DEVELOPMENT AND APPLICATION OF AGENT-BASED DISEASE SPREAD SIMULATION MODEL: THE CASE OF SUWON, KOREA

Hyun-Jin Jung  Yo-Han Kim
Gi-Sun Jung  Yun-Bae Kim
Young Kim
Nokhaiz Tariq Khan

Department of Industrial Engineering  Department of Systems Management Engineering
Sungkyunkwan University  Sungkyunkwan University
Seo-bu Street 2066  Seo-bu Street 2066
Suwon, 16419, S. KOREA  Suwon, 16419, S. KOREA

Jin-Soo Park

Department of Management Information Science
Yongin University
134, Yongin-daehak-ro
Cheoin, Yongin, S. KOREA

ABSTRACT

The spread of diseases such as the Middle East Respiratory Syndrome (MERS) and avian influenza inflicts a significant socioeconomic problem, and highlights the need for a systematic analysis of disease spread patterns. In Korea, however, most domestic research utilize equation based approaches that treat the entire country as a single entity, providing limited applicability. In this study, we propose an agent-based disease diffusion model that reflects not only the nature of the disease, but also the structural and statistical characteristics of each region’s population. Based on the 2010 Census data, we developed a synthetic population model of Suwon city in Korea. The spread of disease and various response strategies were analyzed based on the contact network based on the socioeconomic activities of residents. The proposed model is expected to play an important role in formulation of effective disease-related policies, reflecting the mobility and socio-economic structure of today’s urban society.

1 INTRODUCTION

Diseases are transmitted through various pathogens. Among them, infectious diseases classified as highly infectious diseases have caused a great deal of harm to individuals and countries from past to present. In Korea, recent outbreaks of avian influenza and MERS have raised public’s awareness and call for measures against such infectious diseases.

Numerous studies have been conducted to predict the spread of infectious diseases worldwide. These studies can be largely classified into parametric models (Epstein et al. 2007; Tuite et al. 2011; Longini 1988; Mollison 1977; Colizza et al. 2007; Rvachev and Longini 1985) and simulation models (Smith et al. 1995; Del Valle et al. 2006; Chao et al. 2010; Focks et al. 1995; Khelil, Becker, Tian, and Rothermel 2002; Frias-Martinez, Williamson, and Frias-Martinez 2011). Both methods involve development of a model that well represents the characteristics of the population and disease to simulate the spread behavior of an infectious disease. For more detailed analysis of the spread of infectious disease over a city or country, it is
Kim, Jung, Kim, Jung, Kim, Khan, and Park

essential to construct a population model that well captures the individual life patterns of the population of interest. Since lifestyles and daily patterns are different for each country, it is important to first analyze the demographic characteristics of the population in given country and construct a model that reflects the characteristics of the population.

Studies in Korea have been limited to parametric models of the country as a single unit (Department of Preventive Medicine and Public Health 2010; Korea Centers for Disease Control and Prevention 2005). Parametric models provide insight limited to country as a whole, and provide limited insight on the propagation patterns, as well as effects of various reactive and proactive strategies. Research related to development of population modeling have been carried out in other countries (Buckeridge et al. 2004; Li, Chen, and Lin 2003; Chao 1987). But because such population models are custom made to reflect the characteristics of each country, they are not suitable for use in domestic Korean studies. Therefore, it is necessary to analyze population census data to reflect the characteristics of domestic population and to construct a population model that reflects the characteristics of Korea.

This paper describes a study on the construction of demographic model for each region in Korea using stratified data sampling and optimization methodologies developed by Hwang et al. (2014), applied to analysis of the spread of MERS-CoV (Middle East respiratory syndrome coronavirus) in the city of Suwon, South Korea. We refer to the 2010 Korean census statistics and 2% sample data to extract the characteristics of sub-regions within Suwon as well as their residing individuals. For this, it is necessary to reflect various factors such as population size by age and gender, counts of household types, and inter and intra economic/academic activities of regions. Based on this, we constructed a disease diffusion simulation model and experimented on diffusion rate analysis and diffusion suppression strategies. We compare the simulation results of various reactive and proactive strategies against the default scenario.

