

RESEARCH

Open Access



# Co-establishing an infrastructure for routine data collection to address disparities in infant mortality: planning and implementation

Naleef Fareed<sup>1,2\*</sup>, Christine M. Swoboda<sup>1,3</sup>, John Lawrence<sup>1</sup>, Tyler Griesenbrock<sup>1</sup> and Timothy Huerta<sup>1,2,3</sup>

## Abstract

**Background:** Efforts to address infant mortality disparities in Ohio have historically been adversely affected by the lack of consistent data collection and infrastructure across the community-based organizations performing front-line work with expectant mothers, and there is no established template for implementing such systems in the context of diverse technological capacities and varying data collection magnitude among participating organizations.

**Methods:** Taking into account both the needs and limitations of participating community-based organizations, we created a data collection infrastructure that was refined by feedback from sponsors and the organizations to serve as both a solution to their existing needs and a template for future efforts in other settings.

**Results:** By standardizing the collected data elements across participating organizations, integration on a scale large enough to detect changes in a rare outcome such as infant mortality was made possible. Datasets generated through the use of the established infrastructure were robust enough to be matched with other records, such as Medicaid and birth records, to allow more extensive analysis.

**Conclusion:** While a consistent data collection infrastructure across multiple organizations does require buy-in at the organizational level, especially among participants with little to no existing data collection experience, an approach that relies on an understanding of existing barriers, iterative development, and feedback from sponsors and participants can lead to better coordination and sharing of information when addressing health concerns that individual organizations may struggle to quantify alone.

**Keywords:** Public health informatics, Social determinants of health, Maternal health, Infant health, Surveillance systems

## Background

### Introduction

The use of community-based organizations to address social determinants of health has a long history [1]. Infant mortality represents one area where such an approach has been of common use [1–4]. Ohio is ranked as the

ninth-worst state in the United States for infant mortality in 2019, with an estimated rate of 7.3 infant deaths per 1000 births [5]. Further, there have been numerous reports that have identified the notable disparity in infant mortality between white and black infants; the infant mortality rate is almost three times higher among black infants in Ohio (13.9 per 1000 births for black infants in 2018 vs. 5.4 for white infants) [6–11].

In 2012, the state legislature renewed efforts to address this disparity by convening local partners in the nine counties in Ohio that account for the majority of black

\*Correspondence: Naleef.Fareed@osumc.edu

<sup>2</sup> Department of Biomedical Informatics, College of Medicine, The Ohio State University, Institute for Behavioral Medicine Research, 460 Medical Center Drive, Columbus, OH 43210, USA

Full list of author information is available at the end of the article



infant mortality. In addition, state agencies collaborated to use population data to target areas for outreach and services as well as coordinate efforts among the multiple programs in each community. Infant vitality efforts in these nine counties, also known as the Ohio Equity Institute (OEI) communities [12], include group-facilitated prenatal care, home visiting, progesterone administration, breastfeeding support, fatherhood programs, safe sleep, and health education and community engagement activities focused on changes such as tobacco cessation and birth spacing. Coordinating these efforts helped bring focus to the number of independent initiatives to improve infant and maternal health; however, the effort also highlighted an absence of consistent data collection and infrastructure across programs to facilitate evaluation.

In 2017, the Ohio Department of Medicaid (ODM) began to focus on the implementation of a consistent data collection and infrastructure across programs to facilitate evaluation despite access to Medicaid and population data. The vision refocused on using data to increase surveillance to target areas for outreach and services as well as the coordination of efforts among the multiple programs in each community. To address the lack of consistency across data collection, ODM developed the OEI Community-Based Organization (CBO) Evaluation project. The goal of the OEI CBO Evaluation project is to create a statewide system for routine data collection and conduct analysis on the data collected to determine the extent to which the selected interventions serve high-risk pregnant women and assess the effect of these interventions on health care utilization and birth outcomes. The research team is made up of researchers at The Ohio State University and the Ohio Department of Medicaid, both in Columbus, Ohio. The Ohio State team, composed of health and bioinformatics subject matter experts, was tasked with developing a multimodal data collection system. The research team worked closely with leadership individuals from the CBOs, who attended meetings in person, online, or via phone with the research team in Columbus to ensure the system was created with their input and needs considered.

At the start of the evaluation, the research team performed a needs assessment, which noted there was significant heterogeneity among the interventions, at times individuals participated in more than one program, and there was no standard for data collection. Without a multimodal data collection model necessary to support public health and surveillance of infant mortality, there was a great deal of disparity in data collection approaches and infrastructure. Further, there were no standards for reporting such community-based data. This paper describes the multimodal data collection model

necessary to support public health and surveillance of infant mortality and serves as an exemplar for how such efforts might be accomplished in other domains.

Taken together, concerns about the disparities in the existing approaches to data collection and the associated impact precipitated into a call to action by state agencies to develop a robust data collection infrastructure for OEI. The primary objectives of developing our data model for OEI were twofold: 1) to generate reports that quantify the performance of the CBOs funded by OEI based on key metrics and 2) to dynamically visualize these key metrics using a Tableau dashboard.

#### **Prior work**

As part of this project, we sought to design a data model that included characteristics we previously determined to support public health and surveillance of infant mortality in Ohio. Past research encourages using collaborative design processes to ensure user-centered design [13], as products developed with the needs of the end-user in mind improve acceptability and feasibility of their use [14]. Active data collection processes have been shown to improve prevention programs and increase evaluative capacity to address multiple public health problems [14–17]. A robust data collection system for OEI was determined to have the following characteristics: 1) multimodality of data collection [17–20], 2) a standardized data model [21–23], and 3) systematic educational outreach efforts to train those reporting data [24–26]. These characteristics are supported by literature recommending tailored data collection systems created with evaluation in mind, followed by capacity building and training for these data collection systems [15–17].

A multimodal system is one that allows for data collection and delivery across a number of communication channels: phone, email, fax, mail, and direct data entry. Multimodal systems allow for data collection across CBOs that vary in their technological capabilities and resources. CBO staff have varying levels of comfort with using tools to report data; some are more comfortable using information technology, while others gravitate to using paper forms. The final data are electronic, and paper forms have additional data entry and submission steps; however, many of the community organizations do not have the technology to collect data online, and it may be easier to collect data on paper during in-person visits with participants. Prior research indicates improved data quality when multiple data collection modes are available [17–20].

A standardized data model allows for the collection of a common set of core metrics across CBOs with heterogeneous activities and data collection practices, improving cross-program comparability [21].

The wide variation in the data collection among CBOs before the OEI CBO Evaluation project may have created challenges that were difficult for CBOs such as not directly targeting key interventions. Some CBOs may not have collected enough data to determine whether their participants were high risk, identify service provision needs, or evaluate program efficacy. It has been noted that instituting such a model, also known as a common data model (CDM), is challenging, especially with a diverse set of health care organizations that have varying levels of requisite resources to populate the CDM [22]. Notwithstanding, a CDM would enable new research opportunities that help with the improvement of interventions and assessment with outcomes that are based on data that were previously not available or limited to single site studies [23].

