

# 센서스 데이터를 기반으로 만든 전염병 전파 시뮬레이션 모델

황교상 · 이태식<sup>†</sup> · 이현록

KAIST 산업 및 시스템공학과

## Epidemic Disease Spreading Simulation Model Based on Census Data

Kyosang Hwang · Taesik Lee · Hyunrok Lee

Department of Industrial and System Engineering, KAIST

Epidemic models are used to analyze the spreading of epidemic diseases, estimate public health needs, and assess the effectiveness of mitigation strategies. Modeling scope of an epidemic model ranges from the regional scale to national and global scale. Most of the epidemic models developed in Korea are at the national scale using the equation-based model. While these models are useful for designing and evaluating national public health policies, they do not provide sufficient details. As an alternative, individual-based models at the regional scale are often used to describe disease spreading, so that various mitigation strategies can be designed and tested. This paper presents an individual-based epidemic spreading model at regional scale. This model incorporates 2005 census data to build the synthetic population in the model representing Daejeon in 2005. The model's capability is demonstrated by an example where we assess the effectiveness of several mitigation strategies using the model.

**Keywords:** Epidemic, Modeling, Simulation, Census

### 1. 서론

전 세계적으로 대유행한 신종인플루엔자 A를 비롯하여, 최근 중동호흡기증후군 코로나바이러스, 새로운 조류독감 등 전염병에 대한 사회적 관심이 높아지고 있다. 이와 관련하여 거시적인 전염병의 전파 양상 분석, 의료 수요 예측, 그리고 중재 방법에 대한 효과 측정 등을 위해 수리적 모델과 시뮬레이션 모델을 이용한 많은 연구가 진행되었다(Hethcote, 2000, Dorjee *et al.*, 2013). 수리적 모델은 인구를 같은 특성을 가지는 몇 가지 집단으로 구분하여 집단 사이의 인구 이동을 수식으로 표현한다. 따라서 수리적 모델은 모델에 포함하는 인구 수가 많더라도 필요한 계산량이 상대적으로 적어 큰 범위의 지역에 대한 계산을 수행할 수 있으나, 개개인의 특성 등 상세한 세부 현상을 반영하기 어려운 단점이 있다. 우리나라에서는 국가를 하

나의 단위로 다룬 수리적 모델을 이용한 연구들이 주로 진행되었다(Department of Preventive Medicine and Public Health *et al.*, 2010; Kim, 2003; Park, 2009; Korea Centers for Disease Control and Prevention, 2005). 이들은 국민 전체를 대상으로 하는 피해 예측이나 방제 대책의 효과 분석을 목적으로 하였으며 그에 부합하는 모델을 구축하였다. 하지만 격리, 지정된 구역만을 대상으로 하는 지역이동제한 등 특정 개개인을 대상으로 하는 전염병 대응 정책을 모델 내에서 평가할 수 없다. 이러한 개개인의 특성을 반영할 수 없는 수리적 모델의 단점을 극복하기 위하여 시뮬레이션 모델이 사용되었다. 시뮬레이션 모델은 전염병뿐만 아니라 의료서비스 전반에 걸쳐 많이 쓰이는 방법론이다(Brailsford, 2007; Koo *et al.*, 2012; Kang *et al.*, 2013).

전염병 전파 시뮬레이션 모델은 3가지로 구분될 수 있다. 전 세계의 인구 이동에 의한 전염병 전파 양상 분석을 목적으로

이 논문은 2010년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2010-0025224).

이 논문은 2011년도 정부(미래창조과학부)의 재원으로 한국연구재단-공공복지안전연구사업의 지원을 받아 수행된 연구임(2011-0029883).

<sup>†</sup> 연락저자 : 이태식 교수, 305-701 대전광역시 유성구 대학로 291 한국과학기술원 산업공학과, Tel : 042-350-3166, Fax : 042-350-3110,

E-mail : taesik.lee@kaist.ac.kr

2013년 11월 26일 접수; 2014년 2월 25일 수정본 접수; 2014년 3월 24일 게재 확정.

지구를 표현한 모델, 한 국가 내에서의 전염병 전파 양상 분석, 피해 예측, 국가 단위의 대응체계의 효과 분석을 목적으로 국가를 표현한 모델, 지역을 단위로 삼아 지역 내 전염병 전파 양상, 피해 예측, 전염병 대응 효과를 분석하고 지역 간 인구 이동을 매개로 지역간 전파 양상, 피해 예측, 전염병 대응 효과를 분석하는 모델이 있다.

지역 수준의 시뮬레이션 모델을 구축하기 위해서는, 실제와 유사한 인구 모델을 구축하는 것이 중요하다. 지역 수준 모델이 국가 단위의 수리적 모델과 차이를 가지려면 인구 수, 면적, 위상 등의 지리적인 특성과 나이, 성별, 경제 활동 상태 등의 개인의 특성을 반영하여 모델 내 지역과 개인의 heterogeneity를 보장해줘야 한다. 이를 반영하기 위한 가장 객관적이며 접근할 수 있는 근거는 국가에서 시행하는 인구 전수 조사 자료이다. 이 자료를 이용하여 지리적 특성과 개인의 특성을 사실적으로 잘 반영하는 것이 지역 단위를 표현하며 개개인의 특성을 반영하는 시뮬레이션 모델을 특징짓는데 중요한 요소이다. 통상적으로 이러한 모델들에서는 인구 전수 조사 자료를 기반으로 다른 설문조사 등을 참조하여 인구 모델을 구축한다(Lee *et al.*, 2010, Chao *et al.*, 2010, Chen *et al.*, 2010).

