3.2 All Babies Matter Simulation model for Infant Mortality or "ABM-Sim4IM"

3.2.1 Introduction and background ABM

The state of Ohio has one of the highest infant mortality (IMR) rates in the United States. Based on extrapolation of historical trends in IMR rates in Ohio, the state is on a trajectory to very likely miss the Healthy People 2020 target for infant mortality. Due to the wide-ranging impact of preterm birth and infant mortality, several state-level agencies have implemented strategies to reduce infant mortality rates. Some of these strategies have broadly been implemented, such as Ohio Perinatal Quality Collaborative, while others have been targeted based on geography (e.g. Ohio Equity Institute for select counties), race (e.g., Ohio Infant Mortality Reduction Initiative or OIMRI for African-American pregnant women), or socioeconomic status (e.g., Moms2B for low income women).

Many of these infant mortality reduction strategies have been based on the social ecological framework, which has been promoted by the Ohio Collaborative to Prevent Infant Mortality, through calls for a multi-pronged, multi-level approach towards achieving health equity in IMR rates. Historically, efforts to reduce infant mortality have occurred within disciplinary silos. An interdisciplinary approach to infant mortality, that considers the unique role of each set of risk factors (e.g. medical, behavioral, structural), has the potential to be far more effective in reducing IMR rates. Thus, a deliberate effort was made throughout the scope of this project to incorporate the expertise of stakeholders from different disciplines.

This interdisciplinary approach was prioritized in an effort to integrate the strengths of each discipline, but to also address the lack of shared worldview between disciplines. This cross-disciplinary approach to research and practice can lead to a more effective, and responsive set of infant mortality reduction strategies. The social-ecological framework provides a foundation from which the interactions of these factors across levels can be understood, and accounted for in research and interventions, thus resulting in more culturally-responsive, and effective interventions.

A systematic scoping review (provided in Appendix 1) was conducted in parallel to development of more quantitative systems models to ensure knowledge from different professional disciplines was identified (e.g. medicine, nursing, public health, social work) and included in the ABM. The social-ecological model was used as an organizing framework for the scoping review as well as for developing the systems models. This framework is unique in that it explicitly acknowledges the risk factors at multiple levels, as well as the dynamic nature and interactions between factors across multiple levels.

Systems models include system dynamics modeling, agent-based modeling and social network analysis. For this study, agent-based modeling was used to evaluate the impact of interventions for infant mortality. There is an emerging recognition by epidemiologists[22-24], public health systems researchers [25-27] and health disparities researchers [28-30] that to holistically capture complexity, dynamics, feedbacks and nonlinearities in complex health outcomes researchers need to look towards new ways of thinking—system thinking and modeling.

Systems thinking and modeling is an iterative process in which the world is viewed as dynamic rather than static. Decisions are made for the long term rather than the short term, questions are addressed which are broad rather than narrow in scope, and problems are solved by looking at the whole (holistic) rather than the parts (reductionist). The real world, in which decisions play out, is dynamic, complex, nonlinear and provides feedbacks, yet the models and frameworks currently used to make decisions do not offer any opportunity to project the consequences of those decisions in the real-world setting. Often, the decisions made are based on static, simple, linear, and unidirectional models. This problem has been realized and addressed in other areas of science including public health systems research, infectious disease modeling, climate science, and population health research.

Agent-based models (ABM), which is a methodology in systems science, provides an integrated modeling framework that simulates rule-based interactions between agents (e.g., people) and their environment (e.g., social determinants of health, healthcare utilization). In an ABM, rules of interaction between agents are specified based on real-world data, expert opinion, and model assumptions. These interactions may occur between agents of the same type (e.g., person-to-person) or different types (e.g., person-to-organization). Once an ABM has been calibrated to observed data, simulation experiments can be conducted to evaluate impact of interventions, which are representing represented by changing model parameters. In this manner, ABMs offer policy-makers with a "virtual" test bed for evaluating the effectiveness of interventions to reduce preterm birth rates (PTB) and IMR rates in Ohio.

Box 1. Common questions about ABMs.

What is an Agent-Based Model (ABM)? An agent-based model, or ABM, is a methodology in systems science that provides an integrated modeling framework to simulate rule-based interactions between agents (e.g., people) and their environment (e.g., social determinants of health, healthcare utilization).

Who sets the rules in an ABM?
Rules of interaction between agents are specified based on real-world data, expert opinion and model assumptions.

How are ABMs used by policy makers?
The ABM can be used to simulate what may happen under different policy scenarios. Updating policy scenarios occurs via modifying model parameters related to the policy, such as proportion of people affected and effectiveness of the policy.

How reliable are the results of ABMs?
ABMs are developed based on the current understanding of the system and the results of the ABM are only as reliable as the knowledge about the system and its individual components. Therefore, ABMs should be continuously updated as new insights are obtained through epidemiological studies and further research.

Brief overview, set-up and output of the ABM. The purpose of the agent-based model was to identify interventions that reduce infant mortality and preterm birth rates in the Ohio Medicaid population. The primary objectives of the model were:

- Develop richer representation of how pregnant women interact with the healthcare system during pregnancy and after birth.

- Compare the effectiveness of multiple interventions that reduce PTB (gestational age <37 weeks) and IMR (death before 1st birthday).