Systematic educational outreach efforts are necessary to train health care providers on how to report data to a CDM, maintain the fidelity of data collection over time, and foster relationships between community health organizations and research teams [17, 24–26]. One study demonstrated that use of training elements and data trainers allowed for public health workers to have better awareness of data reporting and higher data reporting rates than those who did not have such resources [25]. Another study listed the early misperceptions associated with routine data reporting – which included providers not realizing the value of data reporting, the inability to fit data reporting with existing workflows, and the lack of best practices on reporting data – and noted how these views diminished over time with the support of training and guidelines [24]. Effective public health surveillance requires training covering key concepts, terminology, policies, information technology, and practices related to data reporting [17].

#### **Goal of this effort**

This paper describes the design, development, deployment, and assessment of a data collection system to improve public health efforts by collecting data from multiple community programs. The goals of our effort were to minimize data collection burden while maximizing the ability to engage in evaluative assessment of the effect of participation in OEI CBOs. Our paper serves as a template for the development of similar systems, in other contexts and other locations, in which existing data infrastructure and inter-organizational coordination are insufficient to meet the needs of a public health effort. We describe the design of the OEI data collection system and data model that supports these goals. The system developed is still being used to collect data from CBOs across the state for the Ohio Department of Medicaid.

## **Methods**

### **Study population**

The OEI CBO Evaluation project focused on programs in nine counties in Ohio with large disparities in infant mortality and significant urban populations: Butler, Cuyahoga, Franklin, Hamilton, Lucas, Mahoning, Montgomery, Stark, and Summit. These include the urban areas in the cities of Hamilton, Cleveland, Columbus, Cincinnati, Toledo, Youngstown, Dayton, Canton, and Akron, respectively. Existing and newly developed community programs received funding to provide consistent data to ODM and could use this funding for hiring, training, and general program needs. The population involved in the development of data collection materials consisted of supervisors and program leaders from OEI CBOs, the research team at Ohio State, and the funders at the Ohio Department of Medicaid. The population assessed encompassed every participant who enrolled in one of the funded programs in the nine OEI counties. The OEI data collection system collects data only for those enrolled in one of the CBOs, but the population enrolled in these programs is largely high-risk, mostly minority women who have the highest need for interventions.

### **Prototyping the OEI data collection infrastructure: needs assessment and design**

The research team at Ohio State for OEI consisted of faculty, skilled professionals, and technical experts tasked to develop, deploy, and evaluate the data collection system. First, in early 2018, our team conducted a needs assessment by reviewing previous program data collection materials, reviewing CBO grant applications, conducting individual and group interviews with CBO leaders, and reviewing data collection materials from similar programs identified during literature review to identify program needs and conceptualize a multimodal data collection system. The semi-structured interview guide asked open-ended questions about how organizations were managed, data collection and reporting methods, frequency of data collection, types of data collected, and abilities to add data points or change data collection methods. Based on our engagements with the CBOs, the research team sought to sketch out for the communities what a robust infrastructure might entail. The needs assessment occurred over a two-month period.

Twenty-one interviews were conducted in both individual and group settings, and input was gathered from 57 CBO leaders representing 60 of the 64 OEI organizations. Some leaders represented multiple CBOs, and there were four CBOs that did not participate in interviews due to non-response to interview requests. The research team met and reviewed interview transcripts to categorize responses.

Our needs assessment revealed that the CBOs that collect the most data have data collection activities at intake; each time a participant is seen; and, for some programs, after birth and when exiting the program. Data collection is done either during home visits with a participant, at group classes, or in other one-on-one settings such as public offices. Some programs collect most of the intake data at the first visit, while some collect it over a few initial visits. Our OEI data collection system largely mimics this, having forms for intake, birth, exit, and encounter with online and paper data collection options.

One of the major gaps noted by our needs assessment process was that without a system of this nature, there was excess variation in data collection approaches between CBOs. For example, initial engagement identified some community programs with advanced electronic data collection systems in place, while other CBOs worked entirely with paper records. We found that some CBOs collected only aggregated attendance data, while others maintained databases of extensive medical, behavioral, environmental, and demographic data. In addition, sufficient identifier information was often not collected, precluding post-processing and subsequent matching of data with community infrastructure.

The needs assessment revealed potential benefits of implementing a more robust system: the sponsors would receive more consistent data collection and program evaluation, while CBOs collecting minimal data would receive data collection materials and support to help with evaluation. Most CBOs had performed little evaluation work; employees lacked the time and resources, and the numbers of participants did not allow for robust analysis of outcomes. The absence of this system could be linked to implications for CBOs such as potential suboptimal allocation of resources, not targeting the individuals who may benefit most, and not providing as many referrals or interventions as a participant may need due to being unaware of risks.

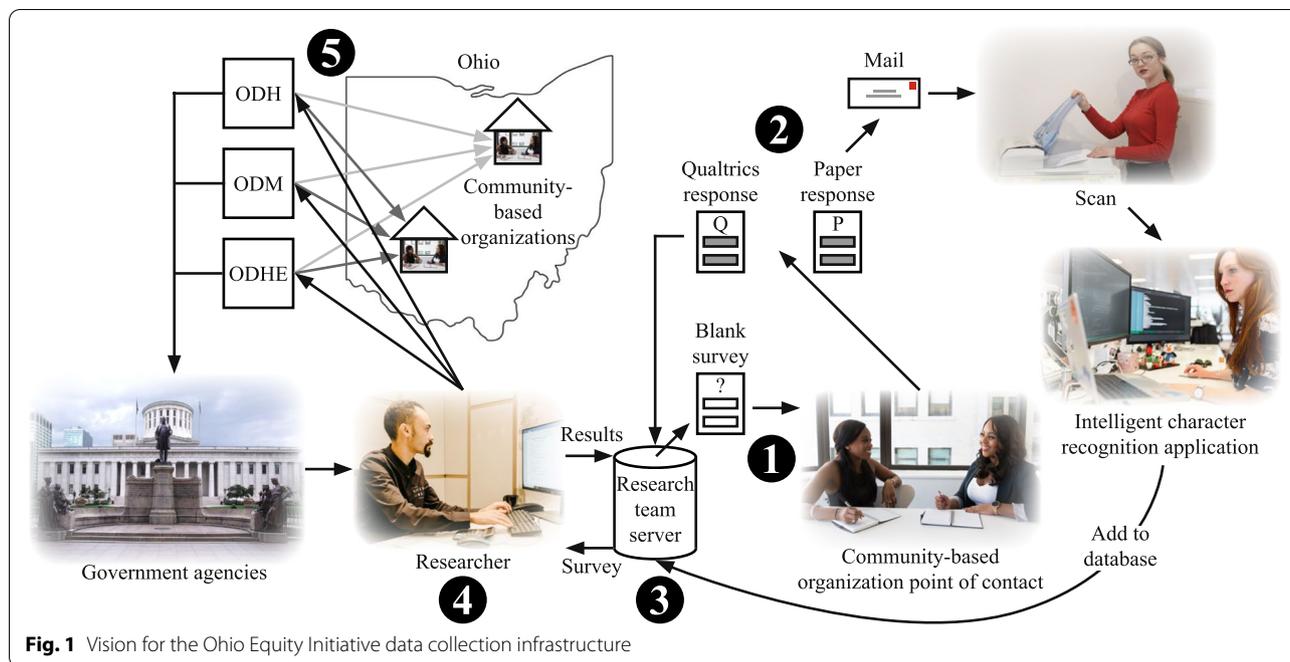
The research team developed an initial variable list with the sponsors that was revised based on the needs assessment. Our initial variable list included variables measuring demographic data, environmental and behavioral risk data, and data about care received by the mother and infant. The demographic data collected included names, birth dates, and Medicaid or Social Security identification numbers to ensure that Vital Statistics and Medicaid data are matched properly, as well as address information to map where participants reside. The needs assessment revealed a desire from the CBOs to not have an overly burdensome number of data points. Most already collected much of the data and would not be able to significantly add more data points. Many clinical variables and birth-related variables that can be found in other

