Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

MODELING AN INFORMATION-BASED COMMUNITY HEALTH INTERVENTION ON THE SOUTH SIDE OF CHICAGO

Chaitanya Kaligotla
Jonathan Ozik
Nicholson Collier
Charles M Macal

Decision and Infrastructure Sciences
Argonne National Laboratory
9700 S Cass Ave
Argonne, IL 60439, USA

Stacy Lindau
Emily Abramsohn
Elbert Huang

Pritzker School of Medicine
The University of Chicago
5841 S. Maryland Ave
Chicago, IL 60637, USA

ABSTRACT

We describe the development and application of a model that simulates the impact of CommunityRx, an information-based health intervention, on the utilization of community-based resources. The model includes a synthetic population of agents matching the sociodemographic characteristics of the South Side of Chicago, along with their activities and behaviors. We simulate the information-based intervention and model agent decision-making about using community resources to maintain health, based on a dynamic dosing of information about community resources, gained through interactions and experience. Through *in silico* experiments, our model aims to demonstrate the flow and spread of information from primary agents to others in the community, and through these dynamic interactions, the impact of an individual-level information intervention on resource utilization.

1 INTRODUCTION

CommunityRx (Lindau et al. 2016), is an information technology-based health intervention, designed to improve population health by systematically connecting people to community-based resources for wellness, disease management and caregiving. This paper describes a computational agent-based model (ABM) to investigate the impact of CommunityRx (CRx), by modeling health maintenance behaviors (use of wellness or health promoting community resources) of agents in a population, as a consequence of information diffused through network interactions.

Broadly, this research integrates clinical trial methodology with agent-based modeling to enable in-silico (computational) experimentation at scale. Our methodological contribution is to introduce the use of agent-based modeling as the integrating analytic framework and computational simulation tool to amplify the impact of, and overcome inherent limitations to, individual-level clinical trials of information-based interventions designed to improve population health.

1.1 The CommunityRx Intervention

CRx was developed in Chicago with the support of a Health Care Innovation Award (1C1CMS330997, 2012-15, ST Lindau, PI) from the U.S. Center for Medicare and Medicaid Innovation (CMMI). The study region for the CRx program covered 16 ZIP codes on Chicago's South Side, an area covering 106 square miles with a population of 1.08*M* people. CMMI funding supported a large observational and case-control

978-1-5386-6572-5/18/$31.00 ©2018 IEEE

study of CRx. The roll-out of the CRx program across the South Side of Chicago over a 3 year period between 2012 and 2015 is shown in Figure 1.

During the 2012-15 study period, each patient was given a "HealtheRx" prescription (hereafter denoted HRx), a 3-page printed list of resources personalized to the patient's gender, home address, health conditions, and preferred language (Lindau et al. 2016). Community resource data for the HRx were populated via an annual census of community service providers using direct observation, through a youth workforce development program called MAPSCorps (Makelarski et al. 2013). The HRx is the primary vector of information diffusion in the intervention; the diffusion of information into the patient's social network represented the secondary diffusion vector.

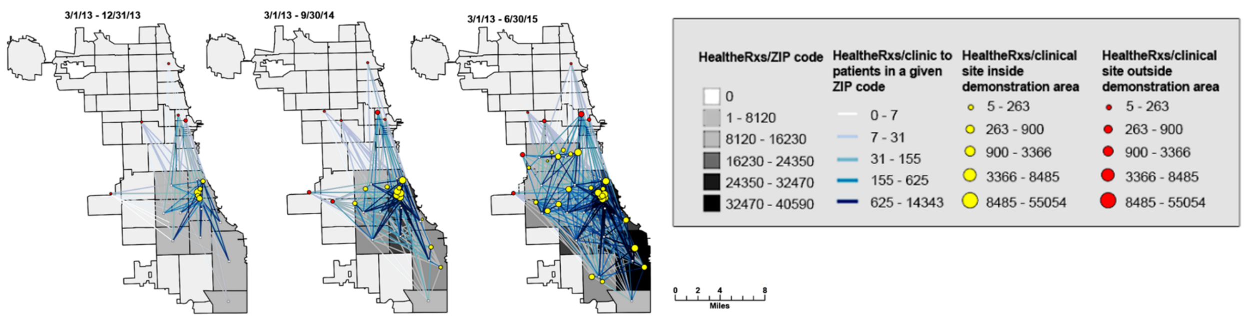

Figure 1: Rollout of CRx program between 2013-2015.

Clinical Trial: There have been two studies of CRx: i) a large scale observational study of 113,000 people of all ages, between 2012 – 15, where a case-control design was used to assess health, healthcare utilization, and cost outcomes by a third party evaluator (Lindau et al. 2016, RTI-International 2017), and ii) a prospective pragmatic clinical trial that assigned Medicaid and Medicare beneficiaries ages 45-74 to intervention and control groups by alternating calendar week assignment. The pragmatic trial was conducted in the primary care and emergency departments, and aimed to evaluate the impact of the CRx intervention on mental and physical health-related quality of life, patient self-efficacy for self-care, and health care utilization and costs.

Computational Trial: Although prospective trials are considered the gold standard in experimental medicine, an information-based intervention delivered to an individual (unlike drugs and devices) spreads non-linearly to other people who did not initially receive the intervention. Traditional linear models used to assess such an intervention are thus bound to undervalue the impact on utilization and health at the community level. Thus, we aim to run computational experiments by use of ABM to study the secondary and interaction effects of the CRx information intervention on the utilization of community resources.