A schematic of the data sources used to develop the model (Fig. 4) and a schematic of the modeling process (Fig. 5) are also provided for the reader to better understand how the model works. The agent-based modeling was developed as follows:

- Variables were linked across multiple data sources to create an analytic dataset (see Fig. 4).
- Analytic dataset was used to virtually represent three types of agents and their characteristics within the simulation model.
- Simulated agent types were individuals (pregnant women, newborn baby), providers (obstetricians, pediatrician) and community health workers.
- For individual agents, pregnancy (conception to birth) and newborn (birth to 1st year of life) periods were simulated at a weekly time scale using real-world data on gestational age and other individual- and area-level characteristics.
- Individual agents accumulated health-related utility, which can be thought of as a benefit, for every visit to their healthcare provider. Additionally, non-health utility was gained as a function of access to and level of resources based on area-level characteristics (e.g., food access, employment). The higher the utility value or "points" an individual accumulates over the course of pregnancy and newborn period, then the less likely they are to experience an adverse outcome, such as preterm birth or infant death.
- The baseline model was calibrated to observed PTB and IMR rate in the Ohio Medicaid population. The calibration process involved estimating threshold values for each time step such that if the simulated utility was above the threshold value, then the pregnancy continued until the next time period. But, if the simulated utility value was below the threshold utility value for a given time period, then the individual would experience a preterm birth. Similarly, utility values for newborn baby was calculated at each time period until their 1st birthday and infant death occurred if the utility value dropped below a baseline threshold value.
- The calibrated model was used to conduct simulation experiments where the impact of interventions was evaluated for individual or combined impact of four interventions: prenatal care, employment, food access, and area-level income. Interventions were modeled by changing individual-level and/or area-level characteristics for agents and simulated PTB and IMR rates were compared against the baseline scenario, which may also be referred to as the "business-as-usual" scenario.

The outputs of ABM Sim4IM were reported as the PTB and IMR rates under simulated interventions. Results were presented for the overall Medicaid population although simulated outcomes were available by exact location, race, county and census tract. For model calibration, results were presented as maps of simulated infant mortality events for comparison with observed maps of infant mortality events and showed how well the calibrated model's output (PTB and IMR rates) matched the observed data in the Ohio Medicaid population.

3.2.2 Methods ABM

The following sections describe the methods to generate the ABM model. Section 3.2.2.1 provides the data sources used for the model. Section 3.2.2.2 provides a detailed description of the model using the OOD protocol (Overview, Design concepts, and Details), which is a common template for describe complex models, such as an agent-based model.

3.2.2.1 Data Sources

Several data sources were used to develop and calibrate the agent-based model (Fig. 4). Fig. 4 is provided as a schematic diagram of how different data sets were brought together and the flow of data throughout the model development process. Area-level data was used based on the 2014 American Community Survey and individual-level data on demographic- and health-related factors was based on a linked dataset that included variables from the Women of Reproductive Age data set (provided by Ohio Department of Medicaid), Ohio Birth/Death Certificates (provided by Ohio Department of Health), and Child Fatality Review data. In addition, spatial data on providers (e.g., Pediatricians and Obstetrician/Gynecologists) was obtained from the spatial analysis task team (Task 3).

Domain knowledge, which was not linked with the analytic dataset, was obtained through conducting a systematic scoping review to identify literature on the social determinants of health related to infant mortality (Appendix 1), incorporating key takeaways from Group Model Building session on infant mortality in Ohio, and reviewing annual report from Fetal Infant Mortality Review (FIMR) Boards across Ohio counties. The FIMR reports, which were obtained from Hamilton County and Columbus Public Health, were used to further identify which interventions to evaluate. The theoretical knowledge obtained from this combination of qualitative and quantitative data formed the basis of the model process and design concepts, which are discussed in more detail below.

[task1_figure4.tiff]

Simulated Agent Characteristics

Several data sources were used in combination to assign values for characteristics of simulated agents in the model. There were three types of agents in the model that were defined by several characteristics (Table 3).

Table 3. Description of each agent type included in the model with information on agent characteristics that were virtually created in the model and the data source for each type of characteristic.

Agent Type	Agent characteristics	Data source for agent characteristics
Woman of reproductive age and Newborns	Demographic factors	Women of Reproductive Age data set (WRA)
	Area-level socioeconomic factors	American Community Survey, 2014
	Pregnancy-related factors and health outcomes	WRA dataset, Child Fatality Review, Birth Certificate data
	Spatial location	WRA data set
Healthcare Provider	Spatial location	Provider dataset (obtained from Task 3)
	Number of Obstetricians/Gynecologists and Pediatricians	Bureau of Labor Statistic and Provider dataset (obtained from Task 3)
Community health worker (CHW)	Number of Community Health Workers	Bureau of Labor Statistics and expert opinion

Demographic factors

For women and newborn agents, the linked WRA dataset was used to define the race of the mother and identify infant deaths. Non-Hispanic White and non-Hispanic Black mothers were only included in the model because women within these categories represented a majority of pregnant women in the Ohio Medicaid population. Additional considerations for focusing on these race categories was the interest of multiple stakeholders in reducing racial inequalities in PTB and IMR rates, review of the literature that largely focused on studying health disparities in these two categories and the ongoing focus of certain State-level interventions specifically targeting African American pregnant women (e.g. OEI).

Socioeconomic factors

For women and newborn agents, several area-level socioeconomic factors were defined using their census tract of mother's place of residence at time of delivery. Values for each factor was obtained from the American Community Survey (ACS,[31]) unless otherwise indicated below. Area-level factors included:

- Unemployment (proportion of households in which head of household or another adult is/are employed on a full-time basis),
- Food access (proportion of tract with low food access and low access to transportation for stores within 0.5 miles (urban setting) and within 10 miles (rural setting),[32]),
- Area-level income (median Household Income, ACS).

These factors were further used to determine whether simulated individuals had access to resources and the resource level (see Appendix 4 for more details).

