

Ohio Opportunity Index



FY22 & FY23
Project
Updates

June 2023

**Ohio Colleges of Medicine
Government Resource Center**

Executive Summary

The Ohio Opportunity Index project was initiated in 2019 because of the need for indices capturing area-level opportunities for healthy living in Ohio. The indices produced by this project are intended to provide population-based health agencies, organizations, and individuals with tools to target assessments, interventions, and evaluations and allow synthesis of the social determinants of health in a more interpretable and actionable way. The project has produced both the Ohio Opportunity Index (OOI) and the Ohio Children's Opportunity Index (OCOI) as data products along with corresponding dashboards. All products continue to be updated to meet ongoing needs of the Ohio Department of Medicaid and other users of the products. This report is focused on the work done over the last two years: (1) exploring methodological and tool enhancements for both indices, and (2) updating the child-focused OCOI.

During fiscal years 2022 and 2023 (FY22 and FY23, respectively), we explored methodological and tool enhancements for both indices. These included:

- A survey designed to learn how Ohio experts believe domains of the OOI and OCOI ought to be weighted in the combination of the individual input measures into an index. The results suggested that, on average, the experts had a preference that differed from a uniform weighting scheme. Notable deviations are that the transportation, environment, and crime domains should be weighted lower than others, while employment, education, and health domains should be weighted higher.
- Quantitative exploration of three different domain weighting schemes. The results showed few differences between the OI scores resulting from the weighting schemes, leading us to choose the simplest approach of equally weighting domains in the combination of data into the indices.
- Evaluation and selection of an improved missing data imputation method. The results suggested that a well-balanced approach to missing data imputation is to average the values of neighboring census tracts.
- Increasing usability through the production of a zip-code version of the OOI.
- Searching for neighborhoods by commonly known names within the dashboard OOI and OCOI exploration tools.
- The development of a web portal to increase accessibility of OOI and OCOI related data and tools. The site provides a brief description of the project and tools, and it hosts OOI and OCOI exploratory dashboards, data sets, tutorials, and reports.

Finally, the primary work during FY23 has been updating the existing OCOI v1 to OCOI v2. The OCOI is a child-focused version of the original OOI (which we often refer to as the general-purpose OOI to appropriately differentiate it from the OCOI). The OCOI is distinguished from the OOI in several important ways:

1. We split the OOI's Health domain into two separate domains, Infant Health and Non-Infant Children's Health.
2. We renamed and broadened the Transportation domain to be called the Access domain, which covers a similar conceptual space, but also includes food and child health care access.
3. We added a Family Stability domain, capturing household income, family structure, and adult mental health and substance use diagnoses.

4. Globally across all domains, the initial development of the OCOI included revision of input measures that could be more child focused.

In addition to reporting on methodological explorations and enhancements, this report documents the process by which the OCOI, initially created in 2021, largely with data as current as 2017, has been updated in 2023. The update includes current input data sets (as current as 2022 in some cases), revision of the measure choices for the purpose of increasing the reliability and availability of input measures in future years, and an improvement to the missing data imputation methods, making the tool more robust for the intended use cases.

When we explored the validity of the OCOI v2, the results provided evidence the tool is a valid representation of factors that explain geographic patterns of health outcomes in Ohio.

Full Report

Contents

Introduction 4

Methodological and Tool Enhancement Activities and Findings 4

 Domain Score Weighting..... 4

 Missing Data Imputation..... 6

 Ohio Opportunity Index Zip Code Version 6

 Neighborhood Search in the OOI and OCOI Dashboards..... 6

 Web Portal for OOI and OCOI Resources 7

Updating the Ohio Children’s Opportunity Index: OCOI v2 7

 Methods..... 7

 Results 10

 Discussion..... 11

Introduction

The Ohio Opportunity Index project was initiated in 2019 because of the need for indices capturing area-level opportunities for healthy living in Ohio. The indices produced by this project are intended to provide population-based health agencies, organizations, and individuals with tools to target assessments, interventions, and evaluations and allow synthesis of the social determinants of health in a more interpretable and actionable way (e.g., see the “Bright Spots” report [published](#) along with the OCOI v1).