materials were removed from the variable list to reduce data collection burden from participants. Many of the variables not collected are linked into the final dataset from Vital Statistics birth records, including birth weight, gestation at birth, and other health variables. The focus of the OEI data collection materials was to learn more about program participation, including its intensity and the services offered, as well as to learn more about the environmental and demographic factors that may play a role in infant and maternal health. A great deal of the data collected by the new materials are not available in any of the other data sources linked. Data collection materials were developed once the variable list was finalized. These variables would subsequently be collected and curated on a database server that concomitantly allowed for the querying of data through of a common data model. All data collection activities were approved by an Institutional Review Board at the Ohio Department of Medicaid.

Figure 1 illustrates our vision for the OEI data collection infrastructure. The figure also shows how the various data collection, curation, and reporting activities are integrated within the infrastructure system. The first step of the system involves a CBO provider collecting information about a participant using one of the previously described collection forms (i.e., intake or birth). The subsequent step is for the provider to use a data entry mechanism to report the data to the research team. Transfer of data is possible through an online data entry system, scanning and faxing forms, mailing forms, secure email of spreadsheet data or forms, or uploading to a secure online portal; in our case, we have used the Qualtrics platform, which accepts data both input into a survey or securely uploaded as a file. The third step involves the curation of a database by the research team on a server that, in step four, can be used by the team to access data via a portal. Step five illustrates how the research team can use the curated database to develop a CDM that can be used by researchers and government agencies to develop queries for reports and dashboards. The CBOs can also use this information to interact with the researchers and government agencies for decision-making purposes. Challenges with character recognition applications led to the use of manual data entry.

#### **Development of the OEI data infrastructure and common data model**

Between February 2018 and July 2018, our team engaged in the development and deployment of the OEI data collection infrastructure. Critical early milestones during this period were developing an agreed upon timeline for data reporting, piloting data collection forms with select CBOs, and refinement of these forms. The Excel sheet and paper forms were sent to select CBOs that



**Fig. 1** Vision for the Ohio Equity Initiative data collection infrastructure

volunteered during their needs assessment interviews to review and pilot test the forms. These CBOs were encouraged to provide feedback on the specific variables measured, ease of use, language and readability, and general look of the forms. After editing the forms based on this feedback, final versions were developed and the online data portal was created using the same questions. Subsequent milestones focused on piloting the data collection system between July 2018 and August 2018 across all CBOs, testing system integration, validating data points, and making system refinements based on the feedback obtained from the previously listed activities. Data collection began in September 2018 for 30 programs, and additional programs were enrolled over time after training and data use agreements were finalized. The data use agreements allowed the organizations to share individualized data with the research team that are aggregated and evaluated before being presented to the sponsors or other programs.

During the development period of the data infrastructure, our team concurrently focused on the design considerations and development of the OEI common data model. The data model architecture for our project contains data about the OEI program, its participants, and non-OEI participants. The model consists of elements from our data collection infrastructure (i.e., OEI system) linked to datasets from state databases (i.e., Ohio Vital Statistics and Medicaid Claims). The online data collection system consists of a database using Qualtrics, which offers the ability to collect data about multiple

participants and from multiple CBOs. CBO members create login information and are directed to a separate landing page for each CBO. On this landing page, participant demographic information can be added and surveys can be filled out and updated about each participant. These Qualtrics surveys follow the same format as the other data collection methods; CBOs create a record of a participant, then have the ability to complete intake, birth, exit, and encounter forms for that participant. These reports are downloaded by the researcher directly and processed and appended. CBOs can also securely send an Excel-readable spreadsheet that is imported and processed, or scanned paper forms that require an additional data entry step into a spreadsheet that is later imported. Another layer to the collection of data through the OEI system is the submission by some CBOs of data formatted for other data collection systems, in our case Care Coordination Systems (CCS) and the Ohio Comprehensive Home Visiting Integrated Data System (OCHIDS). CCS and OCHIDS contain similar data to the OEI system and required some standardizing before appending to the master dataset.

**Training to use the OEI data collection infrastructure**

CBOs were trained on how to use the data collection materials in July and August 2018. These training sessions consisted of webinars with demonstrations of how to use the Qualtrics data portal, the validated Excel spreadsheet, and the paper forms. After these demonstrations, CBOs were given a choice of how to submit data among

those options. After this choice, the CBO was provided with paper forms, the spreadsheet file, or login information to start data collection. After the first month of data collection, the research team conducted the first round of three Plan-Do-Study-Act (PDSA) cycles. The purpose of the PDSA cycles was to gather information about barriers to data collection, answer CBO questions, help CBOs improve their data quality, foster relationships with CBOs, and further improve the data collection infrastructure. The PDSA cycles consisted of emailing all CBOs surveys and inviting them to short phone calls to discuss their concerns with data collection. Phone calls were optional, but they were especially encouraged for organizations that reported many data collection challenges in the survey. As part of these conversations, CBOs were given the opportunity to change data submission preferences, provide input about portal changes desired, and receive additional training.

**Results**

**Data elements**

Data collection occurs at four points in time: enrollment, encounter, birth, and exit. Supplementary Table 1 in Additional File 1 lists the variables from the enrollment form. The intent of the enrollment form is to collect baseline demographic, risk factor, and history of prenatal care information. Supplementary Table 2 in Additional File 1 lists information collected during an encounter with a provider. The intent of the encounter form is to primarily obtain attendance and service utilization information.

Supplementary Table 3 in Additional File 1 lists information gathered about the participant and their newborn infant. The intent of the birth form is to collect risk factor information related to the baby and care utilization by the mother. Birth record data are linked to program data, enabling the researchers to know about infant and maternal health outcomes without overburdening the participant with questions that are personal or may be self-reported inaccurately. Supplementary Table 4 in Additional File 1 lists information collected from the exit form. The intent of the exit form is to collect information about care utilization by the mother and infant. Use of resources from various public programs and risks factor information are also collected. A detailed code book for

all the variables collected in the OEI data collection infrastructure is provided in the appendix.

