An Informatics Infrastructure to Catalyze Cancer Control Research and Practice

Johnie Rose (✉ johnie.rose@case.edu)

Case Western Reserve University School of Medicine <https://orcid.org/0000-0001-7458-9303>

Weichuan Dong

Case Western Reserve University School of Medicine

Uriel Kim

Case Western Reserve University School of Medicine

Joseph Hnath

Case Western Reserve University School of Medicine

Abby Statler

Taussig Cancer Institute

Paola Saroufim

Case Western Reserve University School of Medicine

Sunah Song

Case Western Reserve University School of Medicine

Mustafa Ascha

Case Western Reserve University School of Medicine

Harry Menegay

Case Western Reserve University School of Medicine

Ye Tian

Case Western Reserve University School of Medicine

Mark Beno

Case Western Reserve University School of Medicine

Siran Koroukian

Case Western Reserve University School of Medicine

Research Article

Keywords: Cancer Control, Cancer prevention, Disparities, Database, Community-partnered research, Cancer registry

Posted Date: May 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-464846/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Cancer Causes & Control on April 5th, 2022. See the published version at <https://doi.org/10.1007/s10552-022-01571-0>.

Abstract

Purpose

A disconnect often exists between those with the expertise to manage and analyze complex, multi-source data sets, and the clinical, social services, advocacy, and public health professionals who can pose the most relevant questions and best apply the answers. We describe development and implementation of a cancer informatics infrastructure aimed at broadening the usability of community cancer data to inform cancer control research and practice; and we share lessons learned.

Methods

We built a multi-level database known as The Ohio Cancer Assessment and Surveillance Engine (OH-CASE) to link data from Ohio's cancer registry with community data from the U.S. Census and other sources. Space- and place-based characteristics were assigned to individuals according to residential address. Stakeholder input informed development of an interface for generating queries based on geographic, demographic, and disease inputs and for outputting results aggregated at the state, county, municipality, or zip code levels.

Results

OH-CASE contains 723,410 patient records for Ohioans diagnosed with cancer from 1/1/2006 – 12/31/2017 across 88 counties containing 1215 municipalities and 1197 zip codes. Stakeholder feedback from cancer center community outreach teams, advocacy organizations, public health, and researchers suggests a broad range of uses of such multi-level data resources accessible via a user interface.

Conclusion

OH-CASE represents a prototype of a transportable model for curating and synthesizing data to understand cancer burden across communities. Beyond supporting collaborative research, this infrastructure can serve the clinical, social services, public health, and advocacy communities by enabling targeting of outreach, funding, and interventions to narrow cancer disparities.

Introduction

Among the themes identified in the 2017 Unified Strategy Statement for Health Disparities Research from the American Association for Cancer Research (AACR), American Cancer Society (ACS), American Society of Clinical Oncology (ASCO) and U.S. National Cancer Institute (NCI)[1] was the need for understanding the interplay of factors at multiple levels—from molecular and clinical to socioeconomic and environmental—in driving differences in cancer outcomes across communities. The 2017 Statement further identified the need for timely feedback of actionable data to key stakeholders as part of a cycle of information, action, and monitoring.[1] This feedback can empower community partners to better direct

their interventions and advocacy. In addition, this process of feeding back data can foster a shared approach to posing research questions and interpreting results as part of community-partnered research. [2, 3] When deep understanding of methods and data limitations, as possessed by many in academia, meets the operational experience, cultural familiarity, and social capital of partners embedded in the community, powerful synergies can emerge. These synergies offer the potential to build academic understanding of cancer disparities while enhancing operational effectiveness of community partners who are fighting the cancer control and prevention battle on the ground.

The Public Health Informatics Institute (PHII) and Institute for Alternative Futures (IAF) recommends the creation of queriable data resources accessible by non-academic partners.[4] Federal[5, 6] and state[7] efforts to make county-level cancer burden and social determinants of health data available publicly are valuable public resources, but these portals may not provide the granularity needed for research, operational planning, outreach, and advocacy in community cancer control.

Here, we describe our approach to development of an evolving cancer informatics infrastructure in Ohio designed to support community-partnered cancer control research and to serve as a resource for stakeholders who need access to digestible, sub-county, multi-level cancer and community data to inform their activities. Importantly, this infrastructure was built using widely available data sources and open source development tools. The prototype represents a model which is a) expandable to include other individual- and community-level data sources and b) transportable to other regions and states.

