**ICEES**

**ICEES** offers access to observational clinical data on all patients in the Carolina Data Warehouse for Health (CDWH) with an asthma-like phenotype (defined below). The data additionally contain data derived from several public databases on chemical exposures (e.g., airborne pollutants) and sociological exposures (e.g., estimated household income) (**Appendix A**). The exposures data have been integrated with the clinical data at the patient and visit level. The ICEES clinical data were derived from fully identified patient dataset, but the data have been 'binned' or recoded in order to protect patient privacy, while also providing open access to the data via a Translator API and ensuring compliance with §164.514(b) of [HIPAA, 'Safe Harbor' method for patient de-identification of medical records](https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification).

ICEES is designed to offer four basic functionalities (also see [slide deck](https://drive.google.com/open?id=12TgOZMFkWQLMhjZeN4RVzdxvlt1VYcO8)).

*1. Cohort discovery: users define a cohort using any number of defined feature variables as input parameters, and the service returns a sample size.*

*2. Feature-rich cohort discovery: users select a predefined cohort as the input parameter, and the service returns a profile of that cohort in terms of the available feature variables.*

*3. Hypothesis-driven 2 x 2 feature associations: users select a predefined cohort and two feature variables, and the service returns a 2 x 2 feature table with a corresponding Chi Square statistic and P value.*

*4. Exploratory 1 X N feature associations: users select a predefined cohort and a feature variable of interest, and the service returns a 1 x N feature table with corrected Chi Square statistics and associated P values.*

ICEES can be used for scientific inference and discovery, although important caveats must be considered. The main considerations when working with ICEES are outlined below.

*1. All feature variables have been binned or recoded (see*[*templates*](https://drive.google.com/open?id=12TgOZMFkWQLMhjZeN4RVzdxvlt1VYcO8)*).*

*2. The integrated feature tables are designed for different 'study' periods (currently defined as calendar years).*

*3. The integrated feature tables are designed to provide access to either patient-level data or visit-level data.*

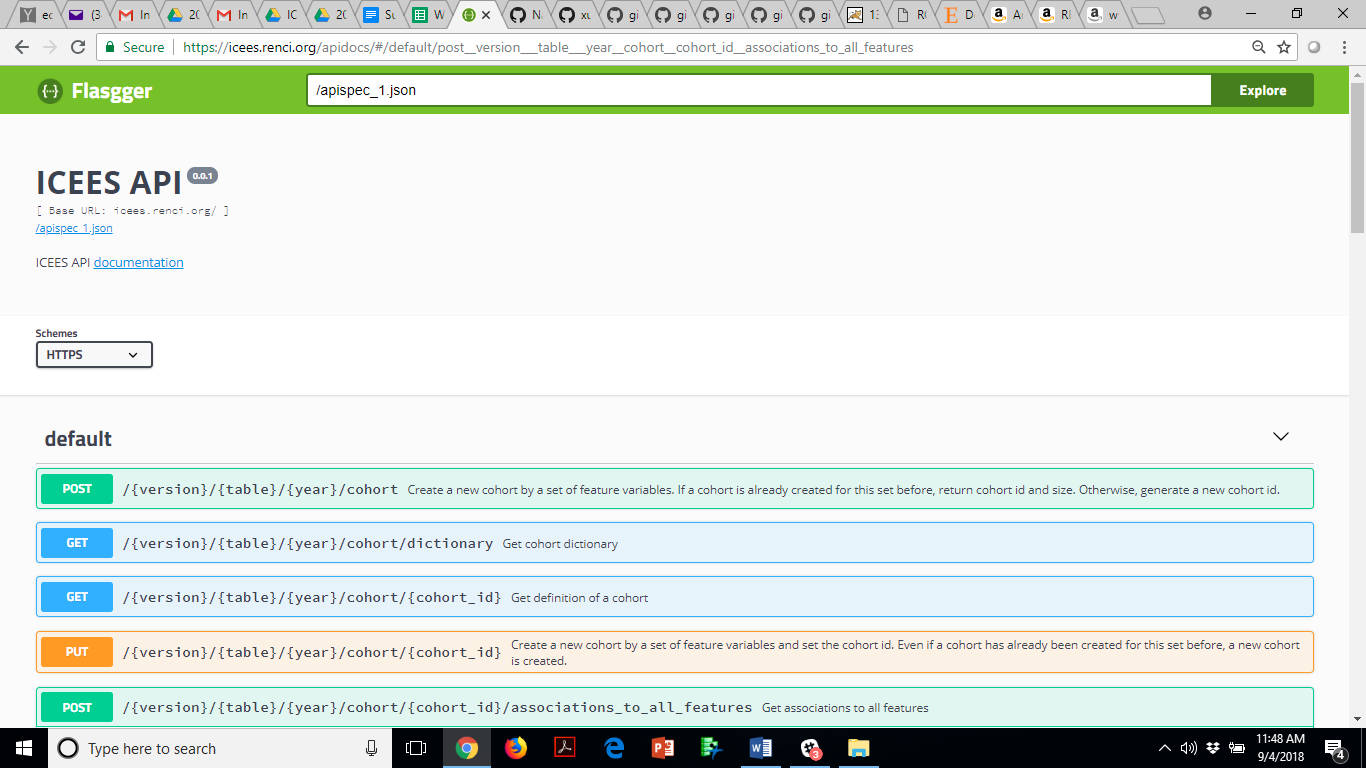
*4. All inferences must be made with respect to the binning strategy, 'study' design, and type of integrated feature table.*