**Data collection infrastructure**

To develop the OEI system and integrate our data model into the system, we accounted for five critical constraints based on the needs assessment interviews:

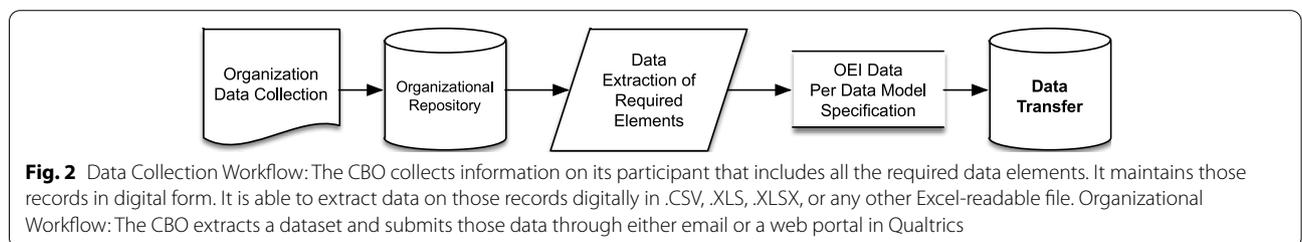
1. Data are currently gathered by CBOs for other purposes. To the extent that the CBO leverages its existing infrastructure to gather these data, we believe that data quality will be higher.
2. Some CBOs may not currently collect data. In these cases, we provided tools that facilitate data entry in a manner that reduces the likelihood of data integrity issues.
3. Some CBOs may not have the capacity to collect data in the local setting. We make no presumption of computational ability in these settings and facilitated data collection in a manner that allows the CBO to meet the reporting criteria with a minimum transaction cost.
4. Data collection is not the focus of these CBOs. Given the community nature of the interventions, the system was developed in a manner that supports ease of use.
5. Data may be collected within a CBO or at a participant’s home. The nature and timing of how CBOs are administered required a data collection approach that is robust across multiple settings.

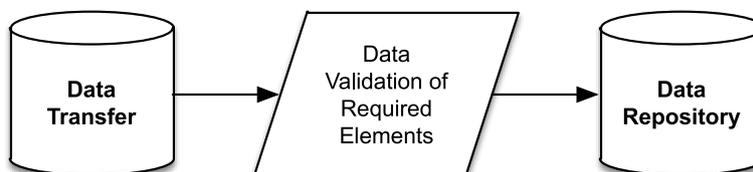
Given those constraints, we developed our multi-pronged, multimodal data collection and integration approach for the OEI system.

**Data collection and integration use cases**

The following section outlines data integration use cases and explicates how data were integrated in a single database.

Use Case 1: The CBO keeps digital records, collects all required data elements, and those elements conform to the data model specifications (Figs. 2 and 3).





**Fig. 3** Integration Workflow: The research team downloads data from email or Qualtrics. Upon the research team's receipt of the data, the files are preprocessed to ensure they conform to data standards. The data are appended to the Data Repository

Use Case 2: The CBO keeps digital records but does not collect data that conforms to the data model (Figs. 4 and 5).

Use Case 3: The CBO keeps digital records, but it does not collect all required data elements and those elements do not conform to the data model specifications.

In these cases, the CBO modifies the workflow in one of two ways:

1. Modify current systems to gather the additional required information and then conform to either Use Case 1 or 2.
2. Adopt the use of digital or paper forms and associated workflows.

For cases in which a CBO does not keep digital records, it chooses one of three options:

1. Use an online data submission system that mimics the paper forms.
2. Use a paper form and send it to our research team for processing.
3. Use a validated Excel spreadsheet created by our research team that mimics one of the forms from one of the other collection modalities.

Use Case 4A: The CBO does not currently keep digital records and chooses to use an online data submission system (Fig. 6).

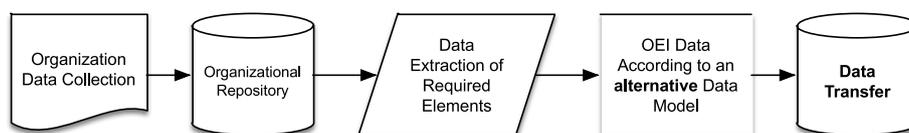
We provided an online system that allowed the CBO to enter data. The system required a computer with an internet connection. The website required site-specific authentication, and we provided credentials to the CBOs that chose this model. In cases when this approach is used, the workflow looks as follows:

Use Case 4B: The CBO does not currently keep digital records and chooses to use a paper-based data collection system (Fig. 7).

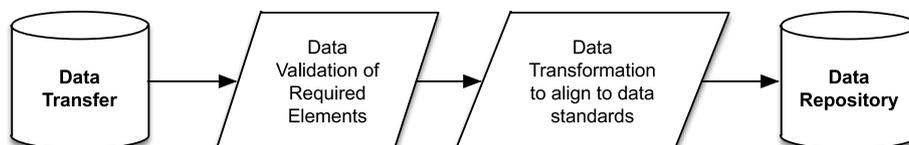
In the case when a CBO chooses to collect data on paper, we provided the CBO paper forms, and their digital equivalent, which they used to meet their reporting obligations. The forms are 8.5 by 11 in., double sided, and singly designed to be used in all cases. CBOs using paper forms could also choose to adopt the model of Use Case 4A and enter data after the fact. In these cases, our team followed the workflow illustrated below:

