

Full set of available features with enumerations can be found at:

<https://github.com/NCATS-Tangerine/icees-api/blob/master/config/features.yml>

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

Feature Variable	Description and binning strategy
CohortID	<i>Cohort ID# & input variables used to define cohort</i>
PatientID	<i>Randomly assigned patient identifier</i>
StudyPeriod	<i>Years 2010, 2011, 2012, 2013, 2014, 2015, 2016</i>
CDWH Clinical Data (years 2010-2019)	
AgeStudyStart	<i>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[†]</i>
AgeStudyStart2	<i>Age at 'study' start date: calculated from birth date and binned by US Census Bureau groups as <5, 5-17, 18-44, 45-64, 65-89 years</i>
Sex	<i>Male, Female, Unknown, Other</i>
Sex2	<i>Male, Female</i>
Race	<i>Caucasian, African American, Asian, Native Hawaiian/Pacific Islander, American/Alaskan Native, Other</i>
Ethnicity	<i>Hispanic (1=Yes, 0=No)</i>
ObesityDx	<i>ICD code^{††} for obesity anytime over 'study' period (1=Yes, 0=No)</i>
ObesityBMI	<i>BMI \geq 30 anytime over 'study' period (1=Yes, 0=No)</i>
Diagnoses: Asthma and asthma-like diagnoses: AsthmaDx CroupDx ReactiveAirwayDx CoughDx PneumoniaDx	<i>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)</i>

<p>Diagnoses with known or suspected sex differences:</p> <p>UterineCancerDx CervicalCancerDx OvarianCancerDx ProstateCancerDx TesticularCancerDx KidneyCancerDx PregnancyDx MenopauseDx EndometriosisDx OvarianDysfunctionDx TesticularDysfunctionDx DiabetesDx AlopeciaDx FibromyalgiaDx AlcoholDependenceDx DrugDependenceDx DepressionDx AnxietyDx AutismDx</p>	
<p>Medications Prescribed or Administered:</p> <p>Medications for asthma-like conditions:</p> <p>Prednisone Fluticasone Mometasone Budesonide Beclomethasone Ciclesonide Flunisolide Albuterol Metaproterenol Diphenhydramine Fexofenadine Cetirizine Ipratropium Salmeterol Arformoterol Formoterol Indacaterol Theophylline Omalizumab Mepolizumab</p> <p>Medications with known or suspected sex differences:</p>	<p><i>One or more prescriptions/administrations of select medications (see list in left column) over 'study' period (1=Yes, 0=No)</i></p>

<p>Anti-depressants</p> <ul style="list-style-type: none"> • Sertraline hydrochloride (Zoloft)* • Citalopram hydrobromide (Celexa) • Fluoxetine hydrochloride (Prozac) • Paroxetine (Paxil) • Trazodone hydrochloride (Desyrel) • Escitalopram oxalate (Lexapro) • Duloxetine hydrochloride (Cymbalta) • Venlafaxine (Effexor) <p>Anti-anxiety drugs (in addition to the above)</p> <ul style="list-style-type: none"> • Propranolol • Hydroxyzine <p>Estrogens</p> <ul style="list-style-type: none"> • Estradiol* • Estropipate* • Estrogen* <p>Progestins</p> <ul style="list-style-type: none"> • Progesterone* • Medroxyprogesterone* <p>Androgens</p> <ul style="list-style-type: none"> • Testosterone* • Androstenedione* • Nandrolone* • Prasterone* <p>Prostate cancer</p> <ul style="list-style-type: none"> • Leuprolide • Goserelin • Triptorelin • Histrelin <p>Breast cancer</p> <ul style="list-style-type: none"> • Tamoxifen <p><i>*Key meds to use as positive controls for sex differences</i></p>	
<p>Laboratory Measures</p> <p>See 'ICEES+ UNC COVID19 FeatureVariables'</p>	<p><i>One or more laboratory measures over “study” period: Lab X (1=Yes, 0=No), X_flag_first (first lab in study period), X_flag_last (last lab in study period)</i></p>
<p>TotalEDInpatientVisits</p>	<p><i>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, ...)</i></p>
<p>US Census Bureau ACS Data (years 2007-2011 & 2012-2016 survey samples)</p>	
<p>EstResidentialDensity</p>	<p><i>US Census Bureau ACS 2012-2016 estimated total population [block group], binned according to US Census Bureau definitions (1=rural [0,2500),</i></p>

