

ICEES – Cohorts & Clinical Feature Variables

ICEES Asthma and Related Common Pulmonary Disorders

- UNC Health electronic health record data, ~160,000 patients (selected in 2016 for 2010)
- NIEHS Personalized Environment & Genes Study (PEGS, formerly the Environmental Polymorphisms Registry) survey and SNP data, ~4,000 participants total (~1,000 with self-reported asthma)
 - ~220 overlapping UNC Health/PEGS patients-participants
 - For more information on PEGS, see:
<https://www.niehs.nih.gov/research/clinical/studies/pegs/index.cfm>
- Available for years 2010-2019
- For feature variable enumeration, see:
<https://github.com/ExposuresProvider/icees-api/tree/master/doc> and
<https://github.com/ExposuresProvider/icees-api-config/blob/master/config/ground-truth-yaml-files>
 - Note that in most cases, clinical feature variables are binned as 0 (no) or 1 (yes)
- For additional information, see:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6748811/>

ICEES Primary Ciliary Dyskinesia (PCD) and Related Rare Pulmonary Disorders

- UNC Health EHR data, ~8,000 patients (selected in 2020)
 - ~200 patients with expert-confirmed PCD, ~200 with expert-confirmed idiopathic bronchiectasis, ~200 with expert-confirmed cystic fibrosis
- Available for years 2010-2021
- For feature variable enumeration, see:
<https://github.com/ExposuresProvider/icees-api/tree/master/doc> and
<https://github.com/ExposuresProvider/icees-api-config/blob/master/config/ground-truth-yaml-files>
 - Note that in most cases, clinical feature variables are binned as 0 (no) or 1 (yes)
- For additional information, see: <https://pubmed.ncbi.nlm.nih.gov/35837616/>

ICEES Coronavirus (COVID) Infection

- UNC Health EHR data
- Three quarterly data pulls
 - SEP2020, ~100,000 patients

- JAN2021, ~214,000 patients
- APR2021, ~100,000 patients – currently available via API
- Patients selected using N3C phenotype; definition varied as pandemic evolved
 - For more information on N3C, see: <https://ncats.nih.gov/n3c>
 - For information on the N3C phenotype, see: https://github.com/National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition
- NIEHS Personalized Environment & Genes Study (PEGS, formerly the Environmental Polymorphisms Registry) survey and SNP data, ~20,000 participants
 - ~500 to ~1,000 overlapping UNC Health/PEGS patients-participants
 - For more information on PEGS, see: <https://www.niehs.nih.gov/research/clinical/studies/pegs/index.cfm>
- One year of prior EHR data were captured for each of the three cohorts
- For feature variable enumeration, see: <https://github.com/ExposuresProvider/icees-api/tree/master/doc> and <https://github.com/ExposuresProvider/icees-api-config/blob/master/config/ground-truth-yaml-files>
 - Note that in most cases, clinical feature variables are binned as 0 (no) or 1 (yes)

ICEES Drug-Induced Liver Injury (DILI)

- DILI Network data on a cohort of patients with Augmentin-induced DILI, ~300 patients
- Patient enrollment and follow up occurred over years 1994-2019
 - For more information on the DILI Network, see: <https://diln.org/>
- Dataset created in 2021
- For feature variable enumeration, see: <https://github.com/ExposuresProvider/icees-api/tree/master/doc> and <https://github.com/ExposuresProvider/icees-api-config/blob/master/config/ground-truth-yaml-files>
 - Note that in most cases, clinical feature variables are binned as 0 (no) or 1 (yes)