국외의 경우, 잘 만들어진 지역 단위의 시뮬레이션 모델들이 존재한다(Smith *et al.*, 1995; Del Valle *et al.*, 2006; Chao *et al.*, 2010; Stein *et al.*, 2013). 하지만 모델의 특성을 특징짓는 인구 모델들은 모두 해당 국가의 인구 모델이기 때문에, 우리나라의 경우에 직접 적용하기 어렵다. 우리나라의 전염병 전파 모델들은 우리나라의 인구정보를 사용하였지만, 수리모델 기반의 국가 수준 모델들로, 개개인의 특성을 반영하는 시뮬레이션 모델이 아니다. 현재보다 더 자세한 수준의 전염병 대응 대책 수립 및 평가에 활용함을 궁극적인 목표로 하여, 본 논문에서는 인구 주택 총 조사를 기반으로 우리나라 인구의 특성이 반영된 인구 모델을 구축하였고, 이 인구 모델을 포함하는 지역 단위의 전염병 전파 시뮬레이션 모델을 개발하였다.

## 2. 선행연구

전염병 전파 모델은 모델링 기법을 기준으로 수리적 모델과 시뮬레이션 모델로 나눌 수 있다. 수리적 모델은 주로 국가 단위의 모델이 많으며 국가와 범 지구적인 표현 범위를 가진다. 시뮬레이션 모델은 주로 지역 단위의 모델이 많으며 표현 범위는 국가와 지역을 아우른다.

범 지구적인 표현 범위를 가진 수리적 모델은 전 세계 항공 네트워크를 이용하여 도시, 국가간 전염병 전파 양상을 볼 수 있는 Global Epidemic Model(Epstein *et al.*, 2007)과 Global Epidemic and Mobility model, GLEaMviz(Broeck *et al.*, 2011)가 있다. 국가를 다룬 수리적 모델로는 Korea Centers for Disease Control and Prevention(2005)의 신종 인플루엔자에 대비하기 위하여 환자 수, 감염자 수, 사망자 수 등을 추정하고 도상 훈련에 필요

한 시나리오를 작성하여 실제 도상 훈련 후 평가에 활용한 모델이 있다. 또한 조류 인플루엔자 방역을 위하여 수리적 모델을 기반으로 조류 인플루엔자 확산 모델을 구축한 연구(Park *et al.*, 2009)와 2009년도 신종인플루엔자 대유행시 국내에서 시행된 휴교, 백신, 항바이러스제 투입 등의 대응 전략을 평가한 연구(Department of Preventive Medicine and Public Health, 2010) 등이 있다.

지역을 단위로 국가와 지역을 모두 다룬 시뮬레이션 모델로는 National Infrastructure Simulation and Analysis Center의 교통 시스템과 인구 이동 패턴 시뮬레이션을 이용하여, 개인 간 접촉을 보다 상세하게 묘사하는 Epidemic Simulation System, EpiSims(Smith *et al.*, 1995; Del Valle *et al.*, 2006)가 대표적이다. Models of Infectious Disease Agent Study에서 개발한 FluTe(Chao *et al.*, 2010)는, 인구 전수 조사 자료를 기반으로 인구 모델을 만들어 지리적 특성을 표현하고 지역간 인구 이동을 구현하여 전염병 확산과 여러 방제 대책의 효과를 추정하는 데에 쓰일 수 있도록 개발되었다. 이 외에도 베이징, 델리와 로스앤젤레스 간의 지역 차이에 의한 전염병 전파의 차이를 분석한 연구(Chen *et al.*, 2010), 감염자가 자발적 격리를 하였을 경우 전염병 전파 양상과 이에 따른 음식 분배 문제를 다룬 연구(Ekici *et al.*, 2010)와 지리정보시스템과 연동하여 시각적인 정보와 의료자원의 수요예측을 제공하는 AsiaFluCap(Stein *et al.*, 2013) 등이 있다.

## 3. 전염병 전파 시뮬레이션 모델

본 논문에서는 접근 가능한 통계를 이용하여 우리나라의 인구 이동 패턴과 지역적 특징이 잘 반영될 수 있는 인구 모델을 구현하고, 인플루엔자가 발병하여 전파될 때의 양상을 시간에 따라, 지역에 따라 파악할 수 있고, 여러 시행 가능한 방제 전략의 효과를 시험해 볼 수 있는 전염병 전파 시뮬레이션 모델을 개발하였다. 모델의 구조적 측면에서, 본 모델은 FluTE의 모델링 프레임워크-인구 구조(population structure), 접촉 확률(contact probability), 전염 확률(transition probability) 정의-를 차용하였다.

인구 모델의 구축을 위해서 통계청에서 제공하는 2005년도 인구 주택 총 조사 데이터(이하 센서스 데이터)와 2005년도 인구 주택 총 조사 2% B형 데이터(이하 샘플 데이터)를 기반으로 가능한 실제에 가깝도록 인구 데이터를 생성하였다. 인구 데이터는 모델 내에서 인구를 생성하는데 기반이 되는 데이터로, 모든 개개인에 대한 신상정보(e.g. 성별, 나이 등)과 위치정보(e.g. 거주 지역, 활동지역 등)를 포함한다. 대한민국의 행정구역은 시/도, 시/군/구, 읍/면/동 순으로 구성되며, 본 연구에서는 행정구역 읍/면/동 단위를 전염병 전파 시뮬레이션에서의 기본 지역 단위로 모델링 하였다. 모델에서 필요한 읍/면/동 단위의 개인정보에 가장 가까운 통계 자료가 샘플 데이터이므로 이를 사용했다. 본 모델에서는 대전광역시에 대한 인구 모델

을 구축하였으며, 동일한 방법으로 대한민국 전체의 또는 다른 지역의 인구 모델을 생성할 수 있다.