The Ohio Colleges of Medicine Government Resource Center (GRC) currently maintains two indices: the Ohio Opportunity Index (OOI) and the Ohio Children’s Opportunity Index (OCOI). Inputs into the indices are theoretically derived from and are organized into several conceptual domains—seven in the OOI and eight in the OCOI—the labels for which vary slightly by the index. The resulting data products are data sets of individual domain scores and the overall Opportunity Index scores, one each for every census tract in Ohio. The OOI is intended as a general-purpose index to be used with all populations in Ohio. The OCOI is a child-focused version intended to be used in projects focused on children.

The Ohio Opportunity Indices have an advantage over related indices available at the national level (e.g., CDC Social Vulnerability Index). Because the indices are focused on Ohio, they are not limited to incorporating health-related measures that are available for all states in the USA. Therefore, the Ohio OOI and the OCOI also incorporate Medicaid claims, Vital Statistics, and Department of Education derived measures. These Ohio-focused inputs support better delineation of areas that facilitate good health as well as those in need of more health-oriented resources. In sum, the Ohio Opportunity Indices are important tools for Ohio policy thinkers concerned with ameliorating disparities in health outcomes.

This report details project activities occurring during FY22 and FY23 for the GRC’s Ohio Opportunity Index project. The activities can be organized into two classes: (1) methodological and tool enhancement activities, and (2) updating the OCOI.

Methodological and Tool Enhancement Activities and Findings

Several project activities related to improving the construction or use of the Opportunity Indices occurred over the entire FY22 and FY23 period. This section details each activity.

Domain Score Weighting

Following the development of the first OOI, the GRC team discovered a theoretical issue with the use of single dimensional latent factor analysis (of which principal components analysis is a special case) to define the appropriate weighting of each domain in their combination to form the overall Opportunity Index. In brief, the results of this type of approach will reflect *lower* weight associated with domains that carry information that is *different* from the remainder of the domains. Factor analysis is the right tool to use when designing a measurement instrument that is intended to measure one single latent dimension (e.g., cigarette craving or depression). However, this is not consistent with the aims of the OOI and OCOI, which is to capture many dimensions that bear on health, and which may balance each other in an overall score. The weights we use should represent the degree to which the domains contribute towards health outcomes. The latent factor approach has no health outcome information on which to anchor its estimation of weights for the combination of the domains into an overall score. Thus, a factor-analysis-

focused approach to weighting is theoretically problematic for the OOI and OCOI development and maintenance.

Discussion of alternatives revealed three options. The first, and simplest option, is to rely on the theoretical underpinning of the domain choices and weight all domains equally. The limitation of this approach derives from the plausibility that the different domains, while all theoretically relevant, may contribute to health to different extents. However, given the complex pathways through which any factor may affect health, it is very hard to quantify the effect of a single domain. The second option is to build regression models of multiple health outcomes as a function of standardized domain scores. The relative contributions of the domain scores towards explaining the outcomes could be used to construct weights. This proved challenging due to complications in finding multiple health outcomes that are not already a part of the indices.

The third option involves asking experts in fields related to the social determinants of health to give their opinions about what the relative weights ought to be. We explored this option but ultimately decided not to use it. In 2021, the team constructed an email list of known peers in Ohio that work in a field related to the social determinants of health. An online questionnaire was distributed to this convenience sample of email addresses. The questionnaire required participants to distribute 100 points across the seven domains of the OOI according to their beliefs about how strongly those domains contribute to health outcomes. Table 1 shows some of the results of this brief study. We found that the respondents did believe that some domains should be weighted differently than others. Notably, they thought the transportation, environment, and crime domains should be weighted lower than others, while employment, education, and health domains should be weighted higher. Table 1 frames this in terms of percentage deviations from a purely uniform distribution of weights, which would be 14.3 points (100/7) for each domain. The maximum observed deviation was a -37% deviation for the crime domain, suggesting that the respondents believe the crime domain has the smallest impact on health among the identified domains.

Table 1. Average expert-assigned weights to the OOI domains and their percentage deviations from a uniform distribution of weights (14.3 points for each domain).