Pregnancy-/Newborn-related factors

The number of prenatal care visits, gestational week at first prenatal care visit, and gestational age (clinical) was used to model the course of pregnancy. Prenatal care visits and prenatal care were distinguished as follows. Prenatal care visit was defined as a visit to an OB/GYN during the course of pregnancy. On the other hand, prenatal care, could also be an intervention strategy for reducing preterm birth if the first prenatal care visit occurred within 12 weeks of gestation. This is what was meant by prenatal care when used as an intervention. Table S1 in Appendix 4 provides additional details on data sources for these variables.

3.2.2.2 Model Development (ODD Protocol)

The following description of model development is based on the ODD Protocol[1]. The ODD Protocol is a best-practice standard within the agent-based modeling community of practitioners. The ODD Protocol facilitates the description of complex systems models, such as the agent-based model, and is particularly helpful in reproducibility and validation studies of research results that are generated from such complex, large and multiscale models.

Overview

Purpose

As previously stated, the purpose of the model is to identify interventions that reduce infant mortality rate and preterm birth in the Ohio Medicaid population. The two primary objectives of the model were as follows:

- Develop richer representation of how pregnant women interact with the healthcare system during pregnancy and after birth and
- Compare the effectiveness of multiple interventions that reduce preterm birth (gestational age <37 weeks) and infant mortality (death before 1st birthday) rates.

In addition, the model is meant to serve the needs and incorporate objectives of multiple stakeholders at multiple levels of organization, such as the macro-level (e.g., Ohio Department of Medicaid, Ohio Departments of Health, Ohio Department of Jobs and Family Services), meso-level (e.g., primary care providers, community health agencies) and micro-level (e.g., individuals, families, and neighborhoods). The model was used to evaluate multiple interventions that could be programmatically connected in a direct or indirect manner to the decision-making and planning process at each of level of organization (i.e., macro-, meso- and micro-level). The description of the model in this report is targeted towards a non-modeling audience and, therefore, technical details of the model are found in Appendix 4.

Entities, state variables and scales

This section describes how the model was set up and how the observed data from multiple sources (Fig. 4) were utilized to determine model parameters.

The three types of agents in the model are:

i. individuals, which included women during their pregnancy and the baby during the newborn period,

ii. health care providers (OB/GYNs and pediatricians), and

iii. community health workers

The pregnancy period is defined as the length of gestation and the newborn period is start at birth and ends at newborn's 1st birthday. During the pregnancy period, a pregnant woman interacts with an OB/GYN type agent whereas during the newborn period, a baby and mother interact with a pediatrician type agent. Another type of agent is the community health worker (CHW). CHWs interact with pregnant women and newborn/mother during either period of time due to the wide range of services they offer to women and infants.

The attributes of providers/community health workers include: individual identifier, type of healthcare worker (OB/GYN, pediatrician or community health worker), spatial location (for OB/GYN and pediatrician only), and resource level. The resource level dynamically changes based on number of pregnant women or newborns serviced by the provider (i.e., number of prenatal care and pediatrician well-checkups attended), number of pathways selected (if any) by the pregnant women/newborn and whether or not additional resources are allocated by the provider to follow-up with women/newborns who missed a scheduled appointment. Currently, the model assumed dummy values for resource level because was not any data on observed metrics for resources from an organizational perspective. The attributes of individuals include: individual identifier, spatial location, race (non-Hispanic Black or non-Hispanic White), prenatal care variables (gestational week of first prenatal care visit and total number of prenatal care visits), and area-level factors. Additional attributes for newborn include gestational age at delivery, whether they were alive 1 year after birth, and, if the newborn died, then their date of death. The codebook for ABM Sim4IM lists additional details for the data used to define these attributes (Table S1).

To identify the variables to be included in the ABM (Table S1), the research team utilized stakeholder expertise, and incorporated rigorous research documenting impact or effectiveness. For the first step of this process, data from the Group Model Building sessions was used. During the GMB, experts identified the full scope of interventions that they felt were important to implement, in order to address infant mortality. As the participants represented a wide-range of professional backgrounds, these interventions spanned a number of different domains and disciplines including factors falling traditionally within the scope of maternal and child health (e.g. progesterone), as well as those from other sectors, (e.g. housing stability programs). In addition, participants also prioritized interventions they felt were most critical, and provided research documenting their effectiveness, when available ("parameter booklet" and "wall of evidence"). This was used as a foundation for the interventions included in the ABM Sim4IM. Following this GMB process, articles from the scoping review were selected in order to determine the extent to which each intervention had been evaluated, and data documenting the interventions' effectiveness in reducing preterm birth or infant mortality, were extracted for smoking cessation and breastfeeding. Neither of these values were used in the final model due to lack of data and data quality issues related to these two interventions in the IMRP datasets.

The spatial resolution of the model was continuous space (i.e., agents were geocoded to their residential address for individuals and to their place of practice for providers). The spatial extent of the model was the geographic boundary of the State of Ohio. All spatial data were converted to the Universal Transverse Mercator coordinate system (Zone 17 for Ohio). The temporal resolution or time step in the model was 1 week. A weekly temporal resolution provided us sufficient granularity for modeling prenatal care and pediatrician visits and tracking various outcomes to output from the model for each agent. These model output were aggregated over the course of pregnancy and newborn's first year of life. Aggregated results were used to evaluate impact of various interventions and present findings of the model in tables and maps. All pregnancies were simulated to complete up to 37 weeks of gestation and all newborns were simulated to complete up to 52 weeks of life after birth.

Model Process and Scheduling

The model process consists of calculating a utility value at each time step for each agent. Utility was defined as a benefit that individuals receive based on their decisions and circumstances. Health utility is defined as the benefit individual gains when making decisions that directly affects their own health (e.g., pregnancy) or the health of those. In the context of this study, health utility was assumed to increase linearly when pregnant women seek prenatal care or gain access to additional resources through interventions, such as receiving adequate prenatal care education. Non-health utility is the benefit individual gains in terms of their overall quality of life. Non-health utility is a function of an individual's area-level characteristics, such as access to resources and level of resources.