	<i>2=urban cluster [2500,50000), 3=urbanized area [50000,inf))</i>
ur	<i>US Census Bureau ACS 2012-2016 estimated urban or rural residence, binned according to US Census Bureau definitions</i>
EstResidentialDensity25Plus *2012-2016 sample only	<i>US Census Bureau ACS 2012-2016 estimated total population aged 25 years or older [block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropNonHispWhite	<i>US Census Bureau ACS 2012-2016 estimated proportion of persons who are non-Hispanic white [block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstPropHouseholdNonHispWhite	<i>US Census Bureau ACS 2012-2016 estimated proportion of households that are non-Hispanic white [block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstPropHighSchoolMaxEducation	<i>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 (pandas qcut) (1, 2, 3, 4)</i>
EstPropHighSchoolDropout	<i>US Census Bureau ACS 2012-2016 estimated proportion persons aged 16-19 y who are neither attending school nor HS graduates [block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstPropHighSchoolDropoutNoWork	<i>US Census Bureau ACS 2012-2016 estimated proportion persons aged 16-19 y who are neither attending school nor HS graduates and are without work [block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstPropHouseholdsNoAuto	<i>US Census Bureau ACS 2012-2016 estimated proportion of households without an automobile [block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstPropHouseholdsNoHealthIns *2012-2016 sample only	<i>US Census Bureau ACS 2012-2016 estimated proportion of persons without health insurance</i>

	<i>[block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstProp5PlusESL	<i>US Census Bureau ACS 2012-2016 estimated proportion of persons 5 years or older who sometimes speak a language other than English at home [block group], binned as quartiles (pandas qcut) (1, 2, 3, 4)</i>
EstMedianHouseholdIncome	<i>US Census Bureau ACS 2012-2016 estimated median household income [block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropMaleLittleWork	<i>US Census Bureau ACS 2012-2016 estimated proportion of males aged 16-64 years who worked less than 26 weeks in previous year [block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropHouseholdSSI	<i>US Census Bureau ACS 2012-2016 estimated proportion of households receiving Supplemental Security Income block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropHouseholdPA	<i>US Census Bureau ACS 2012-2016 estimated proportion of households receiving Public Assistance [block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropFemaleHouseholdNoSpouse	<i>US Census Bureau ACS 2012-2016 estimated proportion of family households headed by a female (no male partner present) [block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropFemaleHouseholdFamilyChild	<i>US Census Bureau ACS 2012-2016 estimated proportion of total households headed by a female with family children aged 18 y or less (no male partner present) [block group], binned as quintiles (pandas qcut) (1, 2, 3, 4, 5)</i>
EstPropFemaleHouseholdAnyChild	<i>US Census Bureau ACS 2012-2016 estimated proportion of total households headed by a female with any children aged 18 y or less (no male partner present) [block group], binned as quintiles (1, 2, 3, 4, 5)</i>
US Census TIGERline Roadway Data (year 2016)	

MajorRoadwayHighwayExposure	<i>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)</i>
MajorRoadwayHighwayExposure2	<i>US Census TIGERline distance in meters from household to nearest major road/highway (1 = 0-49, 2 = 50-99, 3 = 100-149, 4 = 150-199, 5 = 200-249, 6 = >=250 meters)</i>
US DOT Roadway Data (year 2016)	
RoadwayDistanceExposure	<i>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)</i>
RoadwayDistanceExposure2	<i>US DOT distance in meters from household to nearest roadway (1 = 0-49, 2 = 50-99, 3 = 100-149, 4 = 150-199, 5 = 200-249, 6 = >=250 meters)</i>
RoadwayType	<i>UNC DOT roadway classification (e.g., major highway)</i>
RoadwayAADT	<i>US DOT Annual average daily traffic estimate</i>
RoadwaySpeedLimit	<i>US DOT Roadway speed limit</i>
RoadwayLanes	<i>UNC DOT Roadway number of lanes</i>
NC Department of Environmental Quality Data	
CAFO_Distance	<i>NC DEQ distance in meters from household to nearest concentrated animal farming operation (1 = <500, 2 = 500-1000, 3 = 1000-2000, 4 = 2000-4000, 5 = >4000)</i>
LandfillDistance	<i>NC DEQ distance in meters from household to nearest landfill (1 = <500, 2 = 500-1000, 3 = 1000-2000, 4 = 2000-4000, 5 = >4000)</i>
NC Public School Data	
PublicSchoolDistance	<i>TBD</i>