This paper is composed as follows. Chapter 2 describes each model included in the disease diffusion simulation model, and Chapter 3 describes the experiment through the simulation model. Chapter 4 analyzes and evaluates experimental results and epidemic response scenarios, and Chapter 5 describes the conclusions of the study.

2 METHODS

The structure of the infectious disease spread simulation model consists of a generated population that closely follows the statistical characteristics of the given city, mobility model that circulates individuals between daytime and nighttime activity regions, and a transmission model that shapes the characteristics of the disease and propagation mechanics. The population model is constructed using a modification of stratified sampling and supply-and-demand linear programming detailed in Hwang et al. (2014). The diffusion model is a Susceptible-Exposed-Infectious-Recovered (SEIR) model (Smith, Wang, and Li 2001; d'Onofrio 2002; Earn et al. 2000).

![Figure 1: Overview of the simulation model.](image-url)
2.1 Population model

Two distinct types of census information is utilized in formation of population model: the Korean Population and Housing Census (Census data) and A-type data (2% sample of entire population), both from the National Statistical Office. The population generated mimics the characteristics of the individuals in given city, which enables simulation that reflects real world situation as closely as possible. It is important to first note the categorization of residents into four distinct age groups: the 0-6 years old preschool children, 7-18 years old school age population, 19-64 years old economically active population, and 65 year old or older retired population. This is also one of the main factors affecting the spread of disease, since the time and place of activity and patterns of activity are determined largely by age of the individual. The census data contains the population statistics for each of the 39 towns in Suwon, by the aforementioned age category and gender.

The generation of population occurs in two steps. First, we randomly generate individuals with specified gender and age within each of the 39 town in Suwon, while meeting the statistics found in census data. We then derive the statistics of family type from the 2% sample data, which is defined by the number and age category of members in a family. This is possible as the 2% sample data is sampled as family groups, and each individual is labeled with a unique family id. Families are generated to match the ratio of family types in each town. Individuals generated from the first step are assigned to families generated in step two.

2.2 Mobility model

Once the population is generated, we must specify the activity space of individuals by time of day. The time-based activity space has a significant impact on the spread of the epidemic. Population activity time periods are classified into daytime and nighttime, and the activity space for each time period is determined largely by age. The designated activity space may be the same town as the residing town, or it may be in another town or region. In this study, the daytime activity space of 3~6 year olds is designated as nursery school, 7-12 year olds as elementary school, 13-15 year olds as middle school, and 16-18 year olds as high school. The 19 to 64 year olds are categorized as working individuals, and individuals over 65 years old are considered as retired and stay at home during daytime. During the nighttime, the activity space is designated to be each individuals’ residing home, which is already designated in the population generation step in section 2.1.

2.2.1 School Allocation

Allocation of daytime activity space can be solved in two steps, utilizing the supply-and-demand linear programming approach. In the first step, we must solve how many individuals from town A travels to town B for daytime activity, for all combinations of 39 towns. The supply side constraint is defined by number of residents of each town by each activity related age group, while demand side is defined by the regional statistics that outlines the number of school and attending students of each town. The domestic law dictates that preschool to middle school students must attend schools as close to residing town as possible, and thus we define our objective to minimize commuting distance of students while keeping the supply and demand constraints. If a student travels from town A to town B for school, the commuting distance is defined to be the Euclidean distance between the centroid of each town.

In the second step, we generate 182 preschools, 96 primary schools, and 56 middle schools with varying capacity that reflects the regional school statistics derived from the Korea Education and Research Information Service (schoolinfo.go.kr). Students from town A are then sent randomly to one of the schools in town B, with the numbers not violating the solution found in first step. For high school students, students can choose to attend any of the 41 high schools in Suwon, and as such treated in a similar strategy to tackling job commute allocation problem.
2.2.2 Job Allocation

For allocating jobs to individuals, the supply constraint is defined by number of 19~64 year olds in each town, multiplied by the employment rate for each age group. We found that the difference in employment rate pattern between each town to be minimal, and as such a common employment rate pattern is used, shown in Figure 2 below:

![Employment rate by age](image)

**Figure 2:** Employment rate by age.

The demand constraint is derived from the Suwon regional job statistics, which lists number of companies by their size (employee count) for each town. Two additional constraints are derived from regional commuting statistics: number of individuals commuting to own town for work, and number of individuals traveling to own region (cluster of towns) for work. With these constraints, we maximize the dispersion of commutes to different towns to avoid bias towards specific A/B town commute combinations. The solution derived is shown in Figure 3, where orange cells represent the commutes constrained by number of individuals commuting to own town for work, while green cells represents the commutes constrained by number of individuals traveling to own region (cluster of towns) for work. Each row must sum to supply constraint, and each column must sum to demand constraint.