*Access to ICEES is open to all Translator team members and is not subject to regulatory constraints.*

[**ICEES API**](https://icees.renci.org/apidocs/)

**ICEES Preliminary Results**

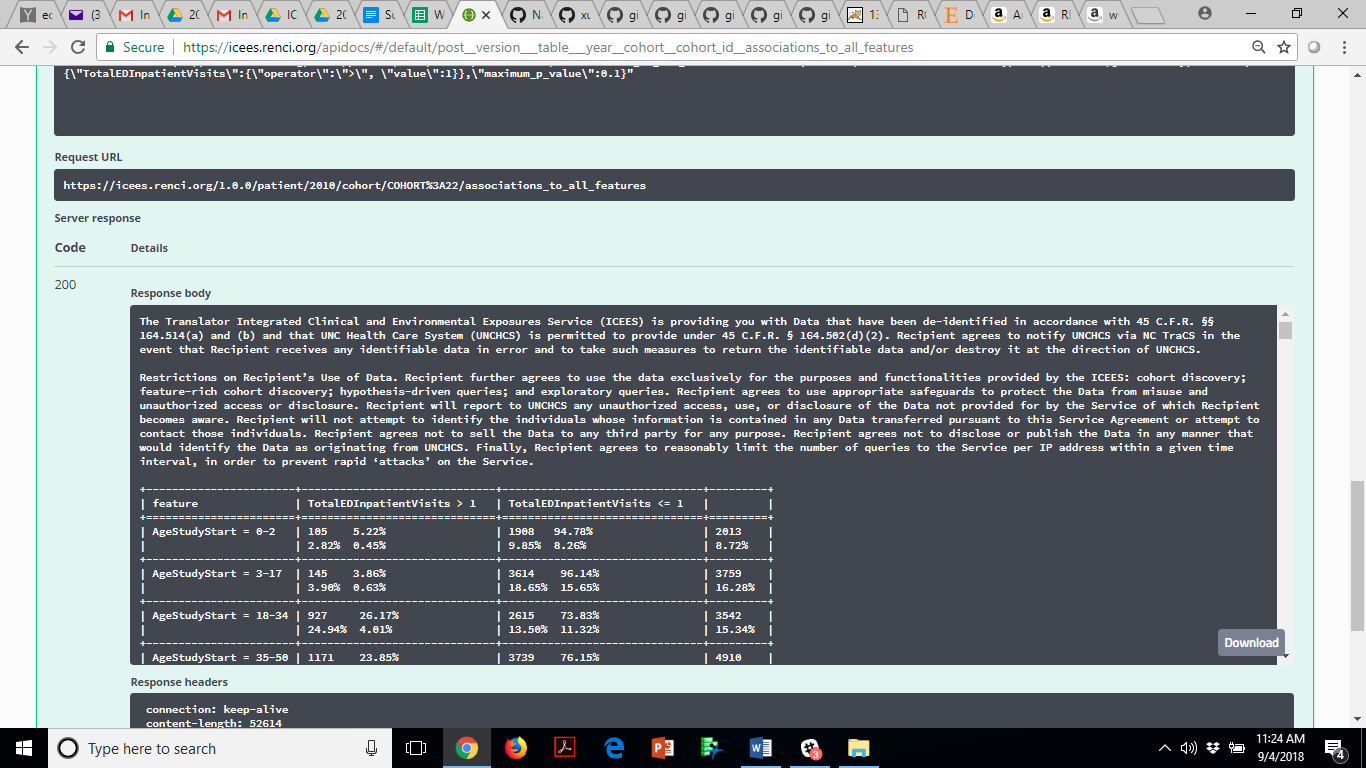
The UI for the ICEES API remains under development (**Figure 1**). While it is functional, it is not very user-friendly. Scientific and technical team members have been working together to improve the design.