Box 2.
Definition of health-related utility
Health-related utility is the benefit an individual gain in terms of their health as a function of decisions related to their health.
How is the concept of utility used in the model?
In the context of this study, imagine that a pregnant woman is given "points" for every prenatal care visit she attends during her pregnancy. Maximum "points" per prenatal care visit are given if she follows recommended guidelines for when prenatal care visits should occur, slightly fewer "points" if she delayed the appointment past the recommended gestational week and no "points" if she missed the visits completely.

In our model, the functional form of total utility was based on the Grossman Model.[33] The Grossman Model says that total utility is the summation of health utility and non-health utility, such that health utility is cumulative and increases linearly over time. Consider the following made-up example about two women who differ in their prenatal care characteristics. In the actual agent-based model, real data on prenatal care characteristics was used from birth certificate and Medicaid Claims data.

Table 4. A simple made-up example illustrating how health-related utility was calculated for two pregnant women with different prenatal care characteristics. In this example, Woman #1 attended 13 prenatal care visits and showed up at her first prenatal care visit at 10 weeks of gestation. Woman #2 attended 10 prenatal care visit and showed up to her first prenatal care visit at 14 weeks of gestation. Assumptions: Pregnant women receive full benefit (1 "point") for each prenatal care visit. Clinical guidelines suggest gestational age at first prenatal care visit should be 12 weeks or less. If first prenatal care visit occurs after 12 weeks of gestation, then partial benefit (<1 "point") is assigned for the visit. PNC=prenatal care visit.

	Woman #1		**Woman #2**	
Gestational week	**Sequence of prenatal care visits**	**Cumulative health related utility (uniteless)**	**Sequence of prenatal care visits**	**Cumulative health related utility (uniteless)**
10	1st prenatal care visit	1		0
11		1		0
12		1		0
13		1		0
14	2nd prenatal care visit	2	1st prenatal care visit	0.75
15		2		0.75
16		2	2nd prenatal care visit	1.75
…	…	…	…	…
34	12th PNC visit	12	8th PNC visit	7.75
36		12	9th PNC visit	8.75
37	13th PNC visit	13	10th PNC visit	9.75
Total health related utility	**-**	**13 (out of possible 13 "points")**		**9.75 (out of possible 10 "points")**

The following table captures the basic objectives of each agent type at different time periods. These objectives are the basis upon which utility functions were constructed for each agent type (Table 5).

Table 5. Qualitative description of how each agent type in the model maximizes its total utility at different time periods in the simulation based on birth status.

Agent Type (examples of who they represent in real-world)	**Pregnancy period (Conception→Birth)**	**Newborn period (Birth→1st Birthday)**
Woman of reproductive age	Follow American Congress of Obstetricians and Gynecologists (ACOG) guidelines for prenatal care. Follow Pathways Model[34] for other issues (e.g., employment)	Follow American Academy of Pediatrics (AAP) guidelines[35] on pediatrician well checkups.
Provider (Pediatrician, OB/GYN)	Minimize missed prenatal care appointments	Minimize missed infant care appointments
Provider (Community healthcare worker)	Complete selected Pathways during birth.	Complete selected Pathways after birth.

The mechanics of the ABM works as follows and further described in the model schematic diagram (Fig. 5).

[task1_figure5.tiff]

During pregnancy, women gain health-related utility through attending a prenatal care visit.

After giving birth women gain health-related utility by attending pediatric well-checkup visits. Health-related utility (U_t^H) cumulatively increases over time because investing in your health is assumed to provide long-term benefits. On the other hand, utility gained through investing in other goods (U_t^Z), such as employment and food, was assumed to not accumulate over time because these goods are consumed immediately upon acquisition. For example, food resources will diminish once it is consumed by agents, but the model does not actually consider the process of resource consumption. Therefore, total utility at a time period t is a summation of these two types of utilities, which can be written mathematically as follows:

$$U_t^T = U_t^Z + \sum_t U_t^H \qquad \text{Eqn. 1}$$

The schedule of recommended prenatal care visits is as follows: monthly visits until 28 weeks of gestation, every 2-3 weeks until 36 weeks of gestation, and then weekly until delivery. Similarly, pediatrician wellness check-up visits are scheduled as follows: 48-72 hours after discharge, 3-5 days after birth, and then at 1, 2, 4, 6, 9 and 12 months. Due to weekly time scale, it was assumed that a visit to the pediatrician takes place within the first week and then follows the AAP guidelines from 1 month after birth and onwards until 12 months. During the pregnancy or newborn period, health-related utility is gained only when the prenatal care or well-checkup visit occurs, respectively. Differently, utility gained through other goods is gained at all time periods as long as the fetus is viable or the newborn baby is alive.

Over the course of the pregnancy and newborn period, individual agents, such as pregnant women and mother/newborn, may require assistance of community health workers for various reasons. As a result, individuals will select pathways for assistance (loosely based on the Pathways Model) that have specified start/end dates and quantified impacts on health or non-health related utility after completing the pathway. Mathematically, Eqn. 1 can be expanded to incorporate additional gains in utility among individuals due to prenatal care visits, where $U_t^{\cdot} = U_t^{pnc\_visit}$ in pregnancy period and pediatric well-checkup visits, where $U_t^{\cdot} = U_t^{ped\_visit}$, in newborn period. Additional health utility may be gained by additional healthy behaviors, such as receiving adequate prenatal care education by attending first prenatal care visit within 12 weeks of gestation ($U_t^{pnc\_educ}$). Eqn. 2 shows how each of these types of health-related were summed up to make up the total health utility value for each agent.

$$U_t^H = U_t^{\bullet} + U_t^{pnc\_educ} \qquad \text{Eqn. 2}$$

Further details on calculating Eqn. 2 are provided in Appendix 4.

