Developing the Population Cancer Assessment and Surveillance Engine (PopCASE)

J. Rose, S. Sahoo, K. Lyytinen, F. Hussain, A. Sarangadharan, G. Kumar, H. Menegay, S. Song, R. Lanese, L. Liu, J. Tsui, S. Koroukian

Case Comprehensive Cancer Center

1. Background
The value of spatially linking cancer registry data to place-based contextual data is increasingly recognized. In addition to enhancing the utility of registry data generally, such linkage can help fulfill the mandate of National Cancer Institute (NCI)-Designated Cancer centers to address the unique cancer burdens of their geographically designated catchment areas in their research and community outreach and engagement activities. However, many registries and cancer centers may lack the substantial resources required to build a registry-centered, multilevel population cancer database linked with multiple space- and place-based data sources.

2. Goals
We describe the ongoing, NCI-funded development of a population cancer data platform, the Population Cancer Assessment and Surveillance Engine (PopCASE), which spatially links patient-level cancer registry data with community data describing the social and health care environment of patients, and an associated application—PopCASE Setup—that will allow central cancer registries or cancer centers with access to patient-level registry data to establish a region-specific instance of PopCASE.

3. Solutions and Methods
PopCASE consists of a PostgreSQL relational database management system, a user interface with support for querying and viewing the data, and a "controller" layer where calculations needed to generate certain results (e.g., age adjustment for user-specified patient subgroups) occur.

PopCASE links registry records by patient residential address to demographic, socioeconomic, household, transportation, risk factor, screening, health care access, and other community metrics. Users can query the data by demographic, geographic, or disease criteria and receive results including case counts, age-adjusted incidence and mortality, stage at diagnosis, time-to-treatment, and dozens of community measures stratified by county, census place ("municipality"), Zone Improvement Plan code tabulation area (ZCTA), and census tract. PopCASE can also be used to generate patient-level research data sets. The use of only data sources in standardized formats universally available across the United States has been a central design principle. PopCASE Setup is a software container packaging the required public data elements and managing the installation of a region-specific PopCASE instance.

4. Outcomes
Eighteen months into the PopCASE project, the database model has been built (see Figure). The controller logic is specified, and the preliminary user interfaces are designed – all under the guidance of a nationwide steering committee.

Here, we provide some statistics describing OH-CASE, the Ohio-specific implementation of PopCASE residing in the Case Western Reserve University Secure Research Environment. OH-CASE contains population data for approximately 11.8 million Ohioans living in 88 counties, 1,204 census places/municipalities, 1,197 ZCTAs, and 2,952 census tracts. Community data from the sources shown in the figure are available for each of these geographic units. The database contains registry records for 880,463 Ohioans with cancer

diagnosed from 2010 through 2022 from the Ohio Cancer Incidence Surveillance System. The total storage requirement for the database is approximately 18 gigabytes.

5. Lessons Learned and Future Directions

The next steps are to revise the interface based on steering committee feedback, complete the controller logic for age adjustment and other real time calculations, and build PopCASE Setup. In the final year of the project, we will use PopCASE Setup to build an instance of PopCASE in a catchment area outside of Ohio.

Figure: Database Model