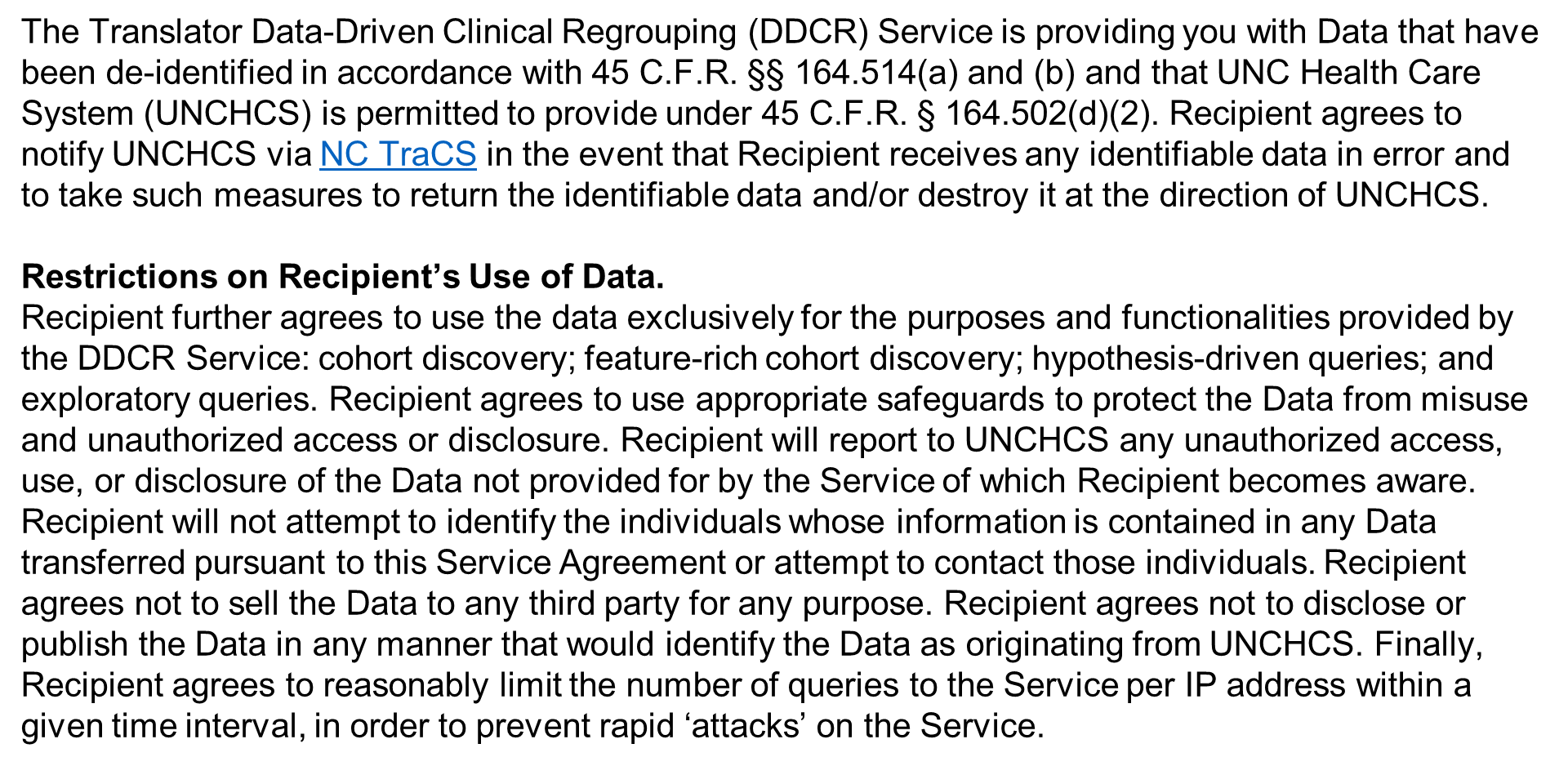
**Figure 1.** The ICEES UI.