Methods

We built the Ohio Cancer Assessment and Surveillance Engine (OH-CASE), a Structured Query Language (SQL) relational database covering the state of Ohio. In addition, we built a point-and-click user interface which facilitates rapid searches of the data by those unfamiliar with SQL. Figure 1 provides an overview of the OH-CASE architecture as described in the following sections.

In order to create a prototype that is reproducible in different locales, we used constituent data sets which are commonly available for most locations in the U.S. and many other countries. OH-CASE is housed in the *Secure Research Environment* (SRE) of Case Western Reserve University. The SRE is a computing environment governed by a risk-based security program that includes implementation of controls meeting recommendations or requirements of regulatory and information security standards including the Health Insurance Portability and Accountability Act (HIPAA), the Federal Information Security Modernization Act (FISMA), and the International Organization for Standardization (ISO). Similar environments exist at other research institutions and academic medical centers

Data Sources

The Ohio Cancer Incidence Surveillance System (OCISS) – OCISS is Ohio's National Program of Cancer Registries (NPCR)[8] supported and North American Association of Central Cancer Registries (NAACCR)[9] certified cancer registry. We obtained registry data for cases diagnosed from 2006 to 2017 (the most

recent year for which data are available) under a data users agreement (DUA) with the Ohio Department of Health (ODH). OCISS contains fields for direct patient identifiers, demographics, insurance type, geocoded address at diagnosis, cancer site, stage, tumor size and histology, receptor status (in the case of breast cancer), initial treatment modalities (including information on positive and negative lymph nodes excised), vital status (including cause of death, if applicable), occupation, and other domains. This individual-level data is linkable to any of the below community-level data sources based on geocoded patient address at diagnosis.

Five-year U.S. Census American Community Survey (ACS) data – The U.S. Census Bureau conducts an ongoing survey of Americans which is deployed to over 3 million households annually. This data is used to provide more timely estimates of key census parameters between decennial censuses.[10] We use published ACS data for five-year periods since estimates over shorter intervals would not provide figures for areas as small as census block groups.[11] These data provide contextual understanding of the demographic (age, race, ethnicity, sex, marital status) makeup, economic circumstances, housing environment, household composition, residential mobility/stability, educational attainment, and transportation resources of neighborhoods. In OH-CASE, we include five-year ACS data tables at the county, zip code tabulation area, census place (which we refer to as ‘municipality’), and census block group levels. A census block group is a subdivision of a census tract and contains anywhere from 600-3,000 people, most typically around 1,500; the census block group is the smallest unit for which sample-based estimates are available between decennial censuses.[12] In a city, this can represent an area of just a few city blocks.

Health Resources and Services Administration (HRSA) Health Professional Shortage Area (HPSA) Data – This publicly-available resource lists geographic divisions which are considered designated health professional shortage areas for primary care, mental health, and dental care based on provider to resident ratios. A HPSA score for each designated health professional shortage areas reflects the degree of shortage of the respective service type.[13] Linking to OCISS data based on geocoded address at diagnosis will allow identification of cancer patients living in HPSA-designated communities, and the degree of shortage based on HPSA score.

Food and Drug Administration (FDA) Certified Mammography Facilities – Updated several times each year, this publicly-available database contains the street addresses of all mammography facilities certified by the American College of Radiology or by states. Such certification is a requirement for providing mammography. [14] This data will allow quantification of mammogram density in a given geography or calculation of distance or travel time to mammography for members of a population of cancer patients.

User Interface Development

We designed a user interface to allow stakeholder groups to easily search elements of the OH-CASE database to fulfill common use cases. The interface was designed with four types of users in mind: health system community outreach staff, advocacy organizations, public health agencies, and

researchers. Hosting focus groups among each type of user, we developed a series of use cases that informed user interface design and prompted discussion of the types of search criteria, results fields, and levels of data aggregation that would provide greatest utility to each group. We presented version 1.0 of the OH-CASE interface to members of each user subset individually or in focus groups to obtain feedback which informs version 2.0 requirements.

The interface itself was designed in RStudio (version 1.2.1335) and a free version of R Shiny package (version 1.2.0). The choice of open source software is intended to foster sustainability and reproducibility.

Results

OH-CASE contains 723,410 unique patient records for Ohioans diagnosed with cancer from 1/1/2006–12/31/2017. Of these, 0.25% had missing values for the key variables of address at diagnosis or diagnosis date. None was missing primary cancer site. The data spans all 88 Ohio counties, 1215 municipalities, 1197 zip codes, 2952 census tracts, and 9238 census block groups—an area containing 337 mammogram facilities.