**Database integration and analysis**

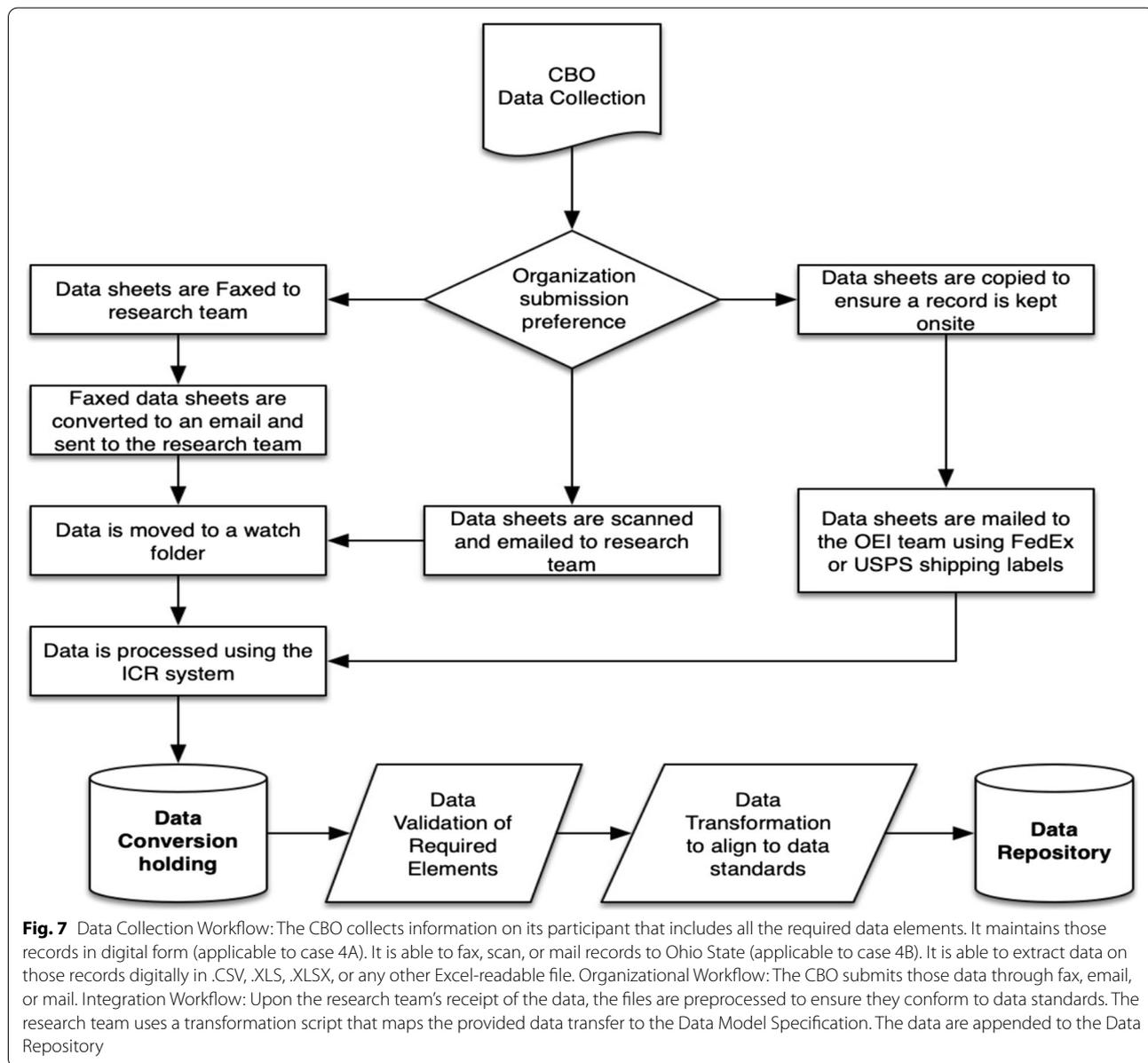
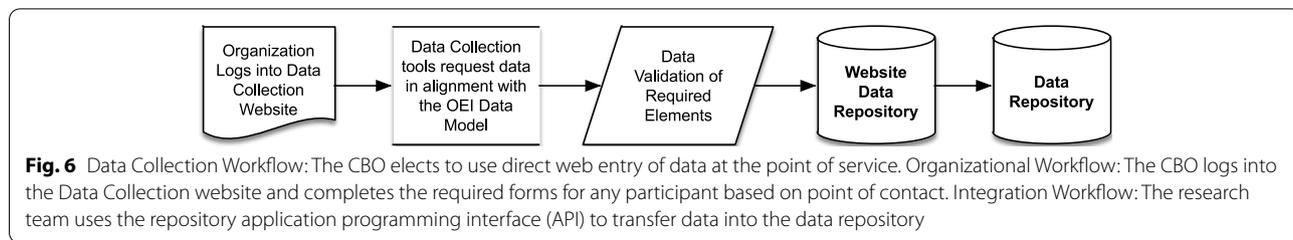
Starting in October 2018, CBOs were asked to submit their previous month's data by the 10th day of each



**Fig. 4** Data Collection Workflow: The CBO collects information on its participant that includes all the required data elements. It maintains those records in digital form. It is able to extract data on those records digitally in .CSV, .XLS, .XLSX, or any other Excel-readable file. Organizational Workflow: The CBO extracts a dataset and submits those data through either email or a web portal in Qualtrics



**Fig. 5** Integration Workflow: The research team downloads data from email or Qualtrics. Upon the research team's receipt of the data, the files are preprocessed to ensure that they conform to data standards. The research team uses a transformation script that maps the provided data transfer to the Data Model Specification. The data are appended to the Data Repository



month. Data from September 2018 were reported for 41 CBOs, and the number of CBOs reporting data increased to 62 by June 2019. From August 1, 2018, to June 1, 2019, the total number of participants reached

across all programs was 10,693, representing 10,074 different people, indicating 619 clients participated in more than one program. Our research team received data from the CBOs and engaged in extensive data

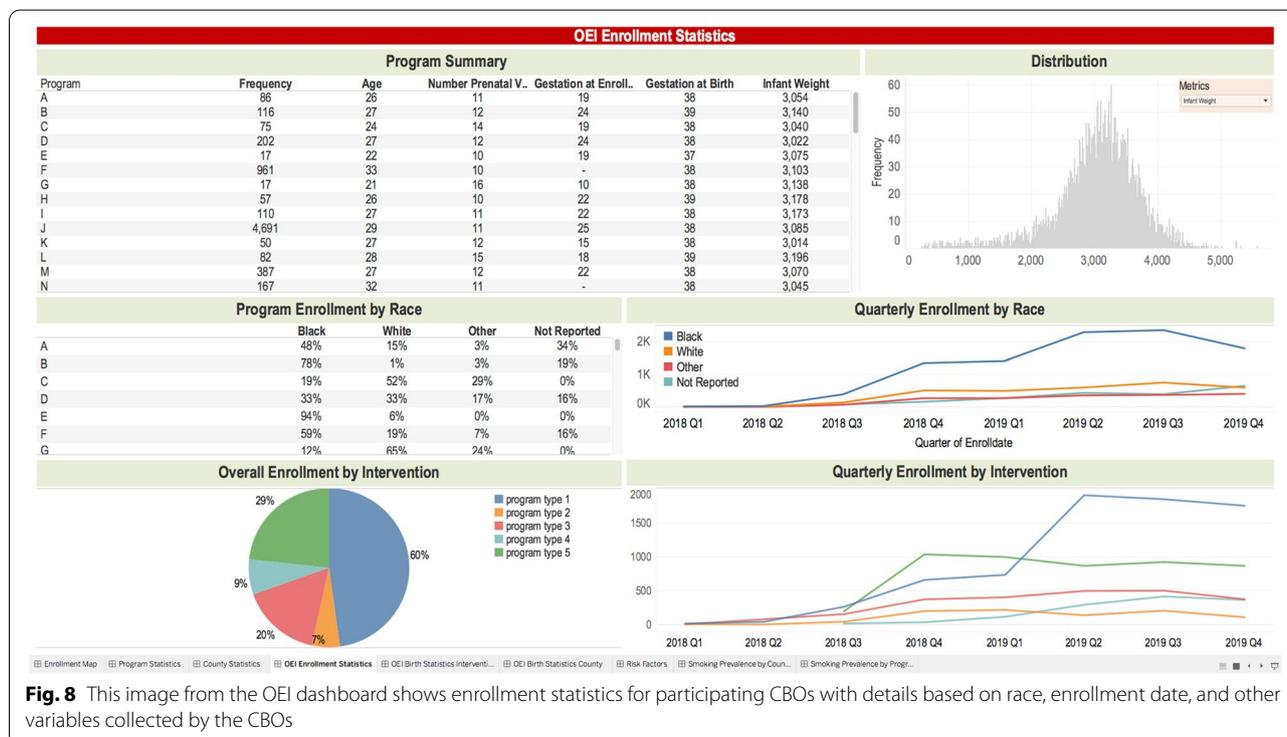
processing steps. These involved assessing data validity, accuracy, completeness, consistency, and uniformity. We also inspected the data for duplication of participant information, where we retained the most complete or most recent version of the forms. Once these steps were complete and the datasets were standardized, they were appended to one another. This appended dataset was then matched to birth records and Medicaid records using key identifiers that included the participant’s Medicaid identification number, name, date of birth, and address; non-matching records were retained for our control group. Of the 8984 total participants in the dataset used to match to birth records in May 2019, there were 2829 participants (31.5%) matched to birth records and 5798 participants (64.5%) matched to Medicaid records. Many of the participants had not yet given birth at the time of this matching. All the datasets in the database contained a one- to three-month delay after an event, except for the Medicaid claims data, which has a six-month delay.