### 3.1 인구 모델

인구 모델 구축의 시작은 인구 데이터 생성이다. 모델에서 인구 구성원 개인은 아래와 같은 데이터 형태로 정의된다:

{가족 id, 연령대, 거주지, 통근통학지}

가족 id는 개인이 어느 가족에 속하는지를 파악하기 위한 id 번호이고, 연령대는 전체 인구의 연령 구성을 0~7세(미취학 인구), 8~19세(취학 인구), 20~64세(경제 활동 인구), 65세 이상(은퇴 인구)로 나누어 각 개인이 속하는 연령대를 정의한다. 거주지와 통근통학지는 각각 야간과 주간 시간대의 활동 장소를 지정하는데 쓰이며, 읍/면/동 단위로 정의한다.

인구 데이터를 구축하기 위해서는 센서스 데이터와 샘플 데이터를 함께 활용하여야 한다. 샘플 데이터는 전국 2%의 인구 샘플에 대해 개인 정보를 담고 있는 인구 데이터로, 성별, 나이, 가족 정보, 거주지, 통근통학지(시/군/구 단위)를 제공한다. 샘플 데이터는 본 모델에서 필요한 정보를 제공하고 있으나, 이를 바탕으로 전체 인구 데이터를 생성할 때, 두 가지 문제점이 있다. 첫째, 샘플 데이터의 크기는 전체 인구 수의 2%이기 때문에, 나머지 98%에 대해서 가상적으로 인구 집단을 구성해야 한다. 둘째, 거주지와 통근통학지가 시/군/구 단위로 정의되어 있기 때문에, 본 모델의 지역 단위인 읍/면/동 단위의 정보를 임의로 생성해 내야 한다. 이러한 문제점으로 인해, 정확한 인구 데이터를 생성하는 것은 불가능하고, 전체 인구에 대한 자료인 센서스 데이터에서 얻을 수 있는 정보를 활용하여 가상 인구 집단을 구성한다. 센서스 데이터로부터 얻을 수 있는 정보는, 각 지역(읍/면/동 단위)에 대한 성별, 연령대별 인구 수이다. 이러한 두 가지 데이터로부터 인구 데이터를 생성하게 되는데 그 기본적인 과정은, 센서스 데이터의 각 읍/면/동 지역의 성별, 연령대별 인구 수만큼 해당 읍/면/동이 속하는 시/군/구의 샘플 데이터로부터 리샘플링(resampling)을 통해 인구 데이터를 생성한다.

FluTE에서는 연령대에 기반하여 접촉률과 주간 시간대의 활동 공간을 지정한다. 이에 따라 본 모델에서는 연령대를 미취학 인구(0~7세), 취학 인구(8~19세), 경제 활동 인구(20~64세), 은퇴 인구(65세 이상)으로 설정하였다. 연령대에 따라 주간 시간대의 활동공간이 정해지는데, 미취학아동은 놀이방이나 유치원에 존재하며, 취학 인구 중 8~13세는 초등학교로, 14~16세는 중학교로, 17~19세는 고등학교로 통학한다. 경제 활동 인구는 직장으로 통근한다. 취학 인구와 경제 활동 인구 중 통근 통학을 하지 않는 개인들과 은퇴 인구는 낮 동안 각자 밤 거주 공간인 집에 거주한다. 인구 주택 총 조사 설문조사에서는 15세부터 65세 이하를 경제 활동 가능 인구로 보았으며, 20세 이상의 인구도 취학연령에 포함되었으나 그 수가 소수이고, 낮의 활동공간이 나이에 의해 결정되기 때문에 취학 인구는 통

학만, 경제 활동 인구는 통근만 하도록 설정하였다.

인구 데이터를 생성하는 과정은 다음과 같다. 센서스 데이터에서 특정 동  $m$ 의 성별  $g$ , 연령대  $a$ 인 인구 수를  $X_{g,a}^m$ 라 하고, 샘플 데이터에서 구  $n$ 의 성별 연령대별 인구 집합을  $Y_{g,a}^n$ 라 하자. 동  $m$ 이 구  $n$ 에 속한다면, 균등 확률 분포(uniform probability distribution)에 따라  $X_{g,a}^m$ 만큼의 Agent를  $Y_{g,a}^n$ 에서 샘플링하고, 샘플링 된 Agent들에게 거주 동 코드  $m$ 을 부여한다. 위의 과정을 모든 구  $n$ , 모든 동  $m$ 에 대해 실시한다. 위의 샘플링 시 개인의 성별과 나이만을 고려했기 때문에 개인이 속하는 가구를 재설정하게 되는데, 이 과정은 제 3.1절 (1)에 자세히 기술하였다. 샘플 데이터에서 제공하는 통근통학지 정보는 개인이 거주 동과 같은 동으로 통근통학하는지 다른 동으로 통근통학하는지이며, 다른 동으로 통근통학하는 경우, 어떤 구에 속한 동으로 통근통학하는지에 대한 정보를 제공한다. 즉, 샘플 데이터의 통근통학지 정보는 거주지-통근통학지가 구 단위로만 알 수 있기 때문에, 모델에서 필요한 개인들의 동 단위 수준의 통근통학지 정보는 새로 생성해야 한다. 이를 위해, 앞의 샘플링 과정에서 부여한 개인의 거주 동 정보를 이용하여, 샘플 데이터의 통근통학지 정보를 보완한 개인의 동 단위 통근통학지 정보를 지정한다. 이 과정은 제 3.1절 (2)에 자세히 기술하였다. 이렇게 생성된 인구 모델은, 야간에 어느 동에서 누구와 함께 같은 집에서 거주하는지와 주간에 어느 동, 어떤 종류의 활동 장소에서 활동하는지에 대한 정보를 제공한다.