	Employment	Transportation	Crime	Housing	Environment	Health	Education
Average points assigned	17.8	10.5	8.9	15.3	11.4	18.0	18.2
Percent deviation from uniform weights (14.3 points)	24%	-27%	-37%	7%	-20%	25.7%	27.6%

Due to the differences observed between a uniform weighting and expert beliefs about contributions towards health of each domain, the expert beliefs became a significant consideration for weighting domains in their combination for an overall index. However, statistical and geographic comparisons of the overall OOI using each weighting scheme suggested that these weight differences had little effect on the overall score. Because the differences were minor and because of the weaknesses (convenience sample) of our expert study, we ultimately chose to use the simpler option of uniform weighting of domains.

Missing Data Imputation

We studied multiple approaches to missing data imputation. The approaches averaged scores in other tracts that were similar to the tract with the missing value. Similarity was determined by geographic proximity, similarity with respect to other measures that had no missingness, and a combination of these two. To evaluate these options, we trained support vector machine models on random subsets of the data for the OCOI v1, then used those models to classify tracts with complete information as either missing or present for each measure. Tracts that were most frequently classified as missing had their value for that measure removed, to create a dataset with known, but hidden values. We retained the true (hidden) value for later checking. We refer to these as the test tracts. The usefulness of this approach was to ensure our test tracts were similar to tracts with actual missing data. We then applied the missing value imputation techniques to the test tracts. For each technique, we checked the distribution of the difference between the true value and the imputed value. The criteria we used to determine the best approach was two-fold. (1) We calculated the average of the difference between the true value and the imputed value. For this check, zero is an ideal and unbiased imputation. However, unbiased does not imply perfect because two very poor imputed values that are far from the truth in opposite directions could average to zero. Thus, we also calculated (2) the variance of the differences between imputed value and truth. For this metric, zero is also ideal and indicates reliability in the imputed values. A zero in both bias and variance implies every value was perfectly imputed. These two criteria (bias and variance) often oppose one another (one going down commonly means the other goes up) and our goal was to find a balanced middle ground. Our conclusion was an average of geographically neighboring census tracts provided the most balanced result. We therefore used this approach in this update of the Ohio Children's Opportunity Index.

Ohio Opportunity Index Zip Code Version

Due to a specific need by the research group involved in the Infant Mortality Research Partnership (IMRP) project, the GRC OOI team produced a zip-code version of the Ohio Opportunity Index. The specific need was to reduce data collection burden on patients and the subtle data security issues involved with geocoding patient addresses to census tracts at run-time for the practitioner-facing perinatal risk calculator developed for the IMRP project. Translating a zip code to an Opportunity Index score simplifies the process since most people know their zip code but do not know their census tract ID number.

The approach to development of the zip-code version was very similar to that for the census tract version. The only difference was the underlying data elements were collected again at the zip-code level of geography. No other part of the process changed.

Neighborhood Search in the OOI and OCOI Dashboards

The OOI and OCOI dashboards were designed to help policy thinkers—particularly those without a background in geographic information systems and analytics—geographically explore the OOI and OCOI data and see patterns that help them think through possible solutions to problems they are facing. A significant limitation of the OOI and OCOI dashboards in this regard is most people, be they policy thinkers or interested citizens, do not think geographically in terms of census tract IDs. Rather, they think in terms of cities, counties, and neighborhoods. The dashboards cover cities and counties well through the county selection options. However, they did not originally have any capacity to highlight neighborhoods by a searched name. In FY22, neighborhood search options were added to solve this

problem. The neighborhood search option opened the door to searching by both neighborhood name and zip code. When a neighborhood or zip code is selected, all census tracts that have overlapping land area with the neighborhood become highlighted in the dashboard, making it easier and faster to explore the area in which the user is interested.

Web Portal for OOI and OCOI Resources

As the OOI and OCOI tools became more known and used, and in response to the many questions and requests arriving for reports and data, the GRC team developed a web page designed to provide most of the requested information to users within a few simple clicks. The web page directly provides brief descriptions of the project, the data products, and the dashboards. It contains links to many useful OOI and OCOI resources, including the current data with quick start guidance, methodological reports, the dashboards, tutorial videos for the dashboards, and contact information for the GRC OOI principal investigators. The web portal is available at <https://grc.osu.edu/Projects/OhioOpportunityIndex>.