1.2 Related Literature and Research Goal

CRx is an information intervention delivered to an individual at a point of healthcare service use. This information diffuses, not only from the individual who receives the intervention to other individuals in their network, but also to and through the social network of the health care professional who delivers the intervention. This diffusion pattern can be likened to an infectious disease - an individual is infected with information that spreads via dyadic interactions and diffuses through the social network of the agent, thus spreading across the population. The underlying stochastic interactions and dynamic interference (an individual agent's risk of information infection is dependent not only on their own behavior, but also on that of that agent's ego network) makes the use of ABM particularly useful (Marshall and Galea 2015).

ABM methods have been used to study transmission dynamics of infectious diseases like MRSA (Macal et al. 2014), information diffusion (Khelil et al. 2002), social network interactions (Bisset et al. 2009) and are particularly suited to study health behaviors (Gorman et al. 2006). The use of ABM to evaluate policies and health interventions has been documented – Kumar et al. (2013), for instance, used ABM to study population health impacts of influenza control policies, and Nandi et al. (2013) demonstrated the use of ABMs to study the cost-effectiveness of health interventions. Our contribution to extant literature is to introduce the methodological use of ABM to study information-based population health interventions.

Research Goal: While this paper focuses on resource utilization, our ultimate research goal is to understand the CRx intervention's impact on health, as a function of agent behavior, resource utilization, and network interconnectedness. Our aim in this paper is to use data from the two CRx studies and other sources, to inform, parameterize and develop the CRx ABM.

We describe the CRx ABM (Section 2), its implemented simulation (Section 3), and demonstrate an experiment to study the diffusion of information as a function of repeated dyadic interactions over time, and measure the intermediate outcome of resource utilization (Section 4). Future work will focus on verification and validation of the model, followed by computational experiments designed to accurately forecast the impact of the CRx intervention on resource utilization and health in the community.

2 AN AGENT-BASED MODEL OF CommunityRx INTERVENTION

In this section, we detail the conceptual model for the CRx ABM, describing our agents, objects, behaviors, activities and interactions.

2.1 Building a Synthetic Population

A synthetic population of agents matching the sociodemographic characteristics of the population on the South Side of Chicago, is built from the Synthetic Populations and Ecosystems of the World (SPEW) data set (Gallagher et al. 2018). These data are publicly available at https://www.stat.cmu.edu/~spew/. Resource and Clinic location data are obtained from the MAPSCorps dataset (place-based data publicly available at www.mapscorps.org) and the CRx database (service level data). These datasets allow us to create a synthetic population and environment with statistical equivalence to the 16 ZIP codes corresponding to the initial CRx study region.

Agents: Agents represent individual persons in a population **P**. Agents are decision makers with respect to health related behaviors, which we model in the following pages. Each agent is characterized by a set of demographic characteristics (age, gender, race,...) which are classified into a finite set of unique demographic buckets (*BucketID*), which remain static during our simulation. Each agent has a designated home, work or school location, and is assigned to a household. We only consider adults (age ≥ 16) in the 16 ZIP codes of interest, resulting in a population of 802,191 agents.

We consider two additional types of entities making up the environment in the CRx ABM – Resources and Clinics. The location of each of these entities is static throughout and is based on MAPSCorps data. Service information for each resource is obtained from the CRx database.

Resources: Resources consist of service providers in the 16 ZIP codes, which signify locations (outside of work, home and school) where agents obtain specific services and perform activities, including health related activities. Let **R** denote the set of all resource agents. A specific service available at a resource is denoted by a service code (*ServiceCode*), which in turn is mapped to specific activities (*ActivityID*) that an agent can perform. Every resource in our data set is associated with relevant service codes, which are then mapped onto corresponding activities. Considering the 16 ZIP codes in the CRx trial, we consider 4903 unique resources that feature in HRxs.

Clinics: Clinics are representative of healthcare providers, and also are the locations where an agent can potentially receive a HRx. Clinics are thus a source of information for agents in **P** about community resources **R**. Let **C** denote the set of all clinic agents.

2.2 Agent Activities and Behavior

Agent activities are developed with the use of the American Time Use Survey (ATUS) dataset (https://www.bls.gov/tus/data.htm). We call the set of all activities available to an agent as the *Master Activity List*, and denote this by *MAL*. Individual activities in the *MAL* are denoted by six digit *ActivityID*s (e.g., 10101 denotes 'sleeping' while 030302 denotes 'Obtaining Medical Care for Household Children'). Each *ActivityID* has some characteristics, a decision type $AB \in 0,1$ and an inertia score γ, which we define later. Each activity in the *MAL* is mapped to relevant service codes ($ActivityID \rightarrow ServiceCode$). By construct, each agent in $\mathbf{P}$ exclusively performs a singular activity *ActivityID* at every time-step t. An agent's set of activities over a day is called a schedule (denoted by *ScheduleID*).