#### (1) 가구 구성

가구 구성은 모델 내에서 접촉을 결정짓는 중요한 요소이기 때문에 각 동의 연령대별 인구 수를 정확하게 만족시키면서, 샘플 데이터의 각 가구 타입별 비율과 최대한 가까운 가구 타입 분포를 생성하고자 한다. 샘플 데이터의 각 가구 타입별 비율은 구별로 개인의 가구 일련번호를 통하여 구할 수 있기 때문에, 각 구에 속하는 동의 가구 타입 수는 최적화 문제의 해를 구별로 각각 구하여 구한다.

#### 집합

- $m \in M$  행정동 집합
- $f \in F$  가구 타입 집합
- $a \in A$  연령대 집합

#### 입력 변수

- $r_f$  샘플 데이터의 가구 타입  $f$ 의 비율
- $P_{ma}$  동  $m$ 에 사는 연령대  $a$ 에 속한 개인들의 인구 수
- $c(f, a)$  가구 타입  $f$ 의 연령대  $a$ 인 가구 구성원 수

#### 결정 변수

- $H$  총 가구 수(자연수)
- $u_{mf}$  동  $m$ 에 배정될 가구 코드  $f$ 의 수(자연수)
- $v_f$  가구 코드  $f$ 의 수(자연수)

가구 타입은 가구 구성원 수와 각 가구 구성원의 연령대에 따라 가구 구성을 표현한다. 가구 타입은 연령대를 자리수로 가지고, 연령대에 해당하는 가구 구성원 수를 값으로 가진다. 20~64세 1인 가구의 가구 코드는 '0010'이며, 0~7세 2명과 20~64세 2명으로 이루어진 4인 가구의 가구 코드는 '2020'이다. 정확한 가구 수를 알 수 없기 때문에 전체 가구 수와 가구 타입별 가구 수는 결정 변수로 놓고 문제의 해를 구한다.

$$\text{Minimize } \sum_f (r_f \times H - v_f)^2 \tag{1}$$

Subject to

$$\sum_f \sum_m u_{mf} = H \tag{2}$$

$$\sum_m u_{mf} = v_f, \quad \forall f \in F \tag{3}$$

$$\sum_f u_{mf} \times c(f, a) = P_{ma}, \quad \forall m \in M, \forall a \in A \tag{4}$$

목적식 (1)은 샘플 데이터의 가구 타입 비율과 생성된 가구들의 가구 타입 비율 사이의 차이를 최소화 한다. 조건식 (2)와 (3)은 결정 변수들 간의 관계를 나타내는 식이며 (4)는 각 동의 연령대별 인구 수에 맞게 가구 타입을 만들도록 한다.

위의 문제를 ILOG사에서 제공하는 상용 소프트웨어 cplex를 이용하여 해 도출시, 특정 가구 타입이 특정 동에만 배치되는 현상이 발생한다. 위에서 구한 가구 타입들의 수는 유지하면서, 하나의 가구 타입이 동별로 고르게 분포할 수 있도록 새로운 최적화 문제를 세우고 해를 구하여 동별 가구 타입 수를 다시 구한다. 위에서 구한  $v_f$  값은 입력 변수로 사용된다.

$$\text{Minimize } \sum_j \sum_m u_{mj}^2 \tag{5}$$

Subject to

$$\sum_m u_{mj} = v_j, \quad \forall j \in F \tag{6}$$

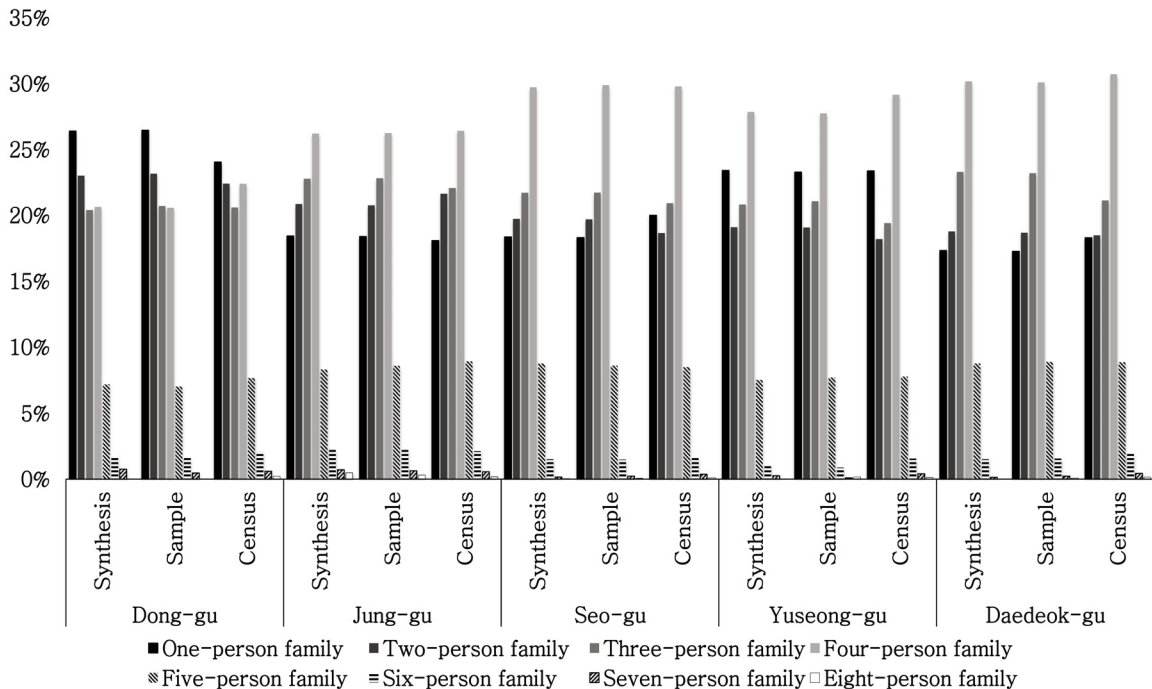
$$\sum_f u_{mf} \times c(f, a) = P_{ma}, \quad \forall m \in M, \forall a \in A \tag{7}$$