Our research team continuously updated a data repository with the above-described data that were appended with information from the OEI system and the external state datasets. Periodic extracts from the data repository were used to generate reports for key stakeholders (e.g., annual reports to ODM) and

used in the Tableau Dashboard to visualize key metrics. The dashboards, built by the research team with regular feedback from stakeholders, display the data in the form of bar charts, pie charts, and tables of relevant information. Each page of the dashboard is able to focus on a different portion of the data, such as the enrollment numbers for the CBOs, gestation data, or risk factors. Making selections within the dashboard modifies the view and presents additional data. By offering the information in a more visual manner, rather than in tabular format, it can be put to use by policymakers and others who may not be as familiar with statistical analysis while allowing them to draw their own conclusions. An example of one of the OEI dashboards is pictured in Fig. 8. (Synthetic data are used in the illustration.)

**Data quality**

During the data processing step, our research team performed assessment of missingness of data for each variable measured. Overall and individual CBOs’ reporting of values for variables were inspected. Data quality assessment consisted of counting the number of variables each CBO collected data for, along with finding the percentage of missing data for eight key participant information variables. CBOs that had poor data quality (those that collected fewer than half of the variables or had high



**Fig. 8** This image from the OEI dashboard shows enrollment statistics for participating CBOs with details based on race, enrollment date, and other variables collected by the CBOs

**Table 1** Counts and proportions of responses by data collection mode

Data Collections Modes	Response Count (%)
CCS or CBO Excel format	26 (41%)
OEI system (Qualtrics)	12 (19%)
OCHIDS	11 (17%)
OEI system (validated Excel file)	8 (13%)
OEI system (paper form)	7 (11%)
Total	64 (100%)

missingness for participant information variables) were contacted by our research team during one of multiple Plan-Do-Study-Act (PDSA) cycles to identify systematic barriers to data reporting. During our PDSA calls with the CBOs, our research team used this information to recommend solutions to help improve data reporting. We experienced notable improvements after each PDSA cycle, and this helped improve data reporting methods over the aggregate reporting period. For example, two CBOs that reported 5 out of 8 (63%) of the contact information variables in January 2019 were reporting all eight of the variables by May 2019. The total dataset had 95.0% of participants' first names and 94.8% of their last names in January of 2019, and 99.3% of first names and 99.2% of last names by May 2019. In January, there were 31 programs sending enrollment forms, four sending birth forms, five sending exit forms, 15 sending encounter forms, and 10 sending group encounter forms. By May, 50 programs sent enrollment, 37 sent birth, 33 sent exit, 41 sent encounters, and 13 sent group encounter forms. Many of the programs that sent only minimal contact information and may not have sent any forms other than enrollment or encounter forms at first were community health worker or Other programs, which may have seen participants only once or twice in public settings. Some of these programs did not collect data at all before this project, and they were able to slowly increase the number of questions they asked over time. Many programs still do not collect every type of form and skip questions, but there were vast improvements throughout the first year of the project.

**Data analysis**

Data created for the OEI reports and dashboard were analyzed using statistical software. Descriptive statistics that include counts, means, and confidence intervals, including bootstrapped estimations of these statistics, were calculated using Stata or Tableau. We also report inferential statistics for some of the metrics.

**Preliminary response counts and rates**

The data model created by our research team has been leveraged to generate reports for key stakeholders and provide them with data visualizations using Tableau dashboards. Table 1 presents a summary of the proportion of data collection by mode for all of the OEI CBOs. Initial collection efforts suggest that CBOs are most comfortable reporting data through CCS or CBO Excel sheet. A combined 32% of CBOs reported through the Qualtrics system and validated Excel sheet modes. The paper form was the least used mode; 11% of CBOs used this mode to report their data.

**Process of establishing the OEI data collection infrastructure**

Reflecting upon the instantiation of the OEI infrastructure, we identified six core elements that represent structural components and approaches to signal to CBOs that we were working within a specified set of expectations and wanted their input and buy-in to the system (Table 2). We also identified five adaptable elements that represent components that can result in the core elements being addressed in different ways, depending on the local realities present among CBOs in regard to resources, organizational culture, and their goals. For example, a decentralized approach can be pursued if CBOs all have similar systems and expertise to standardize data before reporting; however, a more centralized approach is needed if these components are only present in some or no CBOs.

**Discussion**

Data collection continues to be one of the biggest challenges for community-based organizations [27]. The OEI data collection system our research team developed is the

**Table 2** Core and adaptable elements of instantiating OEI data collection infrastructure

Core Elements	Adaptable Elements
1. Clear end-goals to work toward	1. Development of CDM attributes and workflow
2. Needs assessment to systematically identify gaps	2. Types of variables collected
3. Co-creation/user-centered approach to the system to gain buy-in early on and avoid surprises	3. How data are processed: central or decentralized
4. Creation of a common data model	4. How training is performed
5. Multi-channel approach and adjustment to constraints	5. Reporting and analytical framework
6. Training	

foundational platform for a robust public health surveillance system to support interventions focused on infant mortality in Ohio. Even when the CBOs have systems in place, however, there are data points that are difficult to collect for less intensive program types. In our case, programs that meet occasionally in public settings or exist mainly to refer people to other services often see participants only once or twice. They also do not have time to collect extensive data during the encounters, nor has a trusting relationship been built with participants. There is more complete data collection for CenteringPregnancy and Home Visiting programs, so they may collect more complete demographic and risk factor data. Notwithstanding, CBOs that attempt to collect all or almost all of the variables in the OEI data collection system have missing data. Variables frequently missing information in the OEI data repository include data about the other biological parent, certain risk factors, and prenatal care attendance. This missing information is not systematic in nature, and such data quality concerns have motivated continuous changes to the data collection system. One approach has been to reword questions to provide clarity to items that are complex or difficult to understand.

Successes experienced in the first year of data collection included the ability to match to state records and improvements in data quality. Almost a third of participants were matched to birth records at the end of the first year. The lag in birth record reporting, the enrollment of participants early in pregnancy throughout the data collection period, and the collection of data about some participants who were not pregnant at enrollment or experienced pregnancy loss were factors that affected the ability to have all the records matched. Over time, the percentage linked to birth records will improve. In addition, over 60% were matched to Medicaid records. As there is not a requirement that participants receive Medicaid to participate in these programs, some participants may be uninsured, on private insurance, or on insurance associated with school or their parents. Finally, as more CBOs collected data and received support through PDSA cycles, the quality of data collection improved, as there were a higher number of variables collected and lower levels of missing data for participant information variables. Many CBOs who were not collecting sufficient data before this project worked with the research team to increase the number of variables collected. Allowing CBOs to submit data in multiple formats, provisioning reporting materials, and training on reporting data helped increase the number of programs submitting data and the completeness of this information. Despite some reluctance to add many new variables to data collection at needs assessment, the programs were able to comply with data collection because of the support they received

and their desire for evaluation. The effort required to change data collection systems was largely exerted at the beginning of the project; after a few months of using the system, many organizations had few concerns to report during PDSA cycles. Maintaining the system is relatively low effort, requiring periodic training for new CBO employees or if there are system updates, and the centralized data collection infrastructure is more stable than if organizations separately collected and evaluated their own data. CBOs are also now able to use better data for cost-saving QI activities or to secure future grant funding for their efforts.