The utility function for non-health related investments of time and efforts (U_t^Z) that provided benefit to agents consisted of several area-level characteristics (bottom of Table S1 in Appendix 4). These area-level factors were linked to individuals based on census tract of mother's residence at time of delivery. Further details for how each components of Eqn. 3 were calculated were provided in Appendix 4.

$$U_t^Z = U_t^{employment} + U_t^{areaincome} + U_t^{foodaccess} \qquad \text{Eqn. 3}$$

The set of pathways available to individual agents included: food access, employment prenatal care education. These pathways were assumed to be selected based on those women who were indicated for receiving any services related to each pathway in the observed data. It was assumed that enrolling in a pathway would result in a utility gain for the woman depending on the type of pathway. For example, a housing pathway would increase U_t^Z (Eqn. 3) whereas the prenatal care pathway would increase U_t^H (Eqn. 2). Table S2 in Appendix 4 provides detailed information on the impact of interventions on individual-level utility. Interventions during the prenatal and postnatal periods are assumed to alter state variables, namely the utility values. State variables are updated at each time step via asynchronous updating (i.e., immediately updated).

Design concepts

Design concepts, such as the objectives for each agent in the model and the observations recorded during the simulations have been described below but other design concepts, such as basic principles, emergence, and adaptation, were discussed in detail in Appendix 4.

Objective

The objective function is essentially the utility function for each agent type in the model (see Table 5, Eqns. 2 and 3). The form of the utility functions varies for each agent type based on their own objectives.

As an example, for women in the prenatal period, the objective function seeks to increase probability of a full-term live birth and in the postnatal period, the objective function seeks to increase probability of survival until 1st birthday. An adverse birth outcome may occur if a pregnant woman's utility is below a certain threshold value at each week of her pregnancy. Threshold values for pregnant women were obtained by calculating them for those women who full-term in the WRA dataset and for newborns by calculating them for those babies who survived until their 1st birthday. These threshold levels were scaled based on mother's race category and calibrated to match the observed data on preterm birth and infant mortality rates in Ohio.

Observation

The following model outcomes were calculated using simulated pregnancies and the 1st year of life of newborns (Table 6).

Table 6. Model outcomes calculated from the simulation model and relevant data sources from observed data.

Model outcome	Definition	Data source for matching with simulated model outcome
Preterm birth rate	Gestational age <37 weeks	Ohio Department of Medicaid
Infant mortality rate (per 1000 live births)	Death of a newborn before 365 days since birth	Reports and communication with Medicaid analysts

In addition, the total utility was tracked for each agent during the pregnancy and newborn period, latitude/longitude of place of residence, county of residence, race category (non-Hispanic White, non-Hispanic Black), and simulated gestational age (in weeks).

Details

Model Initialization

The model was initialized by selecting a random individual from the analytic dataset and simulating their pregnancy and the first year of life of the newborn. Initial parameter values in all state variables were set to 0 at the start of the simulation. Individuals did not move residence within the model and, therefore, their initial place of residence was assumed for the entire duration of simulated time (i.e., pregnancy period + newborn period). During the calibration process, initial values for the first gestational week when prenatal care was sought were set to be the same as that observed in the analytic dataset. Women who did not receive any prenatal care were excluded from the analytic dataset. After model calibration, this assumption was no longer viable because the pregnancy period and its duration in weeks was simulated by the model. Therefore, the week of first prenatal care visit was assumed to be 8 weeks since conception from which point onwards the model simulated subsequent prenatal care visits.

Model Calibration and Validation

The model calibration procedures involve the following steps. First, pregnancies were simulated until 37 weeks under the assumption that the following agent characteristics were the same as that in the observed data (i.e., the analytic dataset): when prenatal care started and gestational age. In these simulations, utility value was calculated based to Eqn. 1 during the pregnancy period and simulated 50,000 individuals to estimate the average utility value in each week of the pregnancy for pregnancies that were full-term (gestational age > 37 weeks). Second, newborns were simulated during the newborn period if they were born alive and their utility value was recorded according to Eqn. 1. Similar to step 1, the average utility value was recorded for each week of life until 52 weeks after birth for those babies that survived until their 1st birthday. The average utility curves for the pregnancy and newborn period were further estimated by mother's race category. These curves were estimated by simulating approximately 500,000 individuals. The number of simulated individuals can be greater than total number of individuals in the analytic dataset because the analytic dataset was used to generate the synthetic population whereas during the simulation process individuals were randomly picked from the synthetic population.

Polynomial regression models were fitted to the average utility curve for full-term births and newborns alive until their 1st birthday. Separate curves for fitted for each race category (non-Hispanic White and non-Hispanic Black) to account for the observed inequalities in PTB and IMR rates for these race categories in Ohio. Further details on estimating these regression functions are provided in Appendix 4. Using the fitted baseline utility curves for each period, two additional parameters were estimated to fit the simulated model outcomes to observed data for preterm birth rates and infant mortality rate in the Ohio Medicaid population.