Updating the Ohio Children's Opportunity Index: OCOI v2

The Ohio Children's Opportunity Index was initially created in 2021 as a complement to the Ohio Opportunity Index (OOI; [web link](#)). The OOI is intended to be reflective of the geographically distributed factors that influence the health of the general population in Ohio. The OCOI is structurally and operationally very similar to the OOI but has a refocused set of input measures and conceptual domains that are more reflective of factors that bear on, or are indicative of, the health of Ohio's children. The OCOI was developed in collaboration among the Ohio Department of Medicaid (ODM), the Ohio Colleges of Medicine Government Resource Center, and The Ohio State University (OSU) Department of Geography. In conjunction with the OOI, the goal of the OCOI is to support improvements to children's health and well-being and a mission of quality, equity, and population health.

The OOI team aims for a 2-year update cycle of the OOI and OCOI (staggered). This update schedule is intended primarily to update the input measures to the most current available, given attention to standardization of measurement time across measures (i.e., not too many measures come from different sets of years). Additionally, as we learn new approaches to the construction of health indices, we will also update the methods to render a higher quality product. The purpose of this report is to describe the refined methods used to create the updated version 2 (OCOI v2) of the OCOI and provide evidence that it is valid. The major changes to OCOI v2 ([OCOI v1 methods web link](#)) include the following: (1) revisions to the input measures with the goal of including measures that are expected to be more reliably available in the future and therefore maintain a more consistent product; (2) substantial improvements to the missing-data imputation methods to cope more optimally with the common problem of missing geographic data.

Methods

Data and Structure

The Ohio Children's Opportunity Index is composed of fifty-five constituent measures grouped into eight broad categories of opportunity, called domains. These domains were decided before data collection and include: (1) family stability, (2) infant health, (3) child health, (4) access, (5) education, (6) housing, (7) environment, and (8) criminal justice. All domains are aggregates of six to eight constituent measures

with census tract level granularity. The overall opportunity index score is an aggregation of all domain scores.

Conceptually, the domains were selected because of their theoretical relationship to health opportunity. *Family stability* provides consistent safety, resources, and parental support for children to develop healthily. *Access* to healthcare, food, jobs, and a vehicle reflect the ability of people in an area to get the resources they need to survive and be healthy. *Education* captures the availability of resources and the will for kids to learn, including enrollment rate, school performance, third grade reading ability, graduation rate, and education among the area population. *Housing* captures the quality of living conditions, residential mobility, housing cost, and vacancies, which relate to the extent that people in the area are secure with respect to having a consistent place to live. *Environment* reflects the physical conditions of an area that relate to health, including access to parks, walkability, cancer and non-cancer risks of the air quality, the likelihood of lead in the housing supply, and proximity to toxic algal blooms. Criminal justice refers to the degree of reported crime of various types in an area, which can create stressful living circumstances for people of all ages. *Infant health* and *children's health* represent health outcomes of children, reflecting latent conditions of the area, whether physical or social, that have historically led to the observed outcomes.

A list of all fifty-five constituent measures can be found in Appendix A, along with details including a quantitative description of each constituent measure, what domain it belongs to, and the source of the data.

Procedures

After collecting data for all constituent measures, a series of steps were followed to transform and combine data into domain scores and ultimately the Ohio Children's Opportunity Index. The approach was derived from the work of Noble (2006), which is consistent with how both the Ohio Opportunity Index and Ohio Children's Opportunity Index have historically been constructed:

1. **Missing Data Imputation:**

In the case that a constituent measure is missing or cannot be computed or assessed in a particular census tract, then it is necessary to impute a reasonable estimate of the value of that constituent measure for that census tract. In this situation, the average value of the constituent measure among neighboring census tracts was used as the imputation value for the missing census tract. See a previous section titled *Missing Data Imputation*.

