Mounting evidence suggests that more than half of complex disease risk is attributed to differences in an individual’s environment. There is growing consensus on the need to incorporate more social and behavioral characteristics of patients into research and clinical care in order to accurately portray all factors that may impact a person’s health. Many social determinants of health variables are available in public datasets but these datasets can be large and difficult to navigate for those inexperienced with them. This complexity became apparent as Regenstrief Institute, Inc. began developing the Indiana Addictions Data Commons (IADC), a growing resource which seeks to integrate clinical and social data sources to enable comprehensive addictions research. Regenstrief has partnered with several organizations addressing Indiana’s opioid epidemic to create both a technical infrastructure and a governance process to facilitate gathering, managing, sharing, and using population health data for both care delivery and health research in Indiana. The technical infrastructure leverages a common data model that reconciles individual and population-level data. Recently, the IADC has seen a growing demand for datasets which integrate electronic health records and area level measurements from the American Community Survey (ACS).

Problem
ACS is a publicly available dataset that offers researchers an opportunity to incorporate community level variables into research. This annual survey contains a multitude of questions related to social, economic, housing, and demographic variables. Within each category a researcher will find multiple variables represented in a variety of ways. Additionally, the ACS offers aggregated data for the most recent year as well as estimates for the past five years aggregated by demographic variables such as gender, race, or head of household. This results in a large number of permutations for each variable. For example, one variable – Occupants per Room – has 450 permutations to select from. All told, the 2017 ACS 5-Year Estimates dataset contains 25,070 variables across 1,128 tables. From the perspective of a researcher with limited background to ACS variables, it can be difficult to select the data elements that best align to their research question. While data can be easy to download in bulk, identifying and understanding where the data element of interest resides within specific tables can be difficult. The breadth, depth, and complexity of data available in the ACS introduce potential barriers which raise the threshold of knowledge required to incorporate valuable community level variables into project specific datasets.

Solution
To support researchers in their ability to incorporate ACS variables into observational research studies, the team created a new application (built with R shiny) that guides researchers through the hierarchy of ACS content. This interactive tool, named Quick Use of ACS (QUACS), provides a review the 30 ACS tables subjects in an appealing visualization whereby users may drill down into layers of desired information. A resulting list of variables can be exported from QUACS or downloaded directly from the ACS API.

Conclusion
Creation of a tool that allows researchers to rapidly identify and extract ASC variables greatly lowers the threshold for incorporating these important social variables into observational research. With the QUACS, researchers can quickly reduce the 25,070 values to the ten to twelve variables that are relevant to their research study.

References