In order to simulate the occurrence of preterm birth, it was assumed that if individual-level utility values (Eqn.1) were below the baseline utility value multiplied by a scalar constant (ϕ_{preg}) for a given gestational week before completing 37 weeks of gestation, then the baby was born preterm. Otherwise, the pregnancy would continue to the next gestational week. Mathematically, Eqn. 1 during the pregnancy period can be combined with Eqn. S1 to write down the condition under which preterm birth would occur as follows,

$$U_t^T = U_t^Z + \sum_t U_t^H < \phi_{preg} U_x^{pregnancy} \qquad \text{Eqn. 4}$$

Similarly, Eqn. 1 during the newborn period can be combined with Eqn. S2 to write down the condition under which infant death would occur as follows,

$$U_t^T = U_t^Z + \sum_t U_t^H < \phi_{newborn} U_x^{newborn} \qquad \text{Eqn. 5}$$

The third step in the calibration process was to fit the scalar parameter ϕ_{preg} in Eqn. 4 to match the observed preterm birth rate in Ohio's Medicaid population and then in step 4, fit the scalar parameter $\phi_{newborn}$ in Eqn. 5 to match the observed infant mortality rate in Ohio's

Medicaid population. This step-wise approach was necessary because the objectives of the agent-based model was to simultaneously match observed and simulated outcomes simultaneously and then evaluate the impact of interventions for both modeled outcomes.

Best-fit values for ϕ_{preg} and $f_{newborn}$ was estimated by running simulation using a wide range of values for each parameter (50,000 simulated individuals per parameter value) and then calculating the corresponding preterm birth or infant mortality rate. The range of potential best-fit values was narrowed with the observed data through visual inspection of the plot of parameter value and model outcome. This type of approach is also called a coarse grid approach and based on the identified narrower range of best-fit values a fine-grid search was performed for the best-fit value. In this latter step, polynomial regression models were used where the dependent variable was value of ϕ_{preg} or $f_{newborn}$ and the independent variable was the model outcomes of preterm birth rate or infant mortality rate, respectively. The best-fit value for ϕ_{preg} was estimated using the best-fit regression line and the observed outcome value (e.g., 13.9% preterm birth rate). The same approach was followed with estimate $f_{newborn}$, which is also listed in the full description of other model parameters (Table S1).

Simulation Experiments

Simulation experiments were carried out using the calibrated model to evaluate the impact of interventions. Five different interventions were evaluated based on the scoping review of the literature and input from the Group Model Building workshops, which were held as part of the System Dynamics modeling effort. The modification of model parameters to implement each intervention is listed in Appendix 4, Table S2. For each intervention, the simulations were based on 125,000 simulated pregnancies and confidence intervals were calculated using the bootstrap method (1000 samples with 50,000 pregnancies per sample). Lower and upper confidence interval values represent 250th and 975th value after sorting sample estimates of IMR and PTB rates. The percent reduction was calculated in PTB or IMR rates by subtracting the simulated rate value with interventions from the simulated rate value without intervention divided by the simulated rate value without intervention. This ratio was multiplied by 100 to estimate the percent reduction in PTB or IMR rate for each simulated intervention. Results for all simulated births in Ohio were then reported.

3.2.3 Results ABM

3.2.3.1 Overall rates

The observed data used in the model calibration process were as follows: preterm birth rate (PTB, 13.90% in observed data); infant mortality rate (IMR, 7.90 per 1,000 live births). IMR of 7.90 was based on the latest data available from Ohio Department of Medicaid (Figures 7 and 8 in link) and rates calculated in the underlying data used to develop the model. Simulated rates for preterm birth was 13.92% (95% confidence intervals: 13.67-

14.20) and for infant mortality 8.15 deaths per 1000 live births (95%CI: 7.50-8.79) matched well with the observed rates.

3.2.3.2 Spatial patterns

The spatial pattern of preterm birth and infant mortality outcomes was evaluated using the calibrated data, which showed remarkable correspondence with earlier studies that have mapped hotspots for infant mortality (Figs. 6 and 7). These plots are meant as a type of face validation of the model that may be assessed by local public health departments in the respective counties.

Figure 6 goes here (no file provided)

Figure 7 goes here

[task1_figure7.tiff]

From these maps, it appears the overall spatial patterns match reasonably well between observed and simulated pregnancy and newborn outcomes. Most discrepancies between the mapped observed and simulated outcomes are likely due to the observed data not being restricted to the Ohio Medicaid population and lack of overlap in the years for the data, which were 2007-2011 for Figure 6 (observed data) and 2008-2015 for Figure 7 (simulated data).

3.2.3.3 Impact of interventions

Among the several simulations that were possible to evaluate in the calibrated model, the following were selected based on time constraints and computational resources available. Scenarios were simulated where no interventions were implemented in order to calculate baseline PTB and IMR rates. For each intervention scenario, 125,000 pregnancies were modeled from conception until the 1st birthday of newborn. The impact of single interventions was simulated and some combinations of multiple interventions. For illustrative purposes, some interventions were also simulated under varying levels of effectiveness of the intervention.

3.2.3.4 Overall Impacts

The impact of simulated interventions on PTB and IMR rates in the simulated Ohio Medicaid population varied based on the type of intervention (Table 7 and Figures 8 and 9). Those interventions that addressed social determinants of health, such as area-level employment and area-level food access, had the largest impact on PTB and IMR rates. The prenatal care intervention, which involved ensuring that all pregnant women attended their first prenatal care visit by 12 weeks of gestation, slightly lowered PTB rate from 13.92% to 13.67%. On the other hand, addressing area-level factors, such as unemployment rate, had a relatively greater impact on PTB and IMR rates as compared to doing nothing. Also, differential impacts were observed based on model assumptions related to simulated interventions. For example, reducing area-level unemployment by

25% and 50% across all Ohio census tracts resulted in 12% and 29% reduction in IMR, respectively.