The OEI data collection system has been developed to assess statewide infant mortality prevention efforts, as there are many programs attempting to effect change but little coordination among programs. In addition, there is a lack of time and resources for state agencies to create user-centered data collection infrastructure and perform extensive evaluation on their own. The collaboration between multiple teams was necessary to create a robust data collection system that worked for multiple programs. Although there are national reporting requirements for certain infant mortality prevention programs, including Nurse Family Partnership and CenteringPregnancy, there is little comparison of programs or compiling of program data. By collecting data for multiple kinds of programs in Ohio, the effects of any infant mortality prevention efforts can be assessed, and programs of similar types can be compared. Infant mortality is a rare outcome that is difficult to assess statistically, leading to difficulties assessing program success for small individual programs [9]. Compiling and harmonizing data across multiple small programs will allow for comparisons of birth outcomes both throughout the state and among program types to be more feasible, particularly for quantitative assessments where statistical power is needed. For example, if there is an infant mortality rate of 5 per 1000 in a 400 person CenteringPregnancy programmatic intervention, it may not be a statistically significant reduction compared to a general population with a rate of 6 per 1000; however, if the population is 20,000 across multiple CenteringPregnancy programs, the differences in infant mortality rates may reach statistical significance because of an increase in statistical power. The results of these comparisons will help highlight which program types and components have the most potential for improving birth outcomes, allowing funding and program decisions to be made within Ohio. These evaluations may also help other states adopt similar data collection systems and compare efforts among states.

Data collection for community programs also has the potential to lead to coordinated efforts with health

care providers. Infant mortality has multifaceted risks that cannot be addressed through the medical model alone. Many of the environmental, social, and behavioral risks that participants may experience are addressed by community programs through provision of referrals to other services, provision of child care supplies, social support, and education that is not standard for prenatal care in a medical setting [28–31]. Participating in these programs, if shown to be successful at improving maternal and child health, could be encouraged by medical providers. Coordination with community programs could improve health care quality even more if there was potential for information exchange among programs and providers, as participants may report and focus on different risks with a community health worker than they do with their doctors. Additionally, systematic collection of risk factors incorporates these critical social determinants of health to present a comprehensive picture about a patient to the provider and enhances the overall value of care delivered to the patient [32, 33]. This type of information would include aggregate information about patients' environments and the risk factors they experience.

Improvements in public health require information systems for surveillance and effective program implementation and impact, along with partnerships that build public support [4]. Because of the various risks for infant mortality across preconception and throughout pregnancy, infant mortality prevention efforts need to collaborate to achieve lasting change at a large scale, not just in Ohio [3, 4]. Statewide efforts of this nature to standardize data collection could be used as a template for enhanced nationwide infant mortality efforts. The template, moreover, can also be extended to other community services, such as those that target children with disabilities or are related to people with mental and behavioral health conditions. These findings regarding infrastructure building can be applied to almost any conditions with community organizations including mental health, substance abuse, cancer, neurological disorders, HIV/AIDS, and rare diseases to gather more consistent data for improved evaluation.

## Conclusion

There is a need to develop an ecosystem of tools to effectively gather data from program participants, community members, community-based organizations, and local and state authorities to facilitate a greater understanding of how public health interventions impact the communities they serve. There also is a need for participant-facing tools to collect information about service quality and patient-reported outcomes without having to report through a community health worker

or other intermediaries. Potential data collection tools include smartphone applications, text messaging-based data collection, and interactive voice activated applications. In addition, there is a need for organization-facing tools that report information back to organizations about their performance relative to other organizations and information about whether there are gaps in service. Currently, the Tableau dashboards are used by the program sponsors, and some of these results may be reported back to CBOs; however, these organizations do not have control over what data they see. Systems may be needed to provide regular feedback and data access for analysis for individual CBOs. Organizational engagement will be imperative moving forward to develop more efficient systems of bi-directional reporting.

The OEI data collection infrastructure our research team designed and deployed is still in its early stages of system maturity. Our research team continues to learn from its implementation and use as the system evolves over time. The system, although requiring more effort from CBOs, is already demonstrating signs of collective action at the local and state levels to better coordinate and share information on how to best utilize programmatic resources to reduce infant mortality and its associated disparities in across Ohio.

## Abbreviations

CBO: Community-Based Organization; CCS: Care Coordination Systems; OCHIDS: Ohio Comprehensive Home Visiting Integrated Data System; ODM: Ohio Department of Medicaid; OEI: Ohio Equity Institute; PDSA: Plan-Do-Study-Act.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12913-021-07393-1>.

**Additional file 1.** Variable list. This file includes the complete list of variables measured using the enrollment, birth, exit and encounter forms used as part of this study.

**Additional file 2.** OEI Data Dictionary. This file includes the complete data dictionaries for the enrollment, birth, exit, and encounter forms used as part of this study.

## Acknowledgements

The authors would like to thank all the staff members of the Ohio Colleges of Medicine Government Resource Center and the Ohio Department of Medicaid for their input during the development of our infrastructure. We are also grateful for their feedback on our manuscript. We also acknowledge the financial assistance from the Ohio Department of Medicaid under Task Order ODM202012, which supported the building of our infrastructure.

## Disclosure statement

All authors declare no commercial associations or conflicts of interest.

## Authors' contributions

NF, CS, TG, JL, and TH all contributed to the study design, drafting, and review of the manuscript. All authors read and approved the final manuscript.

## Funding

The authors were funded by the Ohio Department of Medicaid.

## Availability of data and materials

Data used in this study was reviewed and approved by the Ohio Department of Medicaid and is not publicly available.

## Declarations

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>CATALYST – The Center for the Advancement of Team Science, Analytics, and Systems Thinking, College of Medicine, The Ohio State University, Institute for Behavioral Medicine Research, 460 Medical Center Drive, Columbus, OH 43210, USA. <sup>2</sup>Department of Biomedical Informatics, College of Medicine, The Ohio State University, Institute for Behavioral Medicine Research, 460 Medical Center Drive, Columbus, OH 43210, USA. <sup>3</sup>Department of Family Medicine, College of Medicine, The Ohio State University, Institute for Behavioral Medicine Research, 460 Medical Center Drive, Columbus, OH 43210, USA.