Assigning Schedules to Agents: We obtain data for our agents' schedules from ATUS 2016 data, which is based on a comprehensive survey of schedules in the general population. ATUS data specifies a set of 10,493 schedules, to which we refer to as *Master Activity Schedule List* and denote by *MASL*. *MASL*s are each associated with specific socio-demographic characteristics corresponding to a unique *BucketID*. We map every ATUS schedule in *MASL* to an individual socio-demographic group *BucketID*. We define a mapping function $f : MAL \rightarrow \mathbf{R}$, where every activity in *MAL* is mapped to resources in $\mathbf{R}$ (via matching *ServiceCode*s), indicating resources where an agent can perform an activity in their schedule. In cases where an individual activity is linked to multiple service codes, we randomly pick one service code assignment based on a probability distribution obtained from from expert informant surveys completed by a diversity of researchers with expertise about the geographic region and its population (each expert completed an exercise to assign likelihoods for using different services, by age group, associated with a specific activity). Thus, each activity in an agent's schedule is mapped to some resource via f. In order to account for this variability in our model, we generate N multiple schedules based on the single *MASL*, thereby stochastically ensuring the diversity in service code assignments for a given activity. Thus our set of generated schedules, *GSL*, can be represented by $|GSL| = N \times |MASL|$. An agent is randomly assigned a *ScheduleID* for every simulated day, from all possible schedules in *GSL* with a matching *BucketID*. Each *ScheduleID* contains a corresponding *BucketID*, a list of *ActivityID*s' with the corresponding start time and end time in seconds and an associated *ServiceCode* for that activity. All schedules start at 0 seconds and end at 86400 seconds, which corresponds to a 24 hour period. This method helps us connect two separate data sets, SPEW and ATUS, and maintain a stochastic variability of the population activities and use of resources on the basis of matching demographic characteristics.

2.3 Agent Behavior: AB Decision Model

We model an agent's health maintenance behavior as a choice model, denoted as an AB Decision Model. Recall that each *ActivityID* an agent undertakes is characterized by a decision type $AB \in 0,1$. Some activities are classified as activities relating to health maintenance behaviors, our main behavioral focus, as $AB = 0$. The set of such activities was obtained through expert informant surveys completed by a diversity of researchers with expertise about the geographic region and its population. Agents with activities with $AB = 0$ face a choice in behavior - either do the activity in question, or not. Agents with $AB = 1$ activities do not face such a decision choice and simply do the activity in question at the specified time.

An agent choosing *Decision A*, implies a choice to do a health-maintenance related activity (use of wellness or health promoting community resource), whereas *Decision B* denotes the choice of not doing the specific health-maintenance related activity. We formulate this behavior on the basis of an agent's real time information dosing about community resources, their activation score (defined below), the characteristics of the activity and an agent's location. The variables which we define below represent the intersection of characteristics of agents, resources and activities. Recall that agents are exposed to information about community resources either from an HRx, or from a peer network, and this exposure to information is used to influence each agent's decision-making behavior. The dynamic nature of information is described in the next subsection.

We model the use of some wellness or health promoting community resource $j \in \mathbf{R}$, by agent $i \in \mathbf{P}$, at time t as a decision model, whose functional form is given by the following relationship between *Activation* and *Activation Threshold*: if $\alpha_i < \frac{\beta^t_{i,j}}{\gamma_j \times \delta_{i,j}}$ then *Decision A*, else *Decision B*, where $\{\alpha, \beta, \gamma, \delta\} \in (0,1)$, $j \in \mathbf{R}$, $i \in \mathbf{P}$, and $t \in 0,1,2,\ldots,T$.

Patient Activation α_i : Alpha Activation Score denotes an individual agent's intrinsic threshold for health related activities. An agent with a high (low) alpha score implies a higher (lower) level of difficulty in performing tasks, and ceteris paribus, is therefore less(more) likely to perform health related activities. This measure represents the heterogeneity of a population towards health behaviors in general. We derive this measure from Skolasky et al. (2011), matching for individual demographics.

Resource Information Score $\beta^t_{i,j}$: The Beta Resource Score denotes an agent's dynamic level of knowledge of the particular resource and its associated benefits. Beta is dynamic and changes over time based on further dosing (exposure to information about the resource). Ceteris paribus, higher (lower) the beta resource score, higher (lower) the likelihood of an agent performing an activity. At each moment in time, we assume each agent has a Beta Resource Score for up to 200 resources known to them, mimicking a cognitive memory load. The assumption of 200 resource memory load is made to represent the ability and bounded knowledge of people to keep things in their mind and make decisions. Furthermore, in future work, we could parameterize the cognitive load and conduct sensitivity analysis.

Resource Inertia γ_j : Gamma γ_j is a measure of resource inertia, which we define as the inherent difficulty associated with performing an activity at a resource. This measure is specific to each activity in *MAL* and represents resource heterogeneity. Ceteris paribus, higher(lower) the gamma (resource unattractiveness) score, lower (higher) the likelihood of an agent performing an activity.

Distance Threshold $\delta^t_{i,j}$: Delta represents the distance thresholds between an individual agent and the location for a resource/activity, accounting for a "location effect" with respect to an agent's decision. Ceteris paribus, higher (lower) the distance to an activity, lower (higher) the likelihood of an agent performing an activity. Using data from (Garibay et al. 2014) we identified distances to which surveyed participants traveled to access community resources of different types. Splitting these distances into tertiles by resource type (common services), we derive resource-specific thresholds for low-medium-high distances.

Thus, an agent chooses to perform a conscious health related activity, *Decision A*, only in cases where the patient's activation score α is less than the Activation Threshold $\frac{\beta}{\gamma \times \delta}$.