In similar fashion to schools, companies are randomly created in each town while keeping faithful to the statistics on number of companies by their size (employee count) for each town. Overall, 67,000 companies ranging from 1~4 employee counts to 1000+ employee counts are created in Suwon. Individuals from town A are randomly assigned to job in town B, while keeping track of commute count solution found in step 1.

2.3 Transmission model

A variety of infectious disease models, such as SEIR, SIR and SIS have been studied (Smith, Wang, and Li 2001; d’Onofrio 2002; Earn et al. 2000; Shulgin, Stone, and Agur 1998; Hethcote 1976). This paper uses the SEIR model. SEIR model identifies each individual as belonging to one of the four states: susceptible, exposed, infectious, and recovered, as shown in Figure 4.

![SEIR model stages](image)

**Figure 4:** Four stages of SEIR disease infection model.
It is assumed in the SEIR model that an individual of the S group can be infected by the I group, and that the E group cannot infect other individuals. Individuals of group I are classified into those who are infected but have no symptoms, and those who have symptoms. Therefore, the S group is more likely to transition to the E group when it comes into contact with the I group than with the I\textsubscript{asym} group. However, in this paper I\textsubscript{asym} and I\textsubscript{sym} are not considered separately, but they are assumed to have the same infectivity. In addition, the SEIR model is a one-way model, in which individuals belonging to the R group are assumed to no longer infect even if they come into contact with individuals of group I, so that the number of infectable populations decreases and the spread of infectious diseases gradually decreases.

### 2.4 Simulation model

In this study, an agent-based simulation (ABS) methodology is applied to the transmission of disease between individuals in generated population. ABS is a technique where agents in a given environment are allowed to interact with each other through an autonomous judgment (Davidsson 2002; Bui and Lee 1999; Railsback, Lytinen, and Jackson 2006). The spread of infectious diseases is a phenomenon suitable for implementation in ABS, as the interaction between individual’s daily life patterns and decisions in a fixed geographical boundary determines movement and contact, leading to exposure, infection, and recovery. Therefore, it is possible to model the characteristics of the individual and the judgment of the general people during disease propagation, and to analyze the behavior change according to various scenarios. We utilized contact probabilities calculated by Chao et al. (2010), as shown in Table 1. SEIR parameter values, as shown in Table 2, are taken from Cowling et al. (2015) that analyzed the spread of MERS in Korea.

#### Table 1: Person-to-person contact probabilities in this study, taken from Chao et al. (2010).

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Child age0-2</th>
<th>Child age3-6</th>
<th>Child age7-18</th>
<th>Adult age19-64</th>
<th>Adult age65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family, infectious is child</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Family, infectious is adult</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Household cluster, infectious is child</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.035</td>
<td>0.035</td>
</tr>
<tr>
<td>Household cluster, infectious is adult</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Neighborhood</td>
<td>0.00000435</td>
<td>0.0001305</td>
<td>0.000348</td>
<td>0.000348</td>
<td>0.000696</td>
</tr>
<tr>
<td>Community</td>
<td>0.0000109</td>
<td>0.0000217</td>
<td>0.0000326</td>
<td>0.000087</td>
<td>0.000174</td>
</tr>
<tr>
<td>Workplace</td>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Daycare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>Elementary school</td>
<td></td>
<td></td>
<td></td>
<td>0.0348</td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td></td>
<td></td>
<td></td>
<td>0.0252</td>
<td></td>
</tr>
</tbody>
</table>

#### Table 2: SEIR parameters used in this study, as estimated by Cowling et al. (2015).

<table>
<thead>
<tr>
<th>Definition</th>
<th>MERS South Korea (2015) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Kim, Jung, Kim, Jung, Kim, Khan, and Park

<table>
<thead>
<tr>
<th>Mean Incubation Period</th>
<th>Period from Exposed state to Infectious state</th>
<th>6.7 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Serial Interval</td>
<td>Period from Exposed state to Recovered state</td>
<td>12.6 days</td>
</tr>
<tr>
<td>Case fatality Risk</td>
<td>Disease cannot be cured and dangerous</td>
<td>21 %</td>
</tr>
</tbody>
</table>

The next section describes the scenario-specific experiments, and results from the methods described in Chapter 2. The model was programmed in Anylogic 8.0, a JAVA based modeling platform.