ICEES was developed with 10 embedded safeguards in order to ensure patient privacy and abide by HIPAA Safe Harbor. Those safeguards include the fact that usage agreements are returned to users with all queries of the service (**Figure 2**).

**A.**



**B.**

****

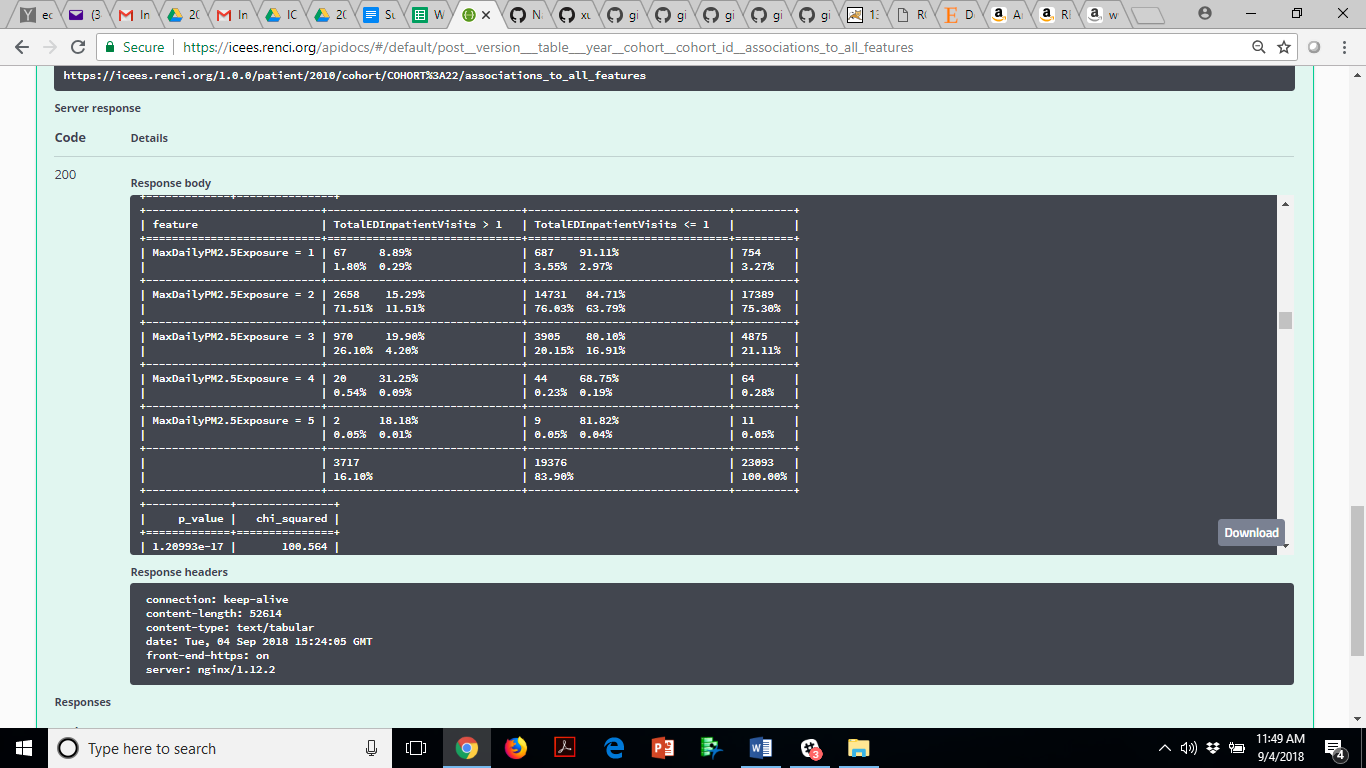
**Figure 2. A.** Screenshot of output from the ICEES API. **B.** A more user-friendly version. Terms and conditions of use. As a safeguard, a set of usage agreements is returned to users as the initial response to all queries of the service.