2.4 Dynamics of Information Diffusion

As described in the previous section, each agent has a dynamic resource score β for any given resource, based on information dosing, which represents the dynamic level of knowledge of the particular resource and its associated benefits. We model repeated information dosing of the Beta Score for an Agent-Resource pair, over time. Our model for the diffusion of information is a function of the source of information dosing, the propensity of information sharing, and the evolution of agent's knowledge with respect to the information.

Source of Information Dosing: Different sources for information dosing are denoted in set $X \in$ {Doctor, Nurse, PSR, Use, Peer}. Let $\varepsilon_x \in (0,1)$ be a dosing parameter for dosing source's effect on recall, for sources $x \in X$. Let ε_1 represent dosing from a physician, ε_2 represent dosing from a HRx delivered in the Emergency Department (ED) typically by a nurse, ε_3 represent dosing from a HRx from a clinic staff or patient service representative (PSR), ε_4 represent dosing from using a resource and ε_5 represent dosing from a peer in any network.

Evolution: Let λ be a decay parameter over time representing the effect of a resource receding from an agent's attention, possibly replaced with knowledge about other resources. Let $n^t_{ij} = 1$ denote the instance of information dosing for agent i about resource j, at time t by some source X. Initialization of Beta follows from function $f_\beta(I)$, which is described in detail in Section 3.2. Then the dosing dynamic is given by: $\forall i, \forall j, \beta^0_{i,j} = f_\beta(I)$, $\beta^t_{i,j} = \lambda \times \left(\beta^{t-1}_{i,j}\right)$ if $n^t_{ij} = 0$, and $\beta^t_{i,j} = \lambda \times \left(\beta^{t-1}_{i,j}\right)^{\varepsilon_x(1-\beta^{t-1}_{i,j})+\beta^{t-1}_{i,j}}$ if $n^t_{ij} = 1$, where $x \in \{1,2,\ldots,5\}$. The parameters of dosing sources (ε_x) dictate the relative values of sources of information, where we also include the effect of diminishing returns as Beta increases. In the absence of robust empirical data, we estimate these parameters based on survey data (where available) about the relative relationships across sources and perform sensitivity analysis on these parameters. An illustrative example is shown in Figure 2, highlighting the relative effects on information dosing across different sources.

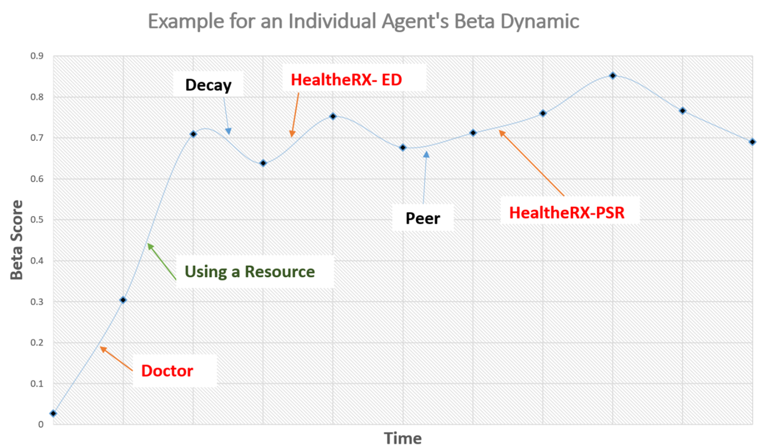

Figure 2: Illustrative example of dosing dynamics showing relative effects of information across sources.

Agent Interactions and the Propensity to Share Information: Agents interact with each other and transmit information about resources when co-located (at work or school for instance). Each activity and individual has an an associated Propensity score, representing the likelihood of an agent sharing information to other agents who are co-located at that time. This is represented as $Propensity_{(ActivityID)} \in [0,1]$ and $Propensity_{(i)} \in [0,1]$. As an agent follows their schedule, they share information about their known places with the other agents collocated at the same resource at the same time. Information sharing is dependent on the person's propensity to share information and the amenability of the activity the agent is currently performing to information sharing (e.g., propensity to share information while an agent is sleeping, is markedly different than when an agent is at work or school). Values for Propensity scores were obtained through a survey of researchers with expertise about the geographic region and its population. This setup leads to endogenously generated networks depicting the primary pathways by which information about community resources diffuses through the population. For instance, agents are the *nodes* in these networks, while *edges* represent the information exchanges via interactions that occur when agents are co-located.

3 CommunityRx SIMULATION AND EXPERIMENTS

The conceptual model of the CRx intervention is implemented as a computational ABM, to use as an experimental tool for studying dynamic behaviors and information diffusion. In this section, we describe our implementation and illustrate an experiment.

3.1 Implementation on Repast HPC

The CRx model is implemented in C++ using the Repast for High Performance Computing (Repast HPC) (Collier and North 2013) and the Chicago Social Interaction Model (chiSIM) (Collier, Ozik, and Macal 2015) toolkits. Repast HPC is an agent-based model framework for implementing agent-based models in MPI and C++ on high performance distributed-memory computing platforms. chiSIM is a framework for implementing models that simulate the mixing of a synthetic population, in this case, the inhabitants and places in the 16 ZIP codes comprising the South Side of Chicago. Each agent has a baseline set of socio-demographic characteristics (e.g., race/ethnicity, age, gender, educational attainment, income). All places are characterized by place type, including households, schools, and workplaces, and have a geographic location. In the synthetic population, agents are assigned to households, workplaces and schools (for those of school age).