Multiple interventions were simulated simultaneously as a demonstration of how policy makers may use of the agent-based model as a decision support tool. For example, the **combined** impact of reducing unemployment rate (by 25% in all Ohio census tracts) was simulated and converting 25% of Ohio census tract from food insecure→food secure. Under this scenario, it was found that both PTB and IMR rates decreased by approximately 28% as compared to doing nothing. Comparatively, the impact of **only** reducing unemployment rates led to a 12% reduction in IMR across Ohio as compared to doing nothing. For illustrative purposes, the impact of three interventions (reducing unemployment rate, increasing neighborhood income, and increasing food access) was also simulated simultaneously, which showed a much greater impact on PTB and IMR rates as compared to implementing only two interventions. The results presented in this report were a subset of all possible intervention combinations, which is even larger when considering the capability of the agent-based model to spatially target interventions and assume various levels of reduction in area-level factors.

Table 7. Simulated impact of interventions on preterm birth (PTB) and infant mortality (IMR) rates compared to doing nothing (i.e. not simulating any interventions). All simulations were based on 125,000 simulated pregnancies and confidence intervals (95% CI) were calculated using the bootstrap method (1000 samples with 50,000 pregnancies per sample). Lower and upper confidence interval values represent 250th and 975th value after sorting sample estimates.

Intervention(s)	"What if?" scenario simulated by intervention	Simulated IMR (per 1000 live births 95% CI)	Simulated PTB (%, 95% CI)
No Intervention (baseline)	What if no interventions were implemented in Ohio?	8.15(7.50,8.79)	13.92(13.67,14.20)
Prenatal Care	What if all pregnant women attended their first prenatal care visit before 13 weeks of gestation?	Not applicable	13.67(13.41,13.93)
Area-level Employment	What if the area-level percent unemployment rate was reduced by 25% across Ohio?	7.14(6.50,7.76)	12.63(12.38,12.89)
Area-level Food Access	What if food insecurity did not exist in 25% of census tracts in Ohio?	6.02(5.47,6.59)	10.97(10.74,11.21)
Area-level Income	What if the median neighborhood household income increased by 1 standard unit (approx. $17,800) across all of Ohio?	7.49(6.83,8.14)	13.63(13.37,13.89)
Area-level Employment +	What if area-level unemployment was reduced by 25% AND 25% of census tracts in Ohio were food secure?	5.96(5.37,6.52)	10.03(9.81,10.24)

Intervention(s)	“What if?” scenario simulated by intervention	Simulated IMR (per 1000 live births 95% CI)	Simulated PTB (%, 95% CI)
Area-level Food Access			
Area-level Employment + Area-level Food Access + Area-level income	What if area-level unemployment was reduced by 25% in all of Ohio AND median neighborhood household income increased by 1 standard unit (approx. $17,800) across Ohio AND 25% of all census tracts in Ohio were food secure?	5.74(5.16,6.35)	9.86(9.63,10.07)

3.2.4 Discussion ABM

The objectives of the All Babies Matter Simulation Model for Infant Mortality Model (ABM Sim4IM) were (1) to develop a richer understanding of the interactions of pregnant women with the healthcare system in the perinatal period, and (2) to compare the effectiveness of interventions. To achieve these objectives, the Grossman Model was adapted to allow the modeler to measure the woman’s health utility based on a variety of trait- and state-specific attributes. Additionally, Bronfenbrenner’s social ecological framework[2] was used as an organizing framework through which to model the presence of specific variables at multiple levels, and the interactions between them.

The ABM-Sim4IM demonstrated the potential for a multitude of individual intervention strategies to reduce preterm birth and infant mortality. Overall, area-level strategies showed the most promise in reducing poor birth and infant outcomes although medical strategies were not fully evaluated due to lack of data and data quality issues. A discussion of each intervention in Table 7 is provided below to assist readers, who may not be familiar with agent-based models, in interpreting and understanding our findings.

Prenatal care: Under the prenatal care intervention, all pregnant women were assumed to have attended their first prenatal care visit at or less than 12 weeks of gestation. In other words, it was assumed that all pregnant women received adequate prenatal care education. The finding of a 2% reduction in PTB rate as compared to doing nothing means that seeking prenatal care early on in the pregnancy was an effective intervention. Apart from this, the model does not make any claims about how to encourage pregnant to seek early prenatal care nor does it target specific subpopulations based on previous pregnancy outcomes. Also, this finding is applicable at the aggregate level (across all of Ohio), which means that effectiveness of the prenatal care intervention may vary by geographic unit of analysis. In other words, there may be rural/urban differences due to the underlying distribution of gestational age at first prenatal care visit in the observed data. Since the agent-based model used the observed data to create virtual agents, it adequately captured and taken into account such differences in our reported findings.

Area-level Employment: Under the area-level employment intervention, the simulated model considered that unemployment rate decreased by 25% across all Ohio census tracts. The finding that PTB and IMR rates decreased by approximately 9% and 12%, respectively, should be interpreted with the following caveats.

First, it does not take into account how long it would take to reduce unemployment rates. There are many additional factors, such as how many people are actively looking for work, availability of job openings, season, and type of job (full-time/part-time), that need to be take into account regarding unemployment rates. It did not account for any of these factors except to suggest a conditional finding. IF unemployment rate was decreased by 25%, THEN the impact on PTB and IMR rates would be as reported in our findings.

Second, this intervention assumes that employment rate would decrease uniformly across Ohio, which again is an unlikely scenario but one that policy makers may want to consider as a "What if?" scenario nonetheless. This assumption also means that unemployment rate would decrease at the aggregate level (i.e., census tract) since the model did not include any information on individual-level employment status. The key point here is that regardless of how unemployment rate is lowered in a real-world setting the simulated reductions in PTB and IMR rates will only materialize when area-level unemployment rent goes down.

Lastly, unemployment rate was correlated with several other area-level factors that were considered in this report. As a result, the reported impact of the employment intervention is only applicable when all those other area-level factors are held constant.