Stakeholder Use Cases

Stakeholder input from the health system community outreach teams of two Cleveland, Ohio academic cancer hospitals, two not-for-profit cancer advocacy organizations, a large local public health agency, and the cancer research community at our institution informed development of the use cases described in Table 1.

Table 1
– Use case examples compiled based on stakeholder input

Stakeholder Group	Use Cases
Cancer center community outreach teams	<ul style="list-style-type: none"> • Prioritize communities needing services from a mobile mammography unit based on high rates of late-stage breast cancer diagnosis, low mammography facility density, and poor transportation access • Compare certain process measures, risk factors, and outcomes of patients across communities, benchmarked against state/national averages • Report outcomes for NCI-designated cancer center catchment areas
Cancer advocacy organizations	<ul style="list-style-type: none"> • Prioritize communities/subpopulations for advocacy and funding activities • Compile descriptive cancer burden statistics for a population of particular interest to a foundation grant funder • Provide data to inform government-targeted advocacy activities
Local public health	<ul style="list-style-type: none"> • Support applications for state and federal funding with cancer burden and demographic statistics • Support cancer cluster investigations stemming from citizen inquiries
Researchers	<ul style="list-style-type: none"> • Provide community-level data to local partners to stimulate hypothesis generation and inform study design as a component of community-engaged/partnered research • Identify the size and demographic makeup of relevant sampling frames as part of determining study feasibility • Compile inclusion enrollment data for population-based study proposals • Use demographic, process, and outcomes data to target primary data collection (e.g. semi-structured interviews) in disparities research

Initial Interface Design – The initial version of the interface allows construction of queries and reporting of results based on OCISS and Census data. The following were integrated into the v1.0 user interface design based on stakeholder input:

Geographic Aggregation Level – Users are first asked to select a geographic level at which they would like query results to be aggregated: one group (statewide), by county, by zip code, or by municipality.

Filter Criteria – Filter criteria from the categories below are applied to individual-level characteristics of OCISS patients. For any field besides date range, multiple options may be selected. Figure 2 shows the filter criteria fields of the OH-CASE interface.

- **Start/End Dates** – Users select a range of diagnosis dates. The interface does not offer the ability to specify exact beginning and ending dates. Specifying a very short date range in combination with

other highly specific filter criteria could greatly increase the risk of re-identifying individuals. Instead, we chose to provide dropdown menus for selecting starting and ending *quarter*.

- Rural/urban continuum – Options are “In counties not lying in a metro area,” “In counties lying in a metro area of 250,000 to 1 million population,” and “In counties lying in a metro area of greater than 1 million population.”
- County – The user can restrict searches to certain counties. We included a preset option for the 15-county region corresponding to our National Cancer Institute (NCI)-designated Cancer Center’s northeast Ohio catchment area.
- Age at diagnosis – Categories corresponding to 14 years and younger, 15–29, 30–44, 45–54, 55–64, 65–74, 75–84, and 85 and older.
- Sex – Female, male, Transsexual/Not otherwise specified, and Other/biologically non-binary
- Race – American Indian or Alaska Native, Asian, Black or African American. Native Hawaiian or Other Pacific Islander, White or Caucasian, Other, and Unknown
- Ethnicity – Hispanic or Latino, Non-Hispanic or Latino, and Ethnicity unknown
- Primary cancer site – We used standardized primary site definitions based on tumor location and histology as specified in the International Classification of Diseases for Oncology, Third Edition (ICD-O-3) / World Health Organization (WHO) 2008 definitions.[15]
- SEER Summary Stage – Options are in situ (excluded by default except in the case of bladder cancer[15]), local, regional, distant, and unknown. OCISS requires reporting of Surveillance Epidemiology and End Results (SEER) Summary[16] but not American Joint Committee on Cancer (AJCC) cancer[17] stage. However, if one were to implement this model in a state where reporting of the more granular AJCC staging is mandatory, the staging filter criteria are easily changed accordingly.
- Breast Cancer Receptor Status – This field is activated if breast cancer is selected as the primary site. Options consist of combinations of estrogen receptor (ER), progesterone receptor (PR), and human epithelial growth factor receptor 2 (HER-2) status as follows:

ER/PR negative, HER2 negative (Triple Negative)

ER and/or PR positive, HER2 negative

ER/PR negative, HER2 positive

ER and/or PR positive, HER2 positive

Search Results – Search results are displayed with one row per selected geographic unit (i.e. one row for the entire state, one row per county, one row per zip code, or one row per municipality). The bottom three rows of each result set contain the following, respectively:

- Total – A summary row for results aggregated across all preceding rows. Counts are summed; rates and proportions are calculated from case and population totals.