3 MODEL EXPERIMENTS AND RESULTS

In this chapter, we conduct experiments on various strategic scenarios to mitigate the spread of disease, and evaluate the effect of each strategy accordingly. Using this model, it is possible to conduct a virtual experiment on national strategies and strategies that can be implemented, making it possible to evaluate strategies that can effectively mitigate the spread of disease. A 10% population is generated for use with this experiment due to memory limitations of our implementation.

Our experiment deploys three different mitigation strategies. In all three scenarios, the simulation begins with one pre-determined, fixed infected individual. In addition, the simulation execution time for each scenario is 180 days, with the parameter for the model shown in Table 3. The basic scenario contains no intervention strategies. The other three strategies are vaccine supply and isolation policy for infected patients. If the disease is allowed to spread without any intervention (scenario 1), the probability of contact with the infected person increases, and the probability that a large number of agents belonging to the susceptible state group becomes infected becomes high.

Table 3: Contact and infection probabilities utilized to simulate spread of MERS in Korea.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Probability</td>
<td>The probability of contacting the same space at the same time per unit time</td>
</tr>
<tr>
<td>Infection Probability</td>
<td>The probability of infection when it is in contact with the infected person at the same time and space</td>
</tr>
<tr>
<td></td>
<td>Table 2. Value</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
</tr>
</tbody>
</table>

Figure 5: Original Scenario results - Number of confirmed persons over time.
3.1 Basic scenario results

Figure 5 above shows the results for the first scenario. Without any intervention strategy, the contact between the non-infected person and the infected person continues to occur, and accordingly, the probability that agents in susceptible state moving to infectious state is increased, and a phenomenon of many infected individuals being generated can be seen.

3.2 Intervention strategies scenario results

Next, we conducted an experiment on strategies to mitigate propagation. The basic assumptions are the same as the basic scenarios, and they are modified according to the details of the scenarios so that the results can be compared and analyzed according to objective criteria. All experiments used fixed random seed to reduce variability.

3.2.1 Vaccination

One of the best pro-active strategies against disease spread is vaccination. For vaccination, we assume that people who receive vaccination do not get sick. We deploy two different vaccination strategies. Both strategies vaccinate similar number of individuals, but in the case of Vaccination 1, middle and high school students, who account for about 10% of the population, are vaccinated. The Vaccination 2 strategy randomly extracts about 10% of the population and vaccinates them. Both assume vaccines have been deployed sufficiently ahead of outbreak.

Table 4: Vaccination strategies analyzed in this study.

<table>
<thead>
<tr>
<th>Description</th>
<th>Vaccination1</th>
<th>Vaccination2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Vaccination to all middle / high school students who make up approximately 10% of the population</td>
<td>Vaccination to approximately 10% of population is randomly extracted</td>
</tr>
<tr>
<td>If vaccination, will not become infected</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.2 Isolation

Isolation is a reactive strategy where infected individuals are sent to an isolated hospital after a certain period of time since entering Infectious state. In such scenarios, the isolation of the infected person reduces the chance of infection by reducing contact with the non-infected person. The Isolation 1 strategy assumes identification and isolation 5 days after entering state I, and isolation strategy 2 assumes identification and isolation 3 days after entering state I.

Table 5: Isolation strategies analyzed in this study.