As of 2020, there are fifteen census tracts with a population of zero. These 15 tracts are usually distinct from other census tracts. For example, some are airports or universities where many constituent measures may be missing or may simply not be relevant (for example, distance to supermarket). Because these census tracts do not represent any population, we initially set the values of all constituent measures of these census tracts to missing, so that they will not impact the calculation of the domain and overall opportunity scores for other tracts. These tracts will be addressed again in the "Create Domain Scores" step.

2. **Measure Transformation:**

All constituent measures should be weighted equally when aggregated together to form the domain and overall opportunity scores. To ensure equal weighting, it is necessary to transform all constituent measures so they all have the same distribution. This transformation was done in

two steps: (1) The “direction” of each constituent measure should be the same, meaning a large value should have the same meaning in terms of opportunity in all constituent measures. Since it was already true of most constituent measures, it was decided that higher values should represent lower opportunity. For constituent measures in which that was not the case, values were multiplied by -1 so that larger values would be correctly associated with less opportunity. (2) Each constituent measure was converted into a z-score so the measure was centered around zero and scaled to have a variance of one.

3. Create Domain Scores:

For every census tract, the values of all transformed constituent measures within each domain were summed together to create a single value for each domain. The summed values were then ranked from 0 to 3167, where a rank of 1 would be assigned to the census tract with the smallest domain value (most opportunity). Tracts with missing values (those with 0 population) were assigned the rank of NA, but are listed last, meaning that the ranking stops at 3152 (3167 – 15). This creates an approximately uniform distribution of scores for each domain. To force the range of ranked values to be greater than or equal to 0 and less than 1 (creating a standard uniform distribution), each ranked value was divided by 3152 to form what we will call “adjusted rank scores.” Lastly, to create the domain scores, the standard uniform distribution of each domain’s adjusted rank scores is transformed into an exponential distribution using the following formula:

$$\text{domain score} = -23 * \log(1 - \text{adjusted rank score} * (1 - e^{-\frac{100}{23}}))$$

At this point, census tracts with zero population will have a missing domain score. Rather than leaving this value as missing, we want to assign these tracts a reasonable domain score. This assignment is done by taking the average of the domain score of the neighboring tracts and imputing that value.

4. Create Overall Opportunity Score:

For each census tract, domain scores were averaged together to form the deprivation index. Higher values of the deprivation index indicate less opportunity. So, by multiplying the deprivation index value by -1, we create the overall Ohio Children’s Opportunity Index, where a large value represents greater opportunity.

Validation

The final OCOI v2 scores were validated using two methods. (1) We assess the construct validity of the OCOI v2 by examining the correlation between OCOI scores and other metrics we expect would vary based on opportunity. Based on availability of data, we calculated the correlation between the OCOI and four metrics: death rates, child asthma rates, child severe mental illness rates, and pre-term birth rates. Child asthma rates, child severe mental illness rates, and pre-term birth rates are all constituent measures of the OCOI. This makes these three metrics weak validation metrics (as a correlation between these and the overall opportunity index should exist). However, death rates were not included in the construction of the OCOI; thus, they supply a good assessment of construct validity. (2) We assessed the convergent validity of the OCOI v2 with other opportunity scores. We examined the correlations between this OCOI v2 and the (a) OCOI v1, (b) the current OOI v2, and (c) CDC’s Social Vulnerability Index (SVI). To compute this correlation, it was necessary to limit included census tracts to only those that had

the same GEOID in both the 2010 and 2020 census tracts. We would expect estimates of opportunity in this OCOI to be positively correlated with other estimates of opportunity and negatively associated with the SVI, which represents an opposing construct: vulnerability.

Results

Many of the results of the OCOI can be found in appendices A and B. Below, we describe some of the results and provide direction for where further information may be found.

Constituent Measures

Appendix A contains a complete list of all 55 constituent measures. This appendix also includes a description of each constituent measure, information about its source and distribution, the number of missing values the measure contains, as well as a choropleth map of the measure's geographic distribution. Appendix A also includes heatmaps which allow you to investigate the relative correlation between all constituent measures.