목적식 (5)는 각 동별 각 가구 타입 수들의 편차를 최소화 한다. 조건식 (6)과 조건식 (7)은 조건식 (3), 조건식 (4)와 같다. 두 문제의 해로부터 구한 동별 가구 타입 수에 맞도록, 가구의 종류와 수는 결정되며, 샘플링 된 개인들은 그들의 거주 동과 연령대에 따라 임의로 가구 일련번호를 부여 받는다.

**Table 1.** *p*-value for Student's t-test between solution and given data for each gu

Gu	Dong-gu	Jung-gu	Seo-gu	Yuseong-gu	Daedeok-gu
<i>p</i> -value	0.999	0.968	0.981	0.999	0.994

<Table 1>은 생성된 인구 데이터와 샘플 데이터의 가구 타입 비율 사이의 *p*-value를 구한 것이다. 5개구에 대해서 모두 0.9이상의 값을 얻었으며, 이는 두 데이터가 통계적으로 유의미한 차이가 없다는 것을 의미한다. <Figure 1>은 생성된 인구 데이터와 샘플 데이터, 그리고 센서스 데이터의 가구 구성원 수별 가구 비율을 나타낸 그래프이다. 인구 데이터와 샘플 데이터는 매우 유사함을 다시 한 번 확인할 수 있다. 인구 데이터와 센서스 데이터는 차이(최대 2.34%p)가 있지만 가구 구성원 수



**Figure 1.** Frequency of family sizes

별 가구 비율의 대소 관계 및 계형이 비슷하다.

(2) 통근통학지

통근통학을 위해 몇 명이 주간에 어느 동에서 어느 동으로 이동하는지에 대한 정보가 있으면, 생성된 인구의 거주 동을 근거로 통근통학 동을 지정할 수 있다. 이 정보는 생산 공급 최적화 문제의 해를 구하여 얻을 수 있다. 통근통학을 위해 다른 동으로 나가는 주간 유출 인구 수를 공급으로, 다른 동에서 들어오는 주간 유입 인구 수를 수요로, 그리고 동간 이동 비용을 1로 정하면 최적화 문제의 해를 통하여 동간 통행량을 구할 수 있다(<Figure 2> 참조). 주간 유출 인구 수는 통근통학 인구 수에서 동내 통근통학 인구 수를 제하여 구하고, 주간 유입 인구 수는 동별 종사자 수에서 동내 통근통학 인구 수를 제하여 구한다. 통근통학 인구 수와 동내 통근통학 인구 수는 새로 샘플링한 인구 정보를 기반으로 산출하고, 동별 종사자 수는 2005년 기준 사업체조사 통계(통근)와 2005년 기준 대전교육통계연보의 학교별 학생 수를 기반으로 산출하여 문제의 해를 구한다. 최적화 문제는 개인의 나이에 근거하여 통근, 초등학교, 중학교, 고등학교별로 따로 해를 구한다. 대전이 아닌 다른 도시로 통근통학하는 개인의 통근통학지는 모두 하나의 코드로 처리하고, 계산에서 제외했다. 또한 실제로 학교가 존재하지 않는 동의 동내 통학자는 모두 동 외 통학으로 코드를 변경했다.

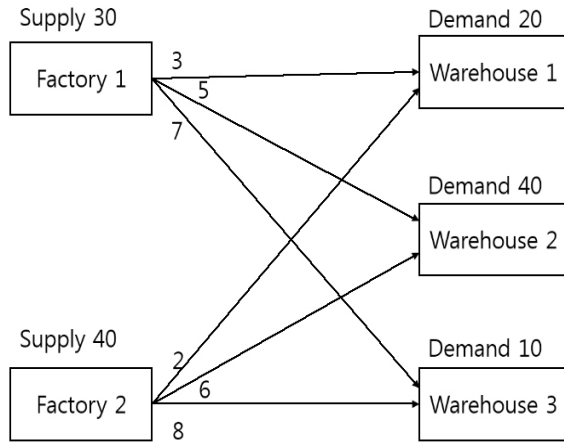


Figure 2. Simple transportation network

집합

- $m_1, m_2 \in M$  행정동 집합
- $n \in N$  행정구 집합
- $n \subset M$

입력 변수

- $T$  총 통근통학자 수
- $C_{m_1}$  동  $m_1$ 에 거주하는 통근통학자 수.
- $p_{m_2}$  동  $m_2$ 별 종사자 수 부분비
- $S_{m_1}$  동  $m_1$ 에서  $m_1$ 으로 통근통학하는 개인의 수