<p>UNC IE 2010 & 2011 CMAQ exposure estimates: PM2.5, ozone^s</p>	
<p>Variable nomenclature format:</p>	
<p>AvgDailyXXExposure_StudyAvg</p>	<p><i>UNC IE average of estimated average daily PM2.5 exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)</i></p>
<p>MaxDailyXXExposure_StudyAvg</p>	<p><i>UNC IE average of estimated maximum daily exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)</i></p>
<p>AvgDailyXXExposure_StudyMax</p>	<p><i>UNC IE maximum of estimated average daily PM2.5 exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)</i></p>
<p>MaxDailyXXExposure_StudyMax</p>	<p><i>UNC IE maximum of estimated maximum daily exposure over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)</i></p>
<p>AvgDailyXXExposure_StudyAvg_qcut</p>	<p><i>UNC IE average of estimated average daily PM2.5 exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)</i></p>
<p>MaxDailyXXExposure_StudyAvg_qcut</p>	<p><i>UNC IE average of estimated maximum daily exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)</i></p>
<p>AvgDailyXXExposure_StudyMax_qcut</p>	<p><i>UNC IE maximum of estimated average daily PM2.5 exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)</i></p>
<p>MaxDailyXXExposure_StudyMax_qcut</p>	<p><i>UNC IE maximum of estimated maximum daily exposure over 'study' period, binned by quintiles (pandas qcut) (1, 2, 3, 4, 5)</i></p>

<p>US EPA 2002-2016 conUS CMAQ exposure estimates: PM2.5, ozone; 2002-2014 conUS CMAQ exposure estimates: CO, NO, NO2, NOx, SO2, acetaldehyde, formaldehyde, benzene**</p> <p><i>(Benzene is only available for select years)</i></p> <p>Variable nomenclature format: AvgDailyXXExposure_2 MaxDailyXXExposure_2</p> <p>AvgDailyXXExposure_2_qcut MaxDailyXXExposure_2_qcut</p>	<p><i>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)</i></p> <p><i>US EPA conUS CMAQ 8-hour ozone maximum exposure estimate, averaged over 'study' period, binned by data values (pandas cut) (1, 2, 3, 4, 5)</i></p> <p><i>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)</i></p> <p><i>US EPA conUS CMAQ 8-hour PM2.5 exposure estimate, averaged over 'study' period, binned by quintiles (pandas.qcut) (1,2,3,4,5)</i></p>
<p>NIEHS EPR Survey and SNP DATA: see “NIEHS_EPR_FeatureVariables.xlsx” file</p>	
<p><i>Variables include EPR survey data, SNP data, and data from Green Team Exposures APIs (generated by independent batch pull, using geocodes associated with EPR)</i></p> <p><i>Differentiated from UNC survey by prefix “EPR”</i></p>	

Abbreviations: ACS = American Community Survey; DOT = Department of Transportation; ED = Emergency Department; IE = Institute for the Environment; PM2.5 = particulate matter $\leq 2.5 \mu\text{m}$ 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, PM_{2.5} 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](#).

† 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, 2002-2015 for PM2.5 and ozone, 2002 only for other chemicals

Note: ICEES is currently restricted to patients with an asthma-like condition from UNC Health Care System and the NIEHS Environmental Polymorphism Registry.

The following codes and parameters were used to identify patients with an 'asthma-like' phenotype:

Diagnostic codes for asthma and asthma-like conditions ICD9 493.% asthma ICD10 J45.% asthma ICD9 464.% croup ICD10 J05.% croup ICD9 496.% reactive airway ICD10 J44.% reactive airway ICD10 J66.% reactive airway ICD9 786.% cough ICD10 R05.% cough ICD9 481.% pneumonia ICD9 482.% pneumonia ICD9 483.% pneumonia ICD9 484.% pneumonia ICD9 485.% pneumonia ICD9 486.% pneumonia ICD10 J12.% pneumonia ICD10 J13.% pneumonia ICD10 J14.% pneumonia ICD10 J15.% pneumonia ICD10 J16.% pneumonia ICD10 J17.% pneumonia ICD10 J18.% pneumonia

Tests and procedures for asthma and asthma-like conditions CPT 94010 spirometry CPT 94070 multiple spirometry CPT 95070 methacholine challenge test CPT 94620 simple exercise stress test CPT

94621 complex exercise stress test CPT 31624 bronchoscopy CPT 94375 flow-volume loop CPT 94060 spirometry (pre/post bronchodilator test) CPT 94070 bronchospasm provocation CPT 95070 inhalation bronchial challenge CPT 94664 bronchodilator administration CPT 94620 pulmonary stress test CPT 95027 airborne allergen panel

Medications prescribed for patients with asthma-like phenotype MEDCTN prednisone MEDCTN fluticasone MEDCTN mometasone MEDCTN budesonide MEDCTN beclomethasone MEDCTN ciclesonide MEDCTN flunisolide MEDCTN albuterol MEDCTN metaproterenol MEDCTN diphenhydramine MEDCTN fexofenadine MEDCTN cetirizine MEDCTN ipratropium MEDCTN salmeterol MEDCTN arformoterol MEDCTN formoterol MEDCTN indacaterol MEDCTN theophylline MEDCTN omalizumab MEDCTN mepolizumab