Received: 19 July 2021 Accepted: 30 November 2021

Published online: 02 January 2022

## References

- Coughlin SS, Smith SA, Fernandez ME. Handbook of community-based participatory research. New York: Oxford University Press; 2017.
- Perry HB, Zulliger R, Rogers MM. Community health workers in low-, middle-, and high-income countries: an overview of their history, recent evolution, and current effectiveness. *Annu Rev Public Health*. 2014;35:399–421. <https://doi.org/10.1146/annurev-publhealth-032013-182354>.
- Bhutta ZA, Cabral S, Chan C, Keenan WJ. Reducing maternal, newborn, and infant mortality globally: an integrated action agenda. *Int J Gynaecol Obstet*. 2012;119(Suppl 1):S13–7. <https://doi.org/10.1016/j.ijgo.2012.04.001>.
- Barfield W, D'Angelo D, Iskander J. CDC grand grounds: public health approaches to reducing U.S. infant mortality. *MMWR Morb Mortal Wkly Rep*. 2013;62(31):625–8.
- America's Health Rankings, 2019. National Infant Mortality. <https://www.americashealthrankings.org/explore/annual/measure/IMR/state/OH>
- Ohio Department of Health, 2019. 2018 Infant Mortality Report. <https://odh.ohio.gov/wps/portal/gov/odh/know-our-programs/infant-and-fetal-mortality/reports/2018-ohio-infant-mortality-report>.
- Swoboda CM, McAlearney AS, Menser TL, Sieck CJ, Hefner JL, Walker DM, Huerta TR. Lessons from a community health worker home-visiting program to reduce infant mortality among black mothers in Ohio. *J Community Med Public Health*, 2019; CMPH-151. doi: <https://doi.org/10.29011/2577-2228/100051>
- Swoboda CM, McAlearney AS, Huerta TR. Risk factors among participants in a community health worker led infant mortality prevention home-visiting program. *J Community Prev Med*. 2018;1(1):1–7.
- Swoboda CM, Benedict JA, Hade E, McAlearney AS, Huerta TR. Effectiveness of an infant mortality prevention home-visiting program on black births in Ohio. *Public Health Nurs*. 2018;35(6):551–7. <https://doi.org/10.1111/phn.12544>.
- Huerta TR, Swoboda CM, Sieck CJ, Hefner JL, Walker DM, McAlearney AS. Ohio Infant Mortality Reduction Initiative Evaluation Report, 2010–2015. Ohio Department of Health, June 2017. Columbus, Ohio.
- Huerta TR, Swoboda CM, Fareed N, Sieck CJ, Hefner JL, Walker DM, et al. Ohio equity institute community-based organization evaluation, year one Report. Ohio Department of Medicaid, June 2019. Columbus, Ohio.
- Ohio Department of Health, 2018. Ohio Equity Institute: About the Ohio Equity Institute. <https://odh.ohio.gov/wps/portal/gov/odh/know-our-programs/maternal-child-health-program-mp-formerly-cfhs/oei/>
- Andrews C, Burleson D, Dunks K, et al. A new method in user-centered design: collaborative prototype design process (CPDP). *J Tech Writ*. 2012;42(2):123–42. <https://doi.org/10.2190/TW.42.2.c>.
- Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. *Transl Behav Med*. 2019 Nov 25;9(6):1057–64. <https://doi.org/10.1093/tbm/iby119>.
- Rice B, Boule A, Baral S, et al. Strengthening routine data systems to track the HIV epidemic and guide the response in sub-Saharan Africa. *JMIR Public Health Surveill*. 2018;4(2):e36. <https://doi.org/10.2196/publichealth.9344>.
- McIntosh S, Perez-Ramos J, Demment MM, et al. Development and implementation of culturally tailored offline mobile health surveys. *JMIR Public Health Surveill*. 2016;2(1):e28. <https://doi.org/10.2196/publichealth.5408>.
- Centers for Disease Control and Prevention. CDC's vision for public health surveillance in the 21st century. *MMWR* 2012;61(Suppl): 1–40.
- Barr PJ, Forcino RC, Thompson R, et al. Evaluating CollaborATE in a clinical setting: analysis of mode effects on scores, response rates and costs of data collection. *BMJ Open*. 2017;7(3):e014681. <https://doi.org/10.1136/bmjopen-2016-014681>.
- De Leeuw E, Hox J. Internet surveys as part of a mixed mode design. In: Das M, Ester P, Kaczmirek L, editors. Social and behavioral research and the internet: advances in applied methods and research strategies. New York: Taylor & Francis Group; 2011. p. 45–76.
- Mauz E, Hoffmann R, Houben R, et al. Mode equivalence of health indicators between data collection modes and mixed-mode survey designs in population-based health interview surveys for children and adolescents: methodological study. *J Med Internet Res*. 2018;20(3):e64. <https://doi.org/10.2196/jmir.7802>.
- Reisinger SJ, Ryan PB, O'Hara DJ, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *J Am Med Inform Assoc*. 2010;17(6):652–62. <https://doi.org/10.1136/jamia.2009.002477>.
- Bacon E, Budney G, Bondy J, et al. Developing a regional distributed data network for surveillance of chronic health conditions: the Colorado health observation regional data service. *J Public Health Manag Pract*. 2018. <https://doi.org/10.1097/PHH.0000000000000810>.
- Cho S, Mohan S, Husain SA, Natarajan K. Expanding transplant outcomes research opportunities through the use of a common data model. *Am J Transplantation*. 2018;18(6):1321–7. <https://doi.org/10.1111/ajt.14892>.
- Grasso C, Goldhammer H, Funk D, et al. Required sexual orientation and gender identity reporting by US health centers: first-year data. *Am J Public Health*. 2019;109(8):1111–8. <https://doi.org/10.2105/AJPH.2019.305130>.
- Muturi SG, Otieno G, Ngatiri G, Muhoho N. Pattern of epidemics monitoring data reporting among health facilities in Nairobi City county, Kenya. *East and Central Africa Medical Journal*. 2017;3(1):7–13.
- Kennedy PJ, Outin S, Rodriguez TR, et al. Building surveillance capacity: Lessons learned from a ten year experience. *J Infect Dis Epidemiol*, 2017; 3: 026. doi: <https://doi.org/10.23937/2474-3658/1510026>.
- Holden RJ, McDougald Scott AM, Hoonakker PL, Hundt AS, Carayon P. Data collection challenges in community settings: insights from two field studies of patients with chronic disease. *Qual Life Res*. 2015;24(5):1043–55. <https://doi.org/10.1007/s11136-014-0780-y>.
- Partridge S, Balayla J, Holcroft CA, et al. Inadequate prenatal care utilization and risks of infant mortality and poor birth outcome: a retrospective analysis of 28,729,765 U.S. deliveries over 8 years. *Am J Perinatol*. 2012;29:787–93.
- Lobel M, Cannella DL, Graham JE, et al. Pregnancy-specific stress, prenatal health behaviors, and birth outcomes. *Health Psychol*. 2008;27:604–15.
- Nagahawatte NT, Goldenberg RL. Poverty, maternal health, and adverse pregnancy outcomes. *Ann NY Acad Sci*. 2008;1136:80–5.
- Hauck FR, Tanabe KO, Moon RY. Racial and ethnic disparities in infant mortality. *Semin Perinatol*. 2011;35:209–20.

32. Amjad S, MacDonald I, Chambers T, Osornio-Vargas A, et al. Social determinants of health and adverse maternal and birth outcomes in adolescent pregnancies: a systematic review and meta-analysis. *Paediatr Perinat Epidemiol.* 2019;33(1):88–99.
33. Nijagal MA, Wissig S, Stowell C, et al. Standardized outcome measures for pregnancy and childbirth, an ICHOM proposal. *BMC Health Serv Res.* 2018;18(1):1–12.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