$D_{m_1n}$  동  $m_1$ 에서 구  $n$ 에 있는  $m_1$ 이 아닌 다른 동으로 통근통학하는 개인의 수

결정 변수

- $y_{m_1m_2}$  동  $m_1$ 에서 동  $m_2$ 으로 통근통학하는 개인의 수 (자연수)
- $z_{m_2}$  동  $m_2$ 의 종사자 수(자연수)

동의 종사자 수는 사업체조사 그리고 대전교육통계연보라는 2005년도 센서스 데이터와 다른 통계에서 값을 참조했기 때문에 절대적인 숫자를 직접 인용하지 않고 비율을 인용하였다.

$$\text{Minimize } \sum_{m_2} (p_{m_2} \times T - z_{m_2})^2 \tag{8}$$

Subject to

$$\sum_{m_1} \sum_{m_2} y_{m_1m_2} = T \tag{9}$$

$$\sum_{m_1} y_{m_1m_2} = z_{m_2}, \quad \forall m_2 \in M \tag{10}$$

$$\sum_{m_2} y_{m_1m_2} = C_{m_1}, \quad \forall m_1 \in M \tag{11}$$

$$y_{m_1m_2} = S_{m_1}, \quad \forall m_1 \in M \tag{12}$$

$$\sum_{m_2 \in n, m_2 \neq m_1} y_{m_1m_2} \geq D_{m_1n}, \quad \forall m_1 \in M, \forall n \in N \tag{13}$$

목적식 (8)은 센서스 데이터가 아닌 다른 통계에서 근거한 동별 종사자 수 부분비와 통근통학지 지정을 통해 얻어지는 동간 종사자 수 비율 사이의 차이를 최소화 한다. 조건식 (9)~조건식 (11)은 동간 통근통학자 수와 각 동의 종사자 수, 통근통학자 수와의 관계를 나타내는 식이며, 조건식 (12), 조건식 (13)은 동간 통근통학자 수를 결정하는 식이다. 조건식 (13)의 부등호는 위에서 수정한, 구가 지정되지 않은 동 외 통학자를 처리하기 위한 장치이다.

위의 문제를 ILOG사에서 제공하는 상용 소프트웨어 cplex를 이용하여 해 도출 시, 제 3.1절 (1)절과 유사하게 특정 동에서 특정 동으로만 통근통학하는 현상이 발생한다. 위에서 구한 각 동별 종사자 수는 유지하면서, 하나의 동에서 여러 동으로 고르게 통근통학 할 수 있도록 새로운 최적화 문제의 해를 구하여 동간 통근통학자 수를 다시 구한다. 위에서 구한  $z_{m_2}$ 는 결정 변수가 아닌 입력 변수로 사용된다.

$$\text{Minimize } \sum_{m_1} \sum_{m_2} y_{m_1m_2}^2 \tag{14}$$

Subject to

$$\sum_{m_1} \sum_{m_2} y_{m_1m_2} = T \tag{15}$$

$$\sum_{m_1} y_{m_1m_2} = z_{m_2}, \quad \forall m_2 \in M \tag{16}$$

$$\sum_{m_2} y_{m_1m_2} = C_{m_1}, \quad \forall m_1 \in M \tag{17}$$

$$y_{m_1 m_2} = S_{m_1}, \quad \forall m_1 \in M \quad (18)$$

$$\sum_{m_2 \in n, m_2 \neq m_1} y_{m_1 m_2} \geq D_{m_1 n}, \quad \forall m_1 \in M, \forall n \in N \quad (19)$$

목적식 (14)는 동간 통근통학자 수의 편차를 최소화시킨다. 조건식 (15)~조건식 (19)는 조건식 (9)~조건식 (13)과 같다. 이렇게 나온  $y_{m_1 m_2}$  에 따라 거주 동이  $m_1$ 인 개인은 통근통학 동  $m_2$ 를 무작위로 부여 받는다.

### 3.2 공간 모델

인구 모델의 구조와 더불어, 각 개인이 존재할 수 있는 공간에 대한 구조도 FluTE의 공간 모델 형식을 차용하였다. FluTE의 공간 모델은 House를 최소 단위로 하여, 점차 범위를 넓혀가도록 계층적으로 구성되어 있다. 본 모델에서도 이러한 계층적 공간 모델을 구성하였으며, 가장 상위 계층부터 동, community, neighbor, house cluster, house로 정의하였다. 행정구역의 기본단위인 동은 2,000명 내외의 인구로 이루어진 가상 인구 단위인 community들로 구성되고, 각 community는 500명 내외로 구성된 4개의 neighbors로 이루어진다. 각 neighbor는 다수 개의 house clusters를 가지고 있으며, 마지막으로 한 House cluster는 4 house로 정의된다. 하나의 house에는 인구 모델에서 동일한 가족 코드를 갖는 개인들이 가족 구성원으로 거주하게 된다(<Figure 3> 참조).

개인들이 주간에 속하게 되는 활동공간들-직장, 학교, 유아원-은 구조적으로 community안에 속하는 것으로 정의한다.

### 3.3 질병 모델

본 모델은 전형적인 SEIR 모델을 따른다. SEIR 모델은 개인의 질병 상태를 4가지로 구분한다. 즉, 전염병이 전파되는 상황에서, 개인의 질병 상태는 면역력이 없어 전염병에 감염될 수 있는 susceptible 상태, 감염되었으나 증상이 발현되기 전인 exposed 상태, 다른 사람을 감염시킬 수 있는 infectious 상태, 그리고 병에서 회복하여 면역력을 획득한 recovered 상태 중 하나

의 상태가 된다. 중 infectious 상태의 경우, 질병의 증상이 나타날 수도(symptomatic), 안 나타날 수도(asymptomatic) 있는 것으로 가정한다. SEIR 모델은 여러 전염병 모델 중 하나이며, 이러한 질병 모델은 여러 가지 질병에 대하여 다양하게 존재한다. 따라서 모델의 대상이 되는 질병에 적합한 모델을 선정하는 것이 필요하다.

