/administrations of select medications (see list in left column) over 'study' period (1=Yes, 0=No)* |
| TotalEDInpatientVisits | *Total # ED/inpatient visits for respiratory issue(s) (defined by same ICD codes used to pull patients with asthma-like conditions) over ‘study’ period (0, 1, 2, 3, …)* |

*Abbreviations: ACS = American Community Survey; DOT = Department of Transportation; ED = Emergency Department; IE = Institute for the Environment; PM25 = particulate matter ≤ 2.5 µm in diameter*

\*The feature variables listed in the table are those for the patient-level tables, which include data on each patient for each year of available data (i.e., data on individual patients are represented as rows in the table). Similar feature variables are available for the visit-level tables, although the variables are sometimes treated differently. For example, PM2.5 exposures are expressed in relation to the 24-hour and two-week period before visits, not in relation to the one-year ‘study’ period, as was done for the patient-level tables. Additional feature variables (e.g., laboratory measures) are available for select years. Further information can be accessed via the [ICEES OpenAPI](https://icees.renci.org/apidocs).

†Maximum age is 89 years, per HIPAA regulations

††The ICD codes for obesity and for the selection of patients with asthma-like conditions can be found in footnote two of the main manuscript.

§From first batch of CMAQ output, hourly estimates, 36-km (2010) or 12-km (2011) resolutions

\*\*From second batch of CMAQ output, US Census tract resolution, 2002-2014 for PM2.5 and ozone, 2002 only for other chemicals