Area-level Food Access: Under this intervention, model outcomes were simulated if food access was increased in Ohio, such that 25% more census tracts in Ohio became food secure. This value can vary based on additional information from the food security literature but for this report 25% was assumed as an illustrative example. The same caveats, which were mentioned for area-level employment, such as how long it would take to realize this intervention in real-world settings and correlation with other area-level factors, were applicable to this intervention. Additional caveats for this intervention is that it does not say anything about where food stores are located, transportation options or availability of fresh and nutritious foods. Instead, our used of the USDA Food Access Research Atlas summarizes aspects of food access and availability into a single binary variable—food insecurity (Yes/No). As a result, even though this intervention may seem to show much higher impacts on PTB and IMR rates, it is likely that such impacts will not occur very quickly because modifying the food environment is a complex problem.

Area-level Income: Under this assumption, it was simulated what would happen if the median household income across all Ohio census tracts was increased by approximately $17,800, which is 1 standard deviation (SD) unit when considering all census tracts in Ohio. Once again, the reported reductions in PTB and IMR rates should be interpreted with the caveats mentioned for other area-level factors. In addition, it did not take into account individual-level income due to lack of any data. Also, area-level income was correlated with unemployment rate. Practically, this means that policy makers may want to address only one of these area-level factors rather than both. Despite this, it was reasonable to simulate these two interventions separately because both factors are not likely to be perfectly and negatively correlated. It also did not consider how to implement this intervention except to suggest that if area-level income increased by 1 SD unit, then PTB and IMR rates may be reduced as reported in Table 7.

For multiple interventions: Among the interventions that modeled impacts of addressing multiple area-level factors, there are two important issues to keep in mind. First, it was assumed that each intervention in the scenario is applied at exactly the same time, which may not be realistic because simulated modifications in area-level factors may be realized at different time scales. Second, it did not take into account correlation between area-level factors within the same scenario. This may be problematic because the benefits gained, in terms of non-health related utility may be counted multiple times for some individuals. Further modifications to the model and how it models social determinants of health may resolve this issue. Despite this there is high confidence in the overall findings because the model was kept as simple as possible yet useful in terms of evaluating impact of interventions. Furthermore, the simulated impact of interventions under this study are only applicable to the Ohio Medicaid population and, arguably, simulated interventions may be less effective among the non-Medicaid population, which is generally healthier that the Medicaid population.

The results further indicate a need for increased cross-sector collaboration to develop effective and responsive interventions that address these critical social determinants of health. Our findings, generated by the agent-based model, provide a mechanism through which policy makers and community stakeholders can identify intervention strategies based on the unique characteristics of the populations most at risk in their communities. The additional capacity of the model to evaluate a wide combination of intervention strategies based on available resources, which may vary in space and time, has yet to be explored further. In addition, the hierarchical nature of the model that was developed leaves much room for exploring how multiple objectives of multiple stakeholders may be achieved through feedback and interaction between individual and providers.

There are several limitations of the ABM-Sim4IM. Statistically, infant mortality is a relatively rare event and it is difficult for rigorous evaluations, such as randomized controlled trials, to achieve the power necessary to measure the effect of a particular intervention. Thus, this model is limited by the accuracy and availability of data used to define the individual parameters. Also, the model does not account for repeated pregnancies that may have ended up in preterm birth, which would increase the risk of preterm birth in future pregnancies. The model also does not take into account medical risk-factors, such as gestational diabetes and preeclampsia, which may place greater risk of poor birth outcomes. In addition, agent-based models are not analytically tractable, which means that there is no mathematical formula that describes the relationship between model inputs and outputs. As a result, interaction effect estimates cannot be studied in the same way as in statistical models because there is no effect estimate for independent variables in the first place. Instead, the best way to think about interaction effects when evaluating the impact of multiple interventions, such as employment and food access, is to think about relative increases in effectiveness. In Table 7, it was found that food access alone reduced IMR by 26% yet when combined with the employment intervention IMR was reduced by 27%. This does not mean that the employment intervention is only slightly effective because when it is simulated as a single intervention IMR was reduced by 12%.

Finally, given the need to balance complexity with parsimony, the identified interventions do not include the full range of possible interventions to promote infant health. There are a number of other initiatives that have been implemented in an effort to reduce preterm birth and infant mortality in Ohio, that are not represented here. The interventions that were selected for incorporation in this model represent those prioritized by either maternal child health stakeholders or those rigorously evaluated in research literature. This was done as a deliberate effort to increase the applicability and validity of the ABM-Sim4IM model. Applicable interventions were identified by infant mortality stakeholders including researchers and maternal and child health practitioners. This was done in conjunction with a systematic review conducted by the research team. The model incorporates the most up-to-date research literature including randomized trials and meta-analyses where available, and it is based on current guidelines for perinatal and infant care, as promoted by ACOG and AAP. Thus, the development of the model, and the assumptions underlying it, are rooted in a strong research and practice foundation. Finally, the model was validated through model calibration, and the evaluation of spatial patterns. Through these validation processes a model was created that can inform the development and implementation of specific intervention strategies that can reduce the disproportionately high rates of infant morbidity and mortality, across the state of Ohio.

3.3 References

1. Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz S, Huse G, Huth A, Jepsen JU, Jørgensen C, Mooij WM, Müller B, Pe'er G, Piou C, Railsback SF, Robbins AM, Robbins MM, Rossmanith E, Rüger N, Strand E, Souissi S, Stillman RA, Vabø R, Visser U, DeAngelis DL. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*. 2006;198, 115–126.
2. Bronfenbrenner U. *Ecological Systems Theory.* 1989.