While we are still analyzing the integrated feature tables that are behind the ICEES API and refining the binning strategy for feature variables, the preliminary results look promising.

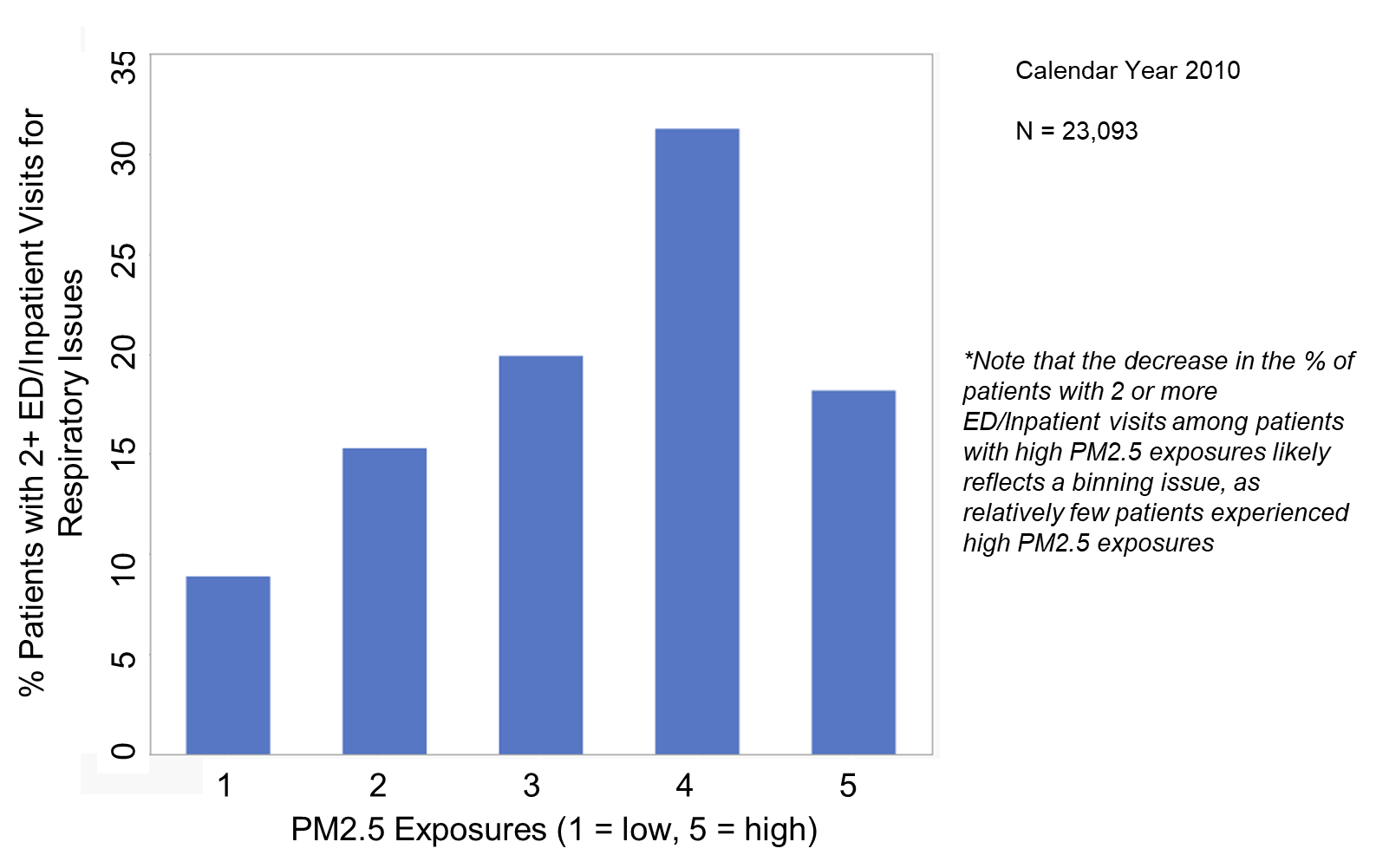
For instance, Green Team has made progress on our original CQ1, which reads: Among pediatric patients with an 'asthma-like phenotype', is exposure to particulate matter <=2.5 micrometers in diameter (PM2.5) and ozone associated with responsiveness to treatment? In other words, are exposures higher in patients who are non-responsive to treatment than in patients who are responsive to treatment?