In a chiSIM based model, such as the CRx model, each agent, that is, each person in the simulated population, resides in a place (a household, dormitory or retirement home/long term care facility, for example) and moves among other places such as workplaces, homes, clinics, and community resources. Each agent has a schedule that determines at what times throughout the day they occupy a particular location. Agents move between places according to their activity schedules. Once in a place, an agent mixes with other agents in some model or domain-specific way. In the case of the CRx model, agents can share information with other co-located receptive agents, who having received that information can then in turn spread that information to other agents as they move.

chiSIM itself is a generalization of a model of community associated methicillin-resistant Staphylococcus aureus (CA-MRSA) (Macal et al. 2014). The CA-MRSA model was a non-distributed model in which all the model components, including all the agents and places, ran on a single computational process. chiSIM retains and generalizes the social interaction dynamics of the CA-MRSA model and allows models implemented using chiSIM to be distributed across multiple processes. Places are created on a process and remain there. Persons move among the processes according to their activity profiles. When a person agent selects a next place to move to, the person may stay on its current process or it may have to move to another process if its next place is not on the person's current process. A load balancing algorithm can be applied to the synthetic population to create an efficient distribution of agents and places, minimizing this computationally expensive cross-process movement of persons and balancing the number of persons on each process (Collier, Ozik, and Macal 2015).

3.2 Beta Initialization

At the start of our model, we posit that an individual agent would have prior knowledge of some resources in their vicinity (to best represent a baseline reality). We implement this prior knowledge through a Beta Initialization Algorithm. First consider two radii, r_{LM} and r_{MH} representing the Low-Medium and Medium-High distance boundary an agent might consider as a resource being near, in the middle, or far. Consider three distributions representing $f(l), f(m), f(h)$ the range of number of resources that are known *a priori*. The conditional assumption is that an agent will know more resources that are closer to them, than compared to medium or far. Then an agent i randomly draws $L \in f(l), M \in f(M), \& H \in f(H)$ representing the number of *a priori* known resources, with an initial Beta score of κ.

Based on expert informant surveys completed by a diversity of researchers with expertise about the geographic region and its population, we assume $r_{LM} = 1$ mile, $r_{MH} = 3$ miles, $f(l) = U[10,100], f(m) = U[1,5], f(h) = U[1,5]$ and $\kappa = 0.02$. This implies an agent knows between 10 and 100 resources in a 1 mile radii around their home location, 1 to 5 resources in the medium (less that 3 miles) and far (more than 3 miles) distance. Each of these initial known resource has a Beta score equal to κ, while all other resources have Beta score of 0. Given our assumption of a cognitive load per agent of 200 resources, an agent at any time point only has Beta scores for the top 200 resources.

3.3 Generating Agent HRx's

The original sets of algorithms that generated a HRx based on patient demographics, home address, health conditions, and preferred language was implemented via a combination of Java code and stored procedures within an MSSQL database. We extracted the central logic and ported it to an open source MariaDB database for the purpose of generating HRx's for our CRx ABM synthetic population. The ported HRx algorithm was validated against historical data from July 1, 2016 to September 30, 2016, when the underlying data on CRx resources used by the algorithm was expected to be relatively stable. For HRxs matching over 95% of the service codes generated, indicating the the HRx was generated based on the patient's reported diagnoses and not prior clinic visits, we found that the algorithm achieved an accuracy of over 92%. That is, the individual service providers produced by the ported HRx algorithm exactly matched the actual generated HRx's over 92% of the time. We used the validated HRx algorithm to pre-generate the HRxs of all 802,191 agents in the model. This was done using parallel queries to the MariaDB database, and took approximately 35 hours on a 2.9 GHz Intel Core i7 laptop. For a larger population this would be run in an HPC cluster environment to take advantage of additional concurrency. While each agent was assigned an HRx, this did not mean that each agent would actually receive one over the course of a simulation. Rather, if the agent did go to a clinic location and was to receive an HRx, the HRx would not have to be generated dynamically. This HRx pre-generation was done to reduce unnecessary run time coupling of the CRx model code with any external libraries needed to interface with databases.

4 MODEL OUTPUT AND DISCUSSION

While the model will require further calibration and validation to reach its full potential as a trusted tool for information-based health intervention analyses, these results nonetheless, provide preliminary insights into the following two general questions: (1) How does an agent's knowledge about resources (β) evolve? and (2) How does the HRx intervention affect the knowledge about HRx resources within the model population?

4.1 Evolution of β Dynamics

In order to better understand how an agent's knowledge about resources evolves, i.e., the β dynamics within each agent, we ran a single ZIP code (60615) version of the model and tracked the dosing events that agents experienced over a simulated 8 week period. This run was executed as part of a small model input parameter sweep using the Extreme-scale Model Exploration with Swift (EMEWS) framework (Ozik et al. 2016) on the Midway2 computing cluster at the University of Chicago. The run was not distributed, running only on on a single process, and took one hour and forty-one minutes to complete. The dosing events result in different effects on the β_{ij}^t score that an agent i retains about resource j at time t, parameterized by the ε_x variable, where x is the source of the dosing $\in$ {Doctor, Nurse, PSR, Use, Peer} (see Eq. **??**). The dosing strengths ordered from weakest individual dose to strongest are {Peer, PSR, Use, Nurse, Doctor}, with associated ε_x values {0.9, 0.25, 0.2, 0.15, 0.05}.