- Ohio benchmark – The corresponding total, when available, from the compiled OCISS data for the entire state of Ohio
- National benchmark – The corresponding total, when available, from the SEER 99% sample data set[17]

Users may select the columns which will be displayed. The columns displayed by default are shown in **Fig. 3** and fall into two broad following categories as follows:

Case Information

- Case counts – In addition to total case counts, crude case counts may be broken down (with percentages) by sex, age category, African American (the predominant minority racial group in Ohio) and Non-African American race, and ethnicity.
- Age-adjusted Incidence rate – U.S. Census American Community Survey population totals based on the user-selected geographies and age range serve as denominators in the calculation of incidence rate per 100,000. For county and larger geographies, age-adjusted incidence and 95% confidence intervals are calculated using the direct method of age adjustment[18]; with the 2000 U.S. standard population as the reference, which is standard practice.[19] For municipalities and zip codes, age-adjusted incidence and 95% confidence intervals are calculated using the indirect method[18], referenced to Ohio-wide age-stratified incidence. Though the indirect method does not as readily allow comparison to rates calculated from different data sets, this method was chosen for smaller geographies because, unlike the direct method, it does not require age strata-specific case counts. The small numbers of cases in individual age strata within small and/or sparsely populated areas could lead to unstable estimates with wide confidence intervals if applying the direct method.[18] Age-adjusted incidence by race and ethnic categories may also be displayed.
- Stage – The proportion of cases diagnosed at a distant (metastatic) stage and with unknown stage are displayed by default for each geographic unit. A high proportion of distant stage disease may indicate poor access to screening or medical care generally. Previous work has shown that patients who are older, African American, have complex care needs, or are insured by Medicaid are more likely to have cancer of unknown stage.[20, 21]
- Median time-to-treatment – Time to initial cancer treatment is operationalized as the date of tissue diagnosis subtracted from date of first treatment of any modality. For multiple cancer types, prolonged treatment delay may be a marker of low quality care or poor access.[22]
- Proportion uninsured – This output is included as a marker of financial access to health services.
- Mortality rate – Age-adjusted mortality rate per 100,000 population is calculated, with direct age adjustment for county-level and larger geographies, and indirect adjustment for smaller geographies (as described for age adjustment of incidence). Median follow-up time is also included to provide context to the mortality figures.

Population Information

- Overall population counts as well as population counts broken down by sex, race, ethnicity, and age categories
- Median household income
- Proportion of the community living in high-Area Deprivation Index (ADI) census block groups – The ADI is a widely-used index of social deprivation calculable from census variables. The index incorporates 17 separate factors covering domains of education, employment, income, housing (costs, crowding), and transportation access.[23] Deciles of ADI values are used where the highest possible ADI ranking is 10, representing the top ten percent of the most deprived block groups of all block groups in Ohio. As a scalable summary measure of deprivation which can be applied to any geography, we have included an algorithm which calculates the proportion of population within any geography living in census block groups within the ninth or tenth decile of ADI.

Results are exportable as a comma-separated value (CSV) file.

The OH-CASE prototype has so far been applied in a number of real-world uses within our community and institutions:

- Quickly accessing population data for grant writing purposes, including understanding local burden and potential study population size as well as completing inclusion enrollment reporting for a population-based study
- Supporting description of catchment area cancer burden by leadership of our NCI-designated Comprehensive Cancer Center
- Creating publicly accessible data briefs regarding specific cancer types
- Enabling original research[24–26]

The **Appendix** provides two use case examples for how OH-CASE may provide insight into variations in cancer burden to inform research or focus the efforts of agencies seeking to mitigate cancer disparities.

Discussion

Here, we have described OH-CASE, a prototype multi-level cancer data infrastructure with a user interface designed to put sub-county-level cancer and community data into the hands of researchers and stakeholders who can apply the data in their local work. This prototype, designed with stakeholder input and built using data sources and tools which are widely and freely available, can be implemented in diverse locales.