감염될 수 있는 그룹(Susceptible, S)에 속하는 개인은 감염시킬 수 있는 그룹(Infectious, I) 내의 개인과 접촉했을 때, 모델상에 정의된 감염 확률에 의해 감염된다. S 그룹의 개인이 I 그룹 외의 개인(S; Exposed, E; Recovered, R)과 접촉하는 경우에는 감염이 발생하지 않는다. 감염이 발생하면, 감염된 개인은 일정 기간 동안 잠복기 상태(Exposed, E)에 속한다. E 그룹의 개인은 증상이 발현되지 않으며 또한 전염력이 매우 약하여 S 그룹의 개인과 접촉을 하더라도 감염시키지 않는 것으로 가정한다. E 그룹의 개인은, 잠복기가 경과하면 전염력을 가지는 상태 (I)가 된다. 이 때, 일정 확률로 증상이 발현되거나( $I_{sym}$ ), 발현되지 않을 수 있으며( $I_{asym}$ ), 증상이 발현한 개인( $I_{sym}$ )의 감염력은 그렇지 않은 개인( $I_{asym}$ )이 가지는 감염력의 2배인 것으로 가정하였다. 감염된 개인(E,  $I_{sym}$ ,  $I_{asym}$  그룹)은 감염된 시점으로부터 일정 감염기간이 지나면 질병으로부터 회복되어 면역력을 획득하며(R 그룹), 이에 따라  $I_{sym}$ ,  $I_{asym}$  그룹 내의 개인과 접촉을 하여도 더 이상 상태에 변화가 없다.

일반적인 전염병 전파의 경우, 초기에 매우 소수의 감염자가 인구 집단 내 감염될 수 있는 개인들을 다수 감염시키는 과정을 통해 감염자 수가 급속히 늘어나다가, 집단 내 감염자 수가 일정 수준을 넘어가면 감염자 수 증가가 둔화되는 양상을 보인다. 즉, 집단 내 새롭게 감염될 수 있는 사람의 수(S 그룹) 자체가 줄어들면서, 감염 건수가 줄어들고, 동시에 감염에서 회복하는 개인의 수가 지속적으로 늘어나면서, 결과적으로 전염병의 확산이 잦아들게 된다.

### 3.4 이산 시간 시뮬레이션

본 모델의 시뮬레이션은, 일정 시간 간격으로 시스템의 상태(지역 내 인구의 전염병 감염 상태)를 업데이트하는 이산 시

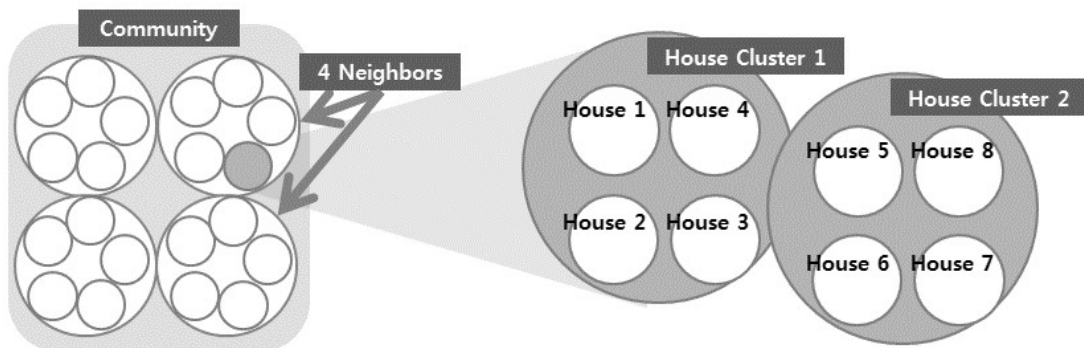


Figure 3. Population structure of FluTE model(Chao et al., 2010)

간 시뮬레이션 방식으로 실행된다. 단위 시간은 12시간으로 주간과 야간으로 나뉘며, 각 개인들은 주간과 야간 시간대에 인구 모델에서 정의하는 공간에 존재하게 된다. 즉, 주간에는 연령대에 따라 결정되는 활동공간에 존재하고, 야간에는 가족 구성원과 함께 집에 존재한다(<Figure 4> 참조).

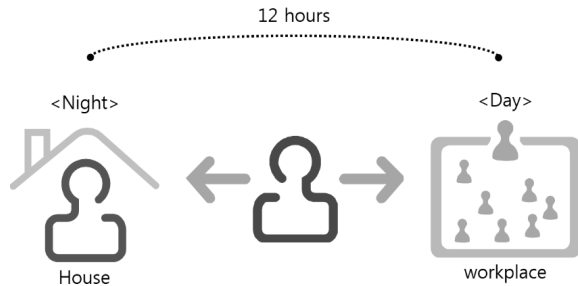


Figure 4. Diagram of the discrete-time simulation

각 단위 시간마다 개인들은 같은 공간에 존재하는 개인들과 일정 확률을 가지고 접촉을 한다. 만약 감염될 수 있는 개인(Susceptible)이 감염자(Infected)와 같은 공간에 존재하여 접촉을 하였을 경우, 일정 확률에 따라 병이 전염될 수 있다.