Figure 3 shows the β dynamics for a single agent's knowledge about 15 different resources over time. We again note our prior assumption that at any one time, an agent can retain at most 200 resources in "memory," where those with lower scores are discarded. The resources are identified by unique IDs in the gray bars above each graph and each panel is additionally labeled A-O. The vertical axes are the β scores and the black dots indicate the β levels at each point in time (measured in simulated hours). The multi-colored vertical lines show the different dosing events that the agent experiences for each particular resource. For the purpose of our investigation, we collapsed the dosing categories into {HRx, Peer, Use} and colored them {blue, green, red}.

The left most panels (A-E) show that the agent receives multiple HRx's that include those resources over the course of the simulation. Panel A shows that the HRx was responsible in informing the agent about that particular resource and multiple HRx doses eventually led to the agent using the resource. Once the agent started using the resource, the β score was seen to be maintained at a high level. This pattern, an HRx nudging a person to use a resource by informing them about it, is a central goal of the HRx intervention. Panel B shows a similar pattern, with two differences. First, the agent is seen to have known about the resource at the very beginning, but through the dynamics of the β scores of other resources, this knowledge was temporarily pushed out of its memory. Second, the agent does start using the resource after learning about it through multiple HRx doses but, unlike in panel A, does so only once within the tracked period. Nonetheless, the resource continues to be peer dosed and its β score does not decay substantially. Panels

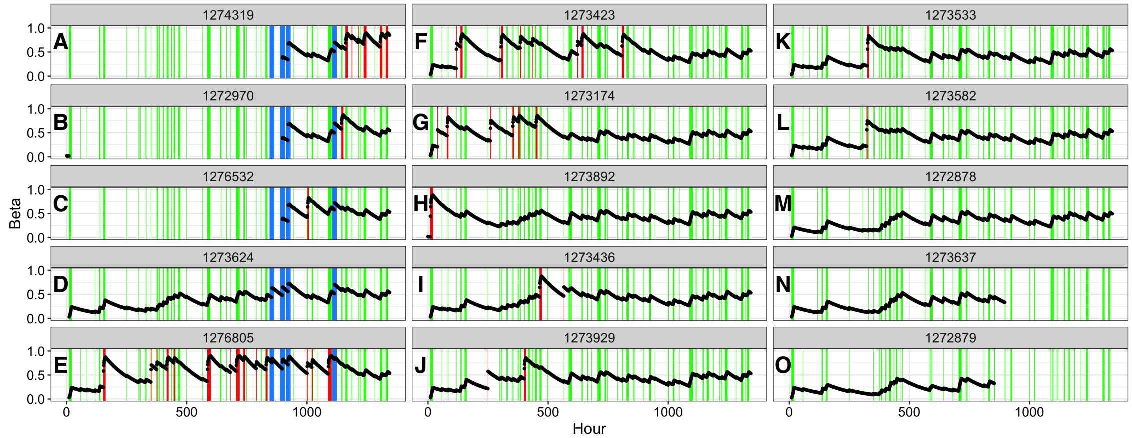

Figure 3: β dynamics for a single agent's knowledge about 15 different resources over time for a single ZIP code (60615) version of the model over an 8 simulated week period. HRx (blue), peer (green) and usage (red) dosing events are indicated by vertical lines.

C-E show additional patterns, including early usage after HRx dosing (C), a resource known throughout the tracked period, buoyed only through peer dosing (D), and use preceding HRx dosing (E).

Panels F-L show patterns involving only peer and resource use dosing, where we can observe the relative strength of resource usage with respect to peer dosing. We also observe (panel M) that peer dosing alone, when sufficiently repeated, is enough to keep the knowledge about a resource in an agent's memory. Panels N and O show the scenarios where long known resources are eventually pushed out of memory when resources such as those in panels A-C are pushed in.

4.2 Information Diffusion of HRx Resources on the South Side of Chicago

From the individual agent's perspective, next we turned to analyzing aggregate model outputs from running the CRx ABM on all 16 target ZIP codes (802,191 agents) with the goal of investigating how the HRx intervention affects the knowledge about HRx resources within the full model population. HRx resources, as opposed to the more general CRx resources that exist in the model, are those that are recommended on the pre-generated HRxs for our synthetic population. We ran two scenarios, one where HRx dosing was completely turned off (no-HRx) and another where it was administered with probability 0.25 per clinic visit (HRx). The clinics that could dispense HRxs were limited to the 25 clinics which have historically issued them. These scenarios were run for one simulated week to include both weekday and weekend activities. The scenario runs were executed as part of a model input parameter sweep using the EMEWS framework (Ozik et al. 2016) on the The Laboratory Computing Resource Center's Bebop cluster at Argonne National Laboratory and took one day, thirteen hours and thirty-four minutes to complete, running on a single process. (Future work will distribute the model across hundreds of processes, resulting in a substantial performance gain.) We found that, compared to the no-HRx scenario, the population in the HRx scenario showed 18% greater knowledge about HRx resources, measured by the total number of HRx resource β scores tracked by the synthetic population (64 million vs. 54 million). Table 1 shows how the individual HRx resources were distributed with respect to the fraction of increased knowledge about them between the two scenarios, defined by $\left(N_j^{HRx} - N_j^{no-HRx}\right)/N_j^{no-HRx}$, where N_j^X is the number of j resource β scores tracked by the population in scenario X. We observe that some resources see a significant increase in exposure.