Such multi-level cancer data infrastructures offer the potential to achieve the widely-espoused[1, 2, 4, 27, 28] goals of empowering local cancer control practitioners with data and enhancing the value and effectiveness of community-academic collaborations. Public and private community agencies leverage finite resources as they attempt to achieve maximal impact within their domains. Academic researchers strive to create generalizable knowledge about the extent and drivers of disparate cancer burdens across

populations, while facing significant pressure to publish and maintain grant funding. While, ultimately, both groups share the same larger goals, the practical pressures on each create a misalignment. Community organizations have limited bandwidth to work with researchers, and there is a danger that these partners serve only as sources of data—a recipe for squandering social capital and damaging sustainability. We believe that accessible cancer and population data can serve as a nidus around which academic-community collaboration can coalesce, leading to better leveraging and application of complex data by community partners and to more contextually-grounded research questions, better interpretation, and improved dissemination by researchers.

From conception to implementation, local action to reduce cancer disparities is most likely to be effective when it is evidence-based but informed by the local wisdom residing in vulnerable communities.[1, 2, 28] For this reason, a number of academic-community partnerships have applied community-based participatory research (CBPR) principles to prioritizing community needs, identifying and implementing appropriate evidence-based interventions, and disseminating best practices.[3, 29–32] Academicians bring expertise in medicine, epidemiology, anthropology, and other fields to the table, providing varying degrees of structure, facilitation, and methodologic expertise. Community partners bring real-world expertise to the table based on their own lived experiences and those of the populations they serve. Their insights can shed light on the programmatic, practical, and human challenges facing the residents of vulnerable communities and the professionals who serve them.

Other federal and state efforts link cancer and community data and make them publicly accessible online. The majority, however, do not provide the ability to parse data at geographic levels smaller than the county. The National Cancer Institute (NCI), for example, offers publicly available interactive tools for mapping and tabulating state- and county-level data on cancer age-adjusted incidence, mortality, screening, and risk factors—filtered by age, sex, race, and ethnicity. These resources draw on data from the National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database, U.S. Census, Behavioral Risk Factor Surveillance System (BRFSS), and other public sources.[5, 6]

In Florida, the website SCAN 360[33] (<https://www.scan360.com/>), a project of Sylvester Comprehensive Cancer Center, serves as a publicly accessible resource for tabulating and visualizing statewide, county-level data on cancer burden, sociodemographics, environmental risk factors, and health behaviors. In Durham County, North Carolina, a restricted set of stakeholders including clinicians and public health practitioners can access the Geographic Health Information System. This innovative system is built from electronic health record (EHR) data from the Duke Health System, vital statistics, property tax, crime, environmental exposure data, and healthcare and community resource data. The Geographic Health Information System is particularly powerful for community health assessment purposes because the Duke system covers the majority of patients in Durham County. In a setting where individual-level clinical data cannot be compiled for a majority of the population, however, a tool such as this may be less useful for community health assessment and surveillance purposes.[34]

OH-CASE is a prototype for a more expansive, statewide infrastructure which will integrate new data sources and new features based on stakeholder feedback. An additional goal of this larger effort, which we refer to as CIDaSh (Cancer Informatics and Data Sharing), is to facilitate sharing and linking of research datasets. In the near term, planned data source additions include a statewide database of tobacco outlets and individual-level treatment quality indicators based on Medicare claims and enrollment data.[35] In other states, it is conceivable that available all-payer databases could be integrated into such an infrastructure.[36] In the medium to long term, we plan to add claims data from the state's breast and cervical cancer program (BCCP), disease-specific registry data (e.g., for Lynch Syndrome), air and water quality data, and community risk-factor data. In the coming months, we will add a feature to generate visualizations summarizing query results. Measures visualized will include age-adjusted incidence, proportion of cases presenting with metastatic disease, time-to-treatment, and cancer-specific mortality results; these visualizations will include charts as well as choropleth maps.

Insights and Lessons Learned

The challenges inherent in a project such as the one described here are many.

- Building the database and interface requires a team of appropriately qualified technical personnel with sufficient bandwidth guided by one or more content experts/champions with a thorough understanding of the source data. It is critical that the two sides agree on a plan for providing appropriately detailed written development specifications. Ambiguity in these specifications can cause unnecessary delays and wasted resources.
- Outputs must be validated systematically and on an ongoing basis. We devised a process for randomly selecting queries to be run using the OH-CASE interface with results validated against source data (the topic of a forthcoming manuscript).
- Substantial collective thought must go into governance—determining a process for controlling and monitoring access to the data and imposing reasonable obligations for appropriate use. We are currently in the process of providing access to our first group of beta stakeholder users within our Comprehensive Cancer Center, and establishing a process by which users will not need to be included as investigators in institutional review board (IRB) protocols based on masking of small numbers in output. Our team is not comfortable with making the OH-CASE interface publicly accessible at this stage. The granular nature of the data requires that users possess a level of understanding that will prevent misinterpretation or inappropriate dissemination of interpreted findings. Users will also likely need some level of access to OH-CASE administrators in order to support appropriate interpretation.
- Sustainability must be a consideration. After the setup phase—even if new data types are not subsequently added—periodic data updates and validation will be required. In addition, a process of granting and controlling user access must be maintained. Ideally, the needed resources will come from institutional support, meaning that champions will need to demonstrate a value proposition to institutional leadership. The mandate to show catchment area insight and impact can be a strong