CQ1 was developed to validate the prototype Translator system, our data sources, and our overall approach; i.e., a wealth of literature demonstrates an association between PM2.5 and ozone exposure and responsiveness to treatment (or asthma exacerbations), so we should be able to replicate this finding. Indeed, the preliminary results are clearly in the intended direction (**Figure 3**).

**A.**



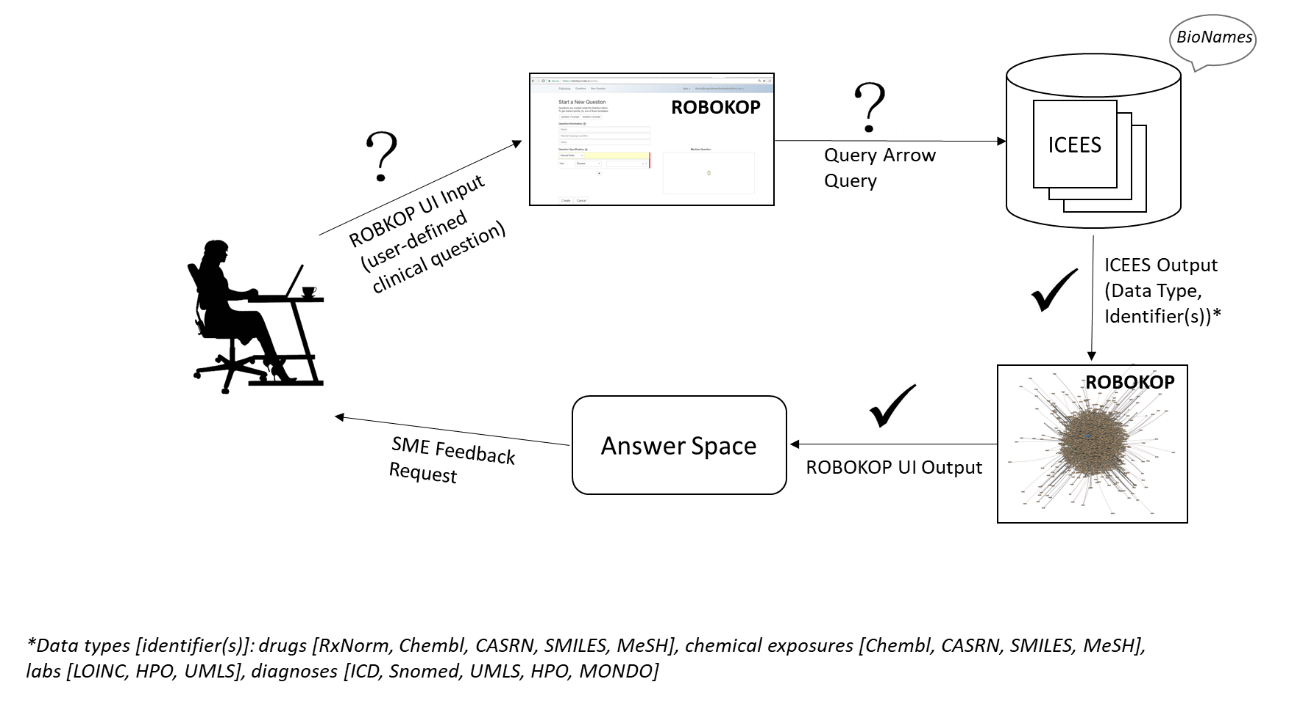
**B.**

****

**Figure 3. A.** Screenshot of output from the ICEES API. **B.** A more user-friendly version of the output. The results demonstrate a significant association between patient-level exposure to PM2.5 and number of ED or inpatient visits for respiratory issues over a one-year ‘study’ period.

We also have developed plans for answering Green Team’s CQ4 and CQ5 using Gamma’s ROBOKOP Reasoner. CQ4 and CQ5 are framed in terms of drug discovery and repurposing. CQ4 reads: Which medications are currently prescribed to pediatric patients with an asthma-like phenotype who are responsive to treatment despite high levels of exposure to particulate matter <=2.5 micrometers in diameter (PM2.5) and ozone? Conversely, which medications are currently prescribed to pediatric patients with an asthma-like phenotype who are *not* responsive to treatment and have high levels of exposure to particulate matter <=2.5 micrometers in diameter (PM2.5) and ozone? CQ5 asks: What protein (gene) targets and biological pathways do those medications act on?

Our current plan (**Figure 4**) is designed as an initial approach for incorporating clinical data into a knowledge graph in order to allow for linkages with other knowledge sources (e.g., data on gene targets, biological pathways, chemical similarities) and high-order reasoning. The fact that we have developed a feasible plan for tackling this challenge deserves emphasis: *this was a complex challenge*. As with ICEES, the solution required a lot of brainstorming and creative thinking among team members. We are confident, however, that our approach will yield meaningful results. Moreover, we believe the approach will have widespread appeal among Translator teams.



**Figure 4.** High-level overview of approach for incorporating clinical data into a knowledge graph.

**Green Team's Asthma-like Cohort**

**Asthma-like cohort**: At present, ICEES is restricted to patients with an ‘asthma-like’ phenotype. However, we are expanding ICEES to include additional patient cohorts (e.g., pain, obesity, diabetes, drug-induced liver injury).

Patients with an asthma-like phenotype were defined as follows:[[1]](#footnote-1)

*1. Patients with a diagnostic code of ‘asthma’ and prescribed or administered medications that are typically used to treat asthma;*

*2. Patients with a diagnostic code for a respiratory condition other than asthma and prescribed or administered medications that are typically used to treat asthma;*