Domain and Overall Scores

Information about the construction of the Domain and Overall scores can be found in Appendix B, which includes both a description of the process and the R code used to construct all domain scores and the overall Opportunity Index.

Appendix B also includes information about the distribution of Domain and Overall scores, including histograms of untransformed and transformed domain scores, histograms of the deprivation and opportunity indexes, as well as a heatmap that visualizes the correlation between different domains.

From the heatmap in Appendix B, we can see the domains Family Stability, Education, Criminal Justice, and Housing are all strongly positively correlated. Meanwhile, the Access domain has a slight negative correlation with the Criminal Justice domain.

In Appendix B, one can also find choropleth maps showing the geographic distribution of the domain scores, the deprivation index, and the overall Opportunity Index.

Validation

Information regarding the validation of the OCOI v2 can be found in the tables below and in Appendix B.

Table 2. Assessment of construct validity.

Outcome	Correlation with OCOI (p-value)	Multiple Regression R-squared
Death Rates	-0.25 (< 0.005)	0.16
Child Asthma Rates	-0.48 (< 0.005)	0.35
Child Severe Mental Illness Rates	-0.03 (0.14)	0.28
Preterm Birth Rates	-0.45 (< 0.005)	0.34

Table 2 shows the correlation between the OCOI and four metrics we expect to be associated with opportunity. Of these four, only Death Rate was not included in the construction of the OCOI, and thus is the strongest validity check of the four. Here, we see Death Rate has a moderately weak, but significant, negative correlation with the Opportunity Index. We also see that Child Asthma Rate and Preterm Birth Rate are both moderately and significantly negatively correlated with the Opportunity Index. These

correlations indicate higher OCOI scores are associated with slightly lower death rates, and lower child asthma and preterm birth rates. We see a very weak and non-significant (at the 0.05 level) correlation between the OCOI and child severe mental illness. Apart from the non-significant negative correlation between the OCOI and child severe mental illness, these results are what we would expect of a measure of opportunity, which lends some construct validity to the Ohio Children’s Opportunity Index.

Table 3. Assessment of convergent validity.

Outcome 1	Outcome 2	Correlation (p-value)
OCOI v2	OCOI v1	0.77 (< 0.005)
OCOI v2	OOI v2	0.76 (< 0.005)
OCOI v2	CDC Social Vulnerability Index	-0.68 (< 0.005)
OCOI v1	OOI v2	0.77 (< 0.005)

Table 3 shows correlations among the OCOI v2, the OCOI v1, the OOI v2, and CDC’s Social Vulnerability Index. The OCOI is strongly and similarly associated with each of the other 3 measures in the expected directions. Notably, OCOI is negatively correlated with the social vulnerability index because while these measures are similar, they represent opposite constructs (vulnerability versus opportunity). Additionally, the reported association between the validated OCOI v1 and OOI v2 provides context for the strength of the other three relationships. The strong observed associations provide evidence for convergent validity of the OCOI.

Discussion

Since its inception in 2019, the Ohio Opportunity Index project and project team maintained and refined the OOI and OCOI products to best meet the needs of its original parent project, the Infant Mortality Research Partnership (IMRP), and its primary sponsor, the Ohio Department of Medicaid. The maintenance and refinement include keeping the indices up to date with the most current data, refinements to the procedures that combine the many data sources into an index, increasing accessibility of the data and tools with a web portal, improving usability of the raw data and the geographic information systems dashboard tools in response to user feedback, and creating alternative versions of the tool for a different geography (zip code) to solve data privacy issues and make the tools useful in more contexts.

Having worked through refinements and updates that resulted in the second version of both the OOI and OCOI, the GRC team is ready to begin a more consistent phase of updating both indices on a regular basis. Our plan is to update each index every other year in a staggered pattern. This year, the OCOI was updated. Next year, the OOI will be updated. Contemporaneous with this update pattern, and despite existing evidence that the tools are valid, we want to begin doing more work towards validating the tools and/or learning where weaknesses in the data tools may lie to improve the Opportunity Indices. The FY24-FY25 Opportunity Index Project Task Order includes several proposals for approaches to these goals. We will discuss them with the project sponsor, ODM, as we proceed into FY24.