For instance, agents referred to one large grocery store located on the South Side of Chicago saw a fractional change of 2.95, meaning there was a 295% increase in the amount of people who knew about this particular business. A large grocery store such as this would be indicated on HRxs for a number of

Table 1: Distribution of service providers based on the fractional increase of agents with knowledge about them (Beta scores) between the no-HRx and HRx scenarios.

Fraction Increase Range	Number of Service Providers
[-1.0, -0.5)	7
[-0.5, -0.2)	104
[-0.2, 0.0)	511
[0.0, 0.2)	362
[0.2, 0.5)	153
[0.5, 1.0)	100
[1.0, 2.0)	57
[2.0, 5.0)	49
[5.0, 20)	11

conditions, including diabetes, hypertension or wellness and may provide a number of services such as a place to buy fresh fruits and vegetables or healthy eating classes.

5 DISCUSSION AND FUTURE WORK

We have demonstrated two key elements of our CRx ABM. The first is the change in an agent's knowledge about resources over time, as a function of their daily activities and interactions with other agents and resources. The second is the effect of the primary vector of information diffusion, the HRx, on resource utilization. We see an 18% average increase in knowledge of HRx resources in the 16 ZIP codes over a one week simulated period.

These elements provide a key insight into the CRx ABM to study the multi-level impact of the CRx intervention. Primary dosing vectors along with secondary network interactions enable an increased knowledge about resources, which in turn can affect an individual agent's future choice of health maintenance behaviors. As agents increasingly choose to partake in health maintenance behaviors, there is an increase in utilization of resources on average. The CRx ABM allows us to focus on an individual agent or an individual resource, or different environmental parameters, allowing us to run computational experiments to robustly evaluate the CRx intervention. The strength of our modeling approach is in the use of diverse data sources to inform and build a representative dynamic interacting system in which we can perform otherwise inefficient or prohibitive computational "what-if" trials that complement more traditional approaches, such as clinical trials. Clinical trial data were used to inform details of the model. Limitations of our study include the sensitivity of results to input data and model assumptions, for e.g., the utilization of a resource being driven by Beta scores that indicate exposure. It might be informative to additionally include a measure that also incorporates sentiment toward a resource, which would reflect an agent's like or dislike.

The CRx ABM is in its early stages and will require additional calibration and validation studies before extending it to evaluate the intervention's effect on health in addition to resource utilization. We also plan to conduct extensive robustness checks and sensitivity analysis. Given the computational burden of running individual models, we will utilize the Extreme-scale Model Exploration with Swift (EMEWS) framework (Ozik et al. 2016) to run complex *model exploration* workflows on HPC resources. These will include dynamic workflows that implement heuristic algorithms such Genetic Algorithms (Holland 1992) or Active Learning (Settles 2012) to efficiently characterize the CRx ABM input parameter space and to run intervention analyses.

In the context of our larger research agenda, this paper demonstrates the use of the CRx ABM to begin modeling the impact of the HRx intervention on information diffusion and resource utilization. In our future work, we will extend our investigations to include the impacts of agent information acquisition to health maintenance behaviors and to overall community health trends.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health / National Institute on Aging [grant numbers R01AG047869 and 3R01AG047869-03S1, ST Lindau, PI]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was completed in part with resources provided by the Research Computing Center at the University of Chicago (the Midway2 cluster) and the Laboratory Computing Resource Center at Argonne National Laboratory (the Bebop cluster). Dr. Huang is supported by grants K24DK105340 and P30DK092949.

REFERENCES

Bisset, K. R., X. Feng, M. Marathe, and S. Yardi. 2009. "Modeling Interaction Between Individuals, Social Networks And Public Policy To Support Public Health Epidemiology". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti et al., 2020–2031. Piscataway, New Jersey: IEEE.

Collier, N., and M. North. 2013. "Parallel Agent-based Simulation With Repast For High Performance Computing". *Simulation* 89(10):1215–1235.

Collier, N. T., J. Ozik, and C. M. Macal. 2015. "Large-Scale Agent-Based Modeling with Repast HPC: A Case Study in Parallelizing an Agent-Based Model". In *Proceedings of the Euro-Par 2015: Parallel Processing Workshops - Euro-Par 2015 International Workshops, Vienna, Austria, August 24-25, 2015, Revised Selected Papers*, 454–465.

Gallagher, S., L. F. Richardson, S. L. Ventura, and W. F. Eddy. 2018. "SPEW: Synthetic Populations and Ecosystems of the World". *Journal of Computational and Graphical Statistics*:1–12. https://doi.org/10.1080/10618600.2018.1442342.

Garibay, L., J. Makelarski, and S. Lindau. 2014. "South Side Population and Vitality Studies Health Study Report". Technical report, South Side Health and Vitality Studies, The University of Chicago. https://sciencelife.uchospitals.edu/wp-content/uploads/2014/07/pop-health-study-final-report.pdf.

Gorman, D. M., J. Mezic, I. Mezic, and P. J. Gruenewald. 2006. "Agent-Based Modeling of Drinking Behavior: A Preliminary Model and Potential Applications to Theory and Practice". *American Journal of Public Health* 96(11):2055–2060. PMID: 17018835.