motivator for National Cancer Institute (NCI) designated Clinical or Comprehensive Cancer Centers.
[37]

Conclusion

By broadening the user base and lowering barriers to usage, the model we have described here has the potential to increase the value of available cancer and community data, enhance the relevance of community-partnered research, and improve the effectiveness of cancer control efforts at the local level. These benefits can accrue through the synergies created when academic data stewards and methodologists include community-based practitioners in the full life cycle of data linkage efforts.

Declarations

Author Contributions: Johnie Rose, Weichuan Dong, Uriel Kim, Joseph Hnath, Uriel Kim, and Siran Koroukian contributed to project conception and design, as well as data acquisition. Database and interface development were performed by Paola Saroufim, Sunah Song, Mustafa Ascha, Harry Menegay, Yi Tian, and Mark Beno. Johnie Rose drafted the manuscript, and all authors provided input. Johnie Rose and Siran Koroukian obtained funding to support the work. All authors read and approved the final manuscript.

Funding: Development of the work described here has been supported by pilot funding from the Cleveland Translational Science Collaborative, the Case Western Reserve University Center for Reducing Health Disparities, and Susan G. Komen of Northeast Ohio. None of these supporters played a role in acquiring data, building the data resource described, writing the manuscript, or the decision to submit the article for publication.

Conflicts of interest/Competing interests: The authors report no conflicts of interest/competing interests.

Data Availability: All data sources cited in the study are publicly available except for individual-level data from the Ohio Cancer Incidence Surveillance System (OCISS), which is used under confidentiality and data use agreement with the Ohio Department of Health.

Code Availability: Code is available upon request.

Ethics Approval: This work was approved by the Institutional Review Boards of Case Western Reserve University and the Ohio Department of Health, and is governed by data use agreements with the Ohio Department of Health for the use of cancer registry data.

Acknowledgments: This study includes data provided by the Ohio Department of Health which should not be considered an endorsement of this study or its conclusions. In particular, we would like to thank Lynn Giljan and Roberta Slocumb at the Ohio Department of Health for their invaluable assistance.

References

- [1] B.N. Polite, L.L. Adams-Campbell, O.W. Brawley, N. Bickell, J.M. Carethers, C.R. Flowers, M. Foti, S.L. Gomez, J.J. Griggs, C.S. Lathan, C.I. Li, J.L. Lichtenfeld, W. McCaskill-Stevens, E.D. Paskett, Charting the future of cancer health disparities research: A position statement from the American Association for Cancer Research, the American Cancer Society, the American Society of Clinical Oncology, and the National Cancer Institute, CA. Cancer J. Clin. (2017). <https://doi.org/10.3322/caac.21404>.
- [2] S. Gehlert, R. Coleman, Using community-based participatory research to ameliorate cancer disparities, Heal. Soc. Work. 35 (2010) 304–9. <https://doi.org/10.1093/hsw/35.4.302>.
- [3] R. Vargas, A.E. Maxwell, A. Lucas-Wright, M. Bazargan, C. Barlett, F. Jones, A. Brown, N. Forge, J. Smith, J. Vadgamma, L. Jones, A community partnered-participatory research approach to reduce cancer disparities in South Los Angeles, Prog. Community Heal. Partnerships Res. Educ. Action. 8 (2014) 471–6. <https://doi.org/10.1353/cpr.2014.0063>.
- [4] M. Edmunds, L. Thorpe, M. Sepulveda, C. Bezold, D.A. Ross, The Future of Public Health Informatics: Alternative Scenarios and Recommended Strategies, EGEMs (Generating Evid. Methods to Improv. Patient Outcomes). 2 (2014) 1156. <https://doi.org/10.13063/2327-9214.1156>.
- [5] National Cancer Institute, State Cancer Profiles, (2020). statecancerprofiles.cancer.gov.
- [6] National Cancer Institute, SEER*Explorer, (2021). <https://seer.cancer.gov/explorer/> (accessed January 3, 2021).
- [7] Sylvester Comprehensive Cancer Center, Scan 360, (2020). <https://www.scan360.com/>.
- [8] Centers for Disease Control and Prevention, National Program of Cancer Registries, (2021). <https://www.cdc.gov/cancer/npcr/index.htm> (accessed January 3, 2021).
- [9] North American Association of Cancer Center Registries, Cancer Registry Standards, (2018). <https://www.naaccr.org/>.
- [10] United States Census, About the American Community Survey, (2021). <https://www.census.gov/programs-surveys/acs/about.html> (accessed January 2, 2021).
- [11] L.P. Dalzell, F.K.L. Tangka, D.S. Powers, B.J. O'Hara, W. Holmes, K. Joseph, J. Royalty, Data sources for identifying low-income, uninsured populations: application to public health—National Breast and Cervical Cancer Early Detection Program, Cancer Causes Control. 26 (2015) 699–709. <https://doi.org/10.1007/s10552-015-0571-y>.
- [12] Federal Register, Fed. Regist. 83 FR 5629 (2018) 56293–56298. <https://www.federalregister.gov/d/2018-24570>.