<table>
<thead>
<tr>
<th>Description</th>
<th>Isolation1</th>
<th>Isolation2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Isolation after 5 days of infectious state</td>
<td>Isolation after 3 days of infectious state</td>
</tr>
<tr>
<td>When infectious state is established, it is isolated to hospital after fixed period</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 ANALYSIS AND DISCUSSION

Our experiment investigates the effectiveness of pro-active measure (vaccination) and a reactive measure (isolation). Results indicate lower peak number of infected individuals with middle/high school vaccination versus random vaccination, highlighting the nature of infected school age children generating more
secondary cases than infected adults (Chao et al. 2010). Considering the practicality and cost-effectiveness of in-school vaccination, our results confirm the validity of vaccination of school age children.

![Figure 6: Vaccination scenario results - number of confirmed persons over time.](image)

For isolation scenarios, our results demonstrate how a delayed detection and isolation can result in significant reduction of its effectiveness, where isolation loses its effectiveness after a 5 day wait until isolation.

![Figure 7: Isolation scenario results - number of confirmed persons over time.](image)

## 5 CONCLUSION

This study investigates the development of agent based disease spread simulation model for Korea. Based on the 2010 census data and 2% micro data, a population model for Suwon city was constructed while considering the statistics of individual’s age, gender, family, residence, and commute area. Through utilization of job and school statistics, the expanded simulation model allows for analysis of disease spread within Suwon city, as well as assess the effectiveness of various countermeasures. The results of our experiment demonstrate how making changes within key strategies, such as selection of vaccination recipients and slower isolation reaction time, can have a significant impact on strength and lifespan of disease outbreak. Our model will also allow for investigation into other research questions, such as amount of time available before intervention becomes necessary, what % of population becoming infected, and whether combination of various strategies or one focused strategy is better overall.
The lack of large dataset for Korea makes it inherently difficult for proper calibration of parameters, and as such the results should be considered for assessing general trends and not the quantitative performance of each strategies.

6 FUTURE WORKS

For future works, our goal is expand this analysis from a city to a nationwide coverage. This would involve addressing issues such as cross city commutes, as well as migration to C++ language for faster run speed. Other areas of investigation include implementation of varying reactive behaviors of individuals to news of outbreak, as well as differing methods of transportation.

ACKNOWLEDGMENTS

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF). National Research Foundation grants [NRF-2015R1D1A1A01059750] and [NRF-2016R1A6A3A11932799] funded by the Ministry of Education, Science and Technology, Korea.

REFERENCES


**AUTHOR BIOGRAPHIES**

**YO-HAN KIM** is a Research Professor with the Dept. of Systems Management Engineering, Sungkyunkwan University. He received a MS in Industrial Engineering from Georgia Tech and Ph.D. in Business and Technology Management from Korea Advanced Institute of Science and Technology. His primary interests include disease spread modeling and modeling of socio-technical systems. His email address is yohan.kim@skku.edu.

**HYUN-JIN JUNG** is a MS student in Dept. of Industrial Engineering, Sungkyunkwan University. She received a BS in Industrial Engineering from Chosun University. She is interested in simulation-based modeling and scheduling and agent based simulation. Her email address is hj Jung217@naver.com.

**YOON-BAE KIM** is a professor with the Dept. of Systems Management Engineering, Sungkyunkwan University. He received a Ph.D. in Industrial engineering from Rensselaer Polytechnic Institute. He is interested in simulation-based modeling, simulation output analysis, market analysis and scheduling. His email address is kimyb@skku.edu.

**GI-SUN JUNG** is a Ph. D student in Dept. of Industrial Engineering, Sungkyunkwan University. He received a BS in System management Engineering from Sungkyunkwan University. He is interested in simulation methodologies and applications and output analysis. His email address is gsjung09@naver.com.

**YOUNG KIM** is a MS student in Dept. of Industrial Engineering, Sungkyunkwan University. He received a BS in System management Engineering from Sungkyunkwan University. He is interested in discrete event systems modeling and simulation based modeling. His email address is thramaster@naver.com.

**NOKHAIZ TARIQ KHAN** is a Ph.D. student in Dept. of Industrial Engineering, Sungkyunkwan University. His research interests include discrete event systems modeling and forecasting. His email address is nokhaiz.tariq@hotmail.com.

**JIN-SOO PARK** is an assistant professor with the Dept. of Management and Information Systems at Yongin University. He is interested in simulation-based modeling and scheduling. His email address is jsf001@yongin.ac.kr.