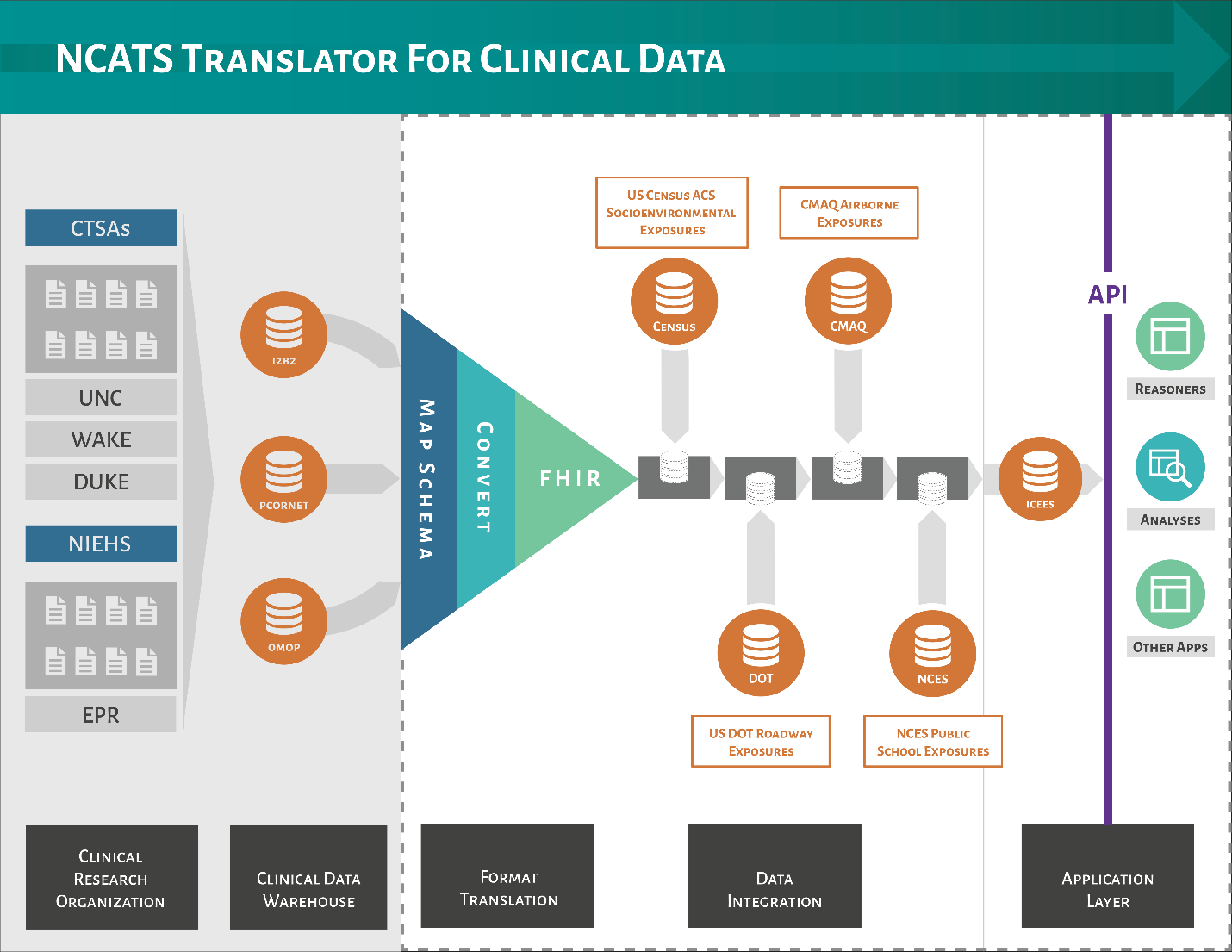
*3. Patients with a diagnostic code for a respiratory condition other than asthma and prescribed tests or procedures that are typically used to diagnosis asthma; or*

*4. Patients with a diagnostic code for a respiratory condition other than asthma and frequent ED visits with albuterol nebulizer administered. (This was not an explicit criterion, but rather is captured in the second criterion.)*

**A Note about Acknowledgements**

**We kindly request that Translator team members provide proper attribution for any products (e.g., manuscripts, podium presentations, software) derived from work related to Green Team's clinical datasets. Attribution should include acknowledgement of the funder (National Center for Advancing Translational Sciences [NCATS], Biomedical Data Translator Program awards, OT3TR002020 and OT2TR002514), the North Carolina Translational and Clinical Sciences (NC TraCS) Institute (NCATS, Center for Translational Science Award, UL1TR002489), UNC Hospitals and Health Care System, and all Green Team members who contributed to the work.**

**Appendix A.** A high-level overview of the ICEES data pipeline and its relationship to i2b2, PCORNET, and OMOP (i.e., the three main clinical data models adopted by CTSAs).



1. 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 diphenydramine MEDCTN fexofenadine MEDCTN cetirizine MEDCTN ipratropium MEDCTN salmeterol MEDCTN arformoterol MEDCTN formoterol MEDCTN indacaterol MEDCTN theophylline MEDCTN omalizumab MEDCTN mepolizumab

   [↑](#footnote-ref-1)