- [13] Health Resources and Services Administration, What is shortage designation, (2021).
<https://bhwh.hrsa.gov/shortage-designation/hpsas>.
- [14] U.S. Food and Drug Administration, Certified Facilities/Certificates, (2020).
<https://bhwh.hrsa.gov/shortage-designation/hpsas>.
- [15] World Health Organization, international classification of diseases for oncology, 3rd ed., 2008.
- [16] National Cancer Institute, SEER Program and Coding Manual, (2021).
<https://seer.cancer.gov/tools/codingmanuals/>.
- [17] American Joint Committee on Cancer, AJCC Cancer Staging Manual, 8th ed., 2018.
<https://cancerstaging.org>.
- [18] L.R. Curtin, R.J. Klein, Direct Standardization (Age-Adjusted Death Rates), Stat. Notes. (1995).
- [19] N.C. Institute, Standard Populations (Millions) for Age-Adjustment, (2020).
<https://seer.cancer.gov/stdpopulations/>.
- [20] S.M. Koroukian, F. Xu, H. Beaird, M. Diaz, P. Murray, J.H. Rose, Complexity of care needs and unstaged cancer in elders: A population-based study, Cancer Detect. Prev. 31 (2007) 199–206.
<https://doi.org/10.1016/j.cdp.2007.04.002>.
- [21] S.M. Koroukian, Assessing the Effectiveness of Medicaid in Breast and Cervical Cancer Prevention, J. Heal. Manag. Pract. (2003) 306–314.
- [22] T.P. Hanna, W.D. King, S. Thibodeau, M. Jalink, G.A. Paulin, E. Harvey-jones, D.E.O. Sullivan, C.M. Booth, R. Sullivan, A. Aggarwal, Mortality due to cancer treatment delay: systematic review and, BMJ. (2020) 4;371:m4087. doi: 10.1136/bmj.m4087. <https://doi.org/10.1136/bmj.m4087>.
- [23] G.K. Singh, Area Deprivation and Widening Inequalities in US Mortality , 1969 – 1998, Am J Pub Heal. 93 (2003) 1137–1143.
- [24] U. Kim, S. Koroukian, A. Statler, J. Rose, The Effect of Medicaid Expansion Among Adults From Low-Income Communities on Stage at Diagnosis in Those With Screening-Amenable Cancers, Cancer. 126 (2020) 4209–19. <https://doi.org/10.1002/cncr.32895>.
- [25] J. Rose, U. Kim, W. Dong, J. Hnath, S. Koopman, Y. Oliver, P. Sage, S. Koroukian, A Qualitative Study of Factors Affecting Timely Breast Cancer Treatment in a High-risk Cleveland Community, in: Case Compr. Cancer Cent. 4th Annu. Disparities Symp., Cleveland, OH, n.d.
- [26] J. Rose, W. Dong, U. Kim, S. Obeng-Gyasi, S. Koroukian, Medicaid expansion associated with earlier stage and improved reconstruction rates in low income breast cancer patients, in: San Antonio Breast Cancer Symp., Virtual, n.d.