Holland, J. H. 1992, April. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, Mass: A Bradford Book.

Khelil, A., C. Becker, J. Tian, and K. Rothermel. 2002. "An Epidemic Model for Information Diffusion in MANETs". In *Proceedings of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '02, 54–60. New York, NY, USA: ACM.

Kumar, S., J. J. Grefenstette, D. Galloway, S. M. Albert, and D. S. Burke. 2013. "Policies to reduce influenza in the workplace: impact assessments using an agent-based model". *American journal of public health* 103(8):1406–1411.

Lindau, S. T., J. Makelarski, E. Abramsohn, D. G. Beiser, V. Escamilla, J. Jerome, D. Johnson, A. N. Kho, K. K. Lee, T. Long et al. 2016. "CommunityRx: a population health improvement innovation that connects clinics to communities". *Health Affairs* 35(11):2020–2029.

Macal, C. M., M. J. North, N. Collier, V. M. Dukic, D. T. Wegener, M. Z. David, R. S. Daum, P. Schumm, J. A. Evans, J. R. Wilder et al. 2014. "Modeling the transmission of community-associated methicillin-resistant Staphylococcus aureus: a dynamic agent-based simulation". *Journal of translational medicine* 12(1):124.

Makelarski, J. A., S. T. Lindau, V. D. Fabbre, C. M. Grogan, E. M. Sadhu, J. C. Silverstein, T. T. T. Tran, M. Van Haitsma, E. Whitaker, and D. Johnson. 2013. "Are your asset data as good as you think? Conducting a comprehensive census of built assets to improve urban population health". *Journal of Urban Health* 90(4):586–601.

Marshall, B. D. L., and S. Galea. 2015. "Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology". *American Journal of Epidemiology* 181(2):92–99.

Nandi, A., I. Megiddo, D. Prabhakaran, and R. Laxminarayan. 2013. "An agent-based simulation modelling approach to extended cost-effectiveness analysis of health interventions". *The Lancet* 381:S96.

Ozik, J., N. T. Collier, J. M. Wozniak, and C. Spagnuolo. 2016. "From desktop to Large-Scale Model Exploration with Swift/T". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder et al., 206–220. Piscataway, New Jersey: IEEE.

RTI-International 2017. "Evaluation of the Health Care Innovation Awards: Community Resource Planning, Prevention, and Monitoring. Third Annual Report.". Technical report. https://downloads.cms.gov/files/cmmi/hcia-communityrppm-thirdannualrpt.pdf.

Settles, B. 2012, June. "Active Learning". *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6(1):1–114.

Skolasky, R. L., A. F. Green, D. Scharfstein, C. Boult, L. Reider, and S. T. Wegener. 2011. "Psychometric properties of the patient activation measure among multimorbid older adults". *Health Services Research* 46(2):457–478.

AUTHOR BIOGRAPHIES

CHAITANYA KALIGOTLA Ph.D., holds postdoctoral appointments at the Decision and Infrastructure Sciences Division at Argonne National Laboratory and the Consortium for Advanced Science and Engineering (CASE) at the University of Chicago. CK's main research focus includes modeling human and social behavior related to the spread of information, opinions and rumors. His email address is ckaligotla@anl.gov.

JONATHAN OZIK Ph.D., is a Computational Scientist at Argonne National Laboratory and Senior Scientist in CASE at the University of Chicago. He leads the Repast agent-based modeling toolkit (repast.github.io) and the EMEWS framework (emews.org) projects. His research focuses on agent-based modeling and large-scale model exploration. His email address is jozik@anl.gov.

NICHOLSON COLLIER Ph.D., is Software Engineer in the Decision and Infrastructure Sciences Division at Argonne National Laboratory and Research Staff in CASE at the University of Chicago. He is the lead developer of the Repast and Repast HPC (repast.github.io) agent-based modeling toolkits and a core developer of the EMEWS (emews.org) framework. His email address is ncollier@anl.gov.

CHARLES M MACAL Ph.D., PE, is a Distinguished Fellow and Group Leader in the Decision and Infrastructure Sciences Division at Argonne National Laboratory. He is a principal investigator for the development of the widely used Repast agent-based modeling toolkit. His email address is macal@anl.gov.

EMILY ABRAMSOHN M.P.H, is the Director of Research and Data Governance for the Lindau Laboratory at the University of Chicago. Her research interests lie in optimizing equitable access to resources for people living in high-poverty communities and women's health. Her email address is eabramsohn@bsd.uchicago.edu.

STACY TESSLER LINDAU M.D and M.A.P.P, is a tenured Professor in the Departments of Obstetrics and Gynecology and Medicine - Geriatrics and Palliative Medicine at the University of Chicago and a practicing physician. She is the principal investigator of CommunityRx. Dr. Lindau is a Health Innovators Fellow at the Aspen Institute. Her email address is slindau@bsd.uchicago.edu.

ELBERT HUANG M.D. and M.P.H, is a Professor of Medicine, Director of the Center for Chronic Disease Research and Policy, and Associate Director of the Chicago Center for Diabetes Translation Research at the University of Chicago. Dr. Huang is a primary care doctor who studies clinical and health care policy issues at the intersection of diabetes, aging, and health economics. His email address is ehuang@medicine.bsd.uchicago.edu.