- [27] R. Parrott, J.E. Volkman, E. Lengerich, C.B. Ghetian, A.E. Chadwick, S. Hopfer, Using geographic information systems to promote community involvement in comprehensive cancer control, *Health Commun.* 25 (2010) 276–285. <https://doi.org/10.1080/10410231003711755>.
- [28] M.H. Chin, A.E. Walters, S.C. Cook, E.S. Huang, Interventions to reduce racial and ethnic disparities in health care, *Med. Care Res. Rev.* 64 (2007) 7S-28S. <https://doi.org/10.1177/1077558707305413>.
- [29] C.D. Meade, J.M. Menard, J.S. Luque, D. Martinez-Tyson, C.K. Gwede, Creating Community-Academic Partnerships for Cancer Disparities Research and Health Promotion, *Health Promot. Pract.* 12 (2011) 456–62. <https://doi.org/10.1177/1524839909341035>.
- [30] J.H. Corbin, M.E. Fernandez, P.D. Mullen, Evaluation of a Community–Academic Partnership: Lessons From Latinos in a Network for Cancer Control, *Health Promot. Pract.* 16 (2015) 345–53. <https://doi.org/10.1177/1524839914558514>.
- [31] T.A. Wynn, C. Anderson-Lewis, R. Johnson, C. Hardy, G. Hardin, S. Walker, J. Marron, M. Fouad, E. Partridge, I. Scarinci, Developing a community action plan to eliminate cancer disparities: lessons learned., *Prog. Community Health Partnersh.* 5 (2011) 161–8. <https://doi.org/10.1353/cpr.2011.0013>.
- [32] L. Noel, F. Phillips, K. Tossas-milligan, K. Spear, N.L. Vanderford, S.G. Eckhardt, Community-Academic Partnerships: Approaches to Engagement, 2019 ASCO Educ. B. (2019).
- [33] Scan360, (2019). <https://www.scan360.com/> (accessed December 29, 2019).
- [34] M.L. Miranda, J. Ferranti, B. Strauss, B. Neelon, R.M. Califf, Geographic health information systems: A platform to support the “triple aim,” *Health Aff.* 32 (2013) 1608–1615. <https://doi.org/10.1377/hlthaff.2012.1199>.
- [35] T.C. Tucker, E.B. Durbin, J.K. McDowell, B. Huang, Unlocking the potential of population-based cancer registries, *Cancer.* 125 (2019) 3729–37. <https://doi.org/10.1002/cncr.32355>.
- [36] National Conference of State Legislators, Collecting health data: all-payer claims databases, (2019). <http://www.ncsl.org/research/health/collecting-health-data-all-payer-claims-database.aspx> (accessed December 29, 2019).
- [37] Department of Health and Human Services, Founding Opportunity Announcement: Cancer Center Support Grants (CCSGs) for NCI-designated Cancer Centers (P30 Clinical Trial Optional), (2019). <https://grants.nih.gov/grants/guide/pa-files/par-20-043.html> (accessed April 14, 2021).

Figures

Figure 1

OH-CASE Architecture – FDA = U.S. Food and Drug Administration; OCISS = Ohio Cancer Incidence Surveillance System; HRSA = Health Resources and Services Administration; ETL = extraction, transformation, and loading

Define your dataset

Visualize

Select population of interest

Describe how to aggregate your data.

Start by describing how to group the data (i.e. the categories over which to report numbers of cases):

Geographic level

Municipality

Next, choose how you want to filter the data to keep only certain patients:

Start Date

2006 Q1

End Date

2017 Q4

Rural Urban Continuum

County

Age at Diagnosis

Sex

Race

Ethnicity

Primary Site

SEER Summary Stage

Figure 2

OH-CASE interface filter criteria fields

Case Information:

- ☒ Crude Case Count
- ☐ Sex
- ☐ Age
- ☐ Race
- ☐ Ethnicity
- ☒ Age-Adjusted Incidence
- ☐ AAI-Race
- ☐ AAI-Ethnicity
- ☒ % Distant Stage
- ☒ % Unknown Stage
- ☒ Median Time to Treatment(Days)
- ☒ % Uninsured at Diagnosis
- ☐ Age-Adjusted Cancer-Specific Mortality Rate
- ☒ Median Follow-up Time(Years)

Population Information:

- ☒ Total Population
- ☐ % Female
- ☐ By Age
- ☐ % African American
- ☐ % Hispanic
- ☒ Median Household Income
- ☒ % Community in High ADI

Figure 3

OH-Column display options – The checked boxes are displayed by default.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [APPENDIX.docx](#)