#### 4. 모델의 활용 : 전염병 대응 전략 평가

본 장에서는 본 모델을 활용하여 여러 전염병 방제 전략의 효과를 시험할 수 있음을 예제를 통해 보인다. 본 예제에서는 대전시에 대해 구성된 인구 모델을 사용하였으며, 전염병 관련 parameter는 아래 <Table 2>와 같이 정의하였다.

시뮬레이션은 4가지 시나리오에 대하여 100회씩 수행하였다. 시뮬레이션은 0일에 무작위로 10명을 선택하여 감염시키고 이후 180일 동안 전염병 전파 추이를 기록한다. 4가지 시나리오는 아무런 대응 전략을 세우지 않고 전염병을 전파시키는 시나리오와 3가지 가상 대응 전략을 시행하는 시나리오로 나뉜다. 가상 대응 전략은 초기 30일간 감염자 격리, 45일~60일 휴교령 시행, 인구의 18%에 대한 사전 백신 접종이다. 격리정책을 시행하면 증상이 나타난 감염 환자는 감염기간이 끝날 때까지 아무와도 접촉을 할 수 없도록 격리되고, 증상이 없는 감염자에 의해서만 전염병이 전파된다. 휴교령을 시행하면, 해당 지역의 모든 학생들은 주간 시간 대에 학교에 가지 않고 집에 머무르며 가족 구성원과만 접촉을 할 수 있다. 백신 접종을 받은 개인은 감염자와 접촉했을 때 감염될 확률이 정해진 비율만큼 감소하는 것으로 가정하는데, 본 예제에서는 백신

Table 2. Disease-specific parameters for the example

Parameter	Definition	Value
Transmission probability	The probability of disease transmission when infectious and susceptible person are contact each other	0.274
Days latent	The days between becoming exposed and becoming infectious	1 days
Days infected	The days between becoming exposed and becoming recovered	5 days
Symptomatic rate	The probability of infected person becoming symptomatic	0.67

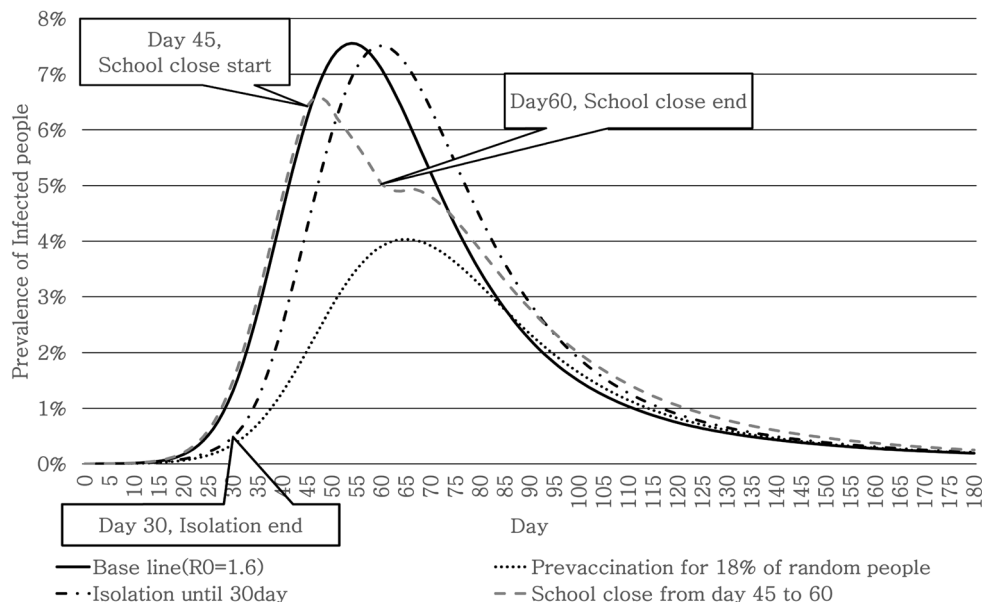


Figure 5. Disease spreading progress for alternative situation

접종 시 감염될 확률을 임의로 0으로 설정하였다.

<Figure 5>는 전염병 대응 전략에 따라 전염병 전파 추이가 어떻게 달라지는지를 보여주는 그래프로서, 실선은 아무런 방제 전략을 취하지 않았을 때의 전파 양상이다. 파선과 점선의 조합은 발생 초기 30일간 감염 환자를 격리하는 조치를 취했을 때 전파 양상이다. 감염 환자수가 적은, 전염병 발생 초기에 격리정책을 실시하면, 그래프에서 나타나는 것과 같이 전염병 전파가 지연되는 효과가 있다는 것을 알 수 있다. 파선은 전염병 발생 후 45~60일 사이에 휴교령을 시행했을 때의 전파 양상이다. 휴교령은 주간 시간대에 취학 연령 집단의 접촉 빈도를 감소시키며 전염병 전파를 막는데 효과적이다. 하지만 휴교령이 끝나면 다시 접촉 빈도가 증가하기 때문에, 휴교령에 의한 전염병 전파 감소 효과는 시행 시기와 기간에 큰 영향을 받을 것으로 예상할 수 있다. 실험 결과에서도 휴교령 직후 전염병 전파가 급감하다가 휴교령이 끝난 후 감소세가 줄어든 것을 알 수 있다. 점선은 사전 백신 접종을 했을 때 전파양상이다. 백신 접종은 감염될 수 있는 인구 수를 감소시키기 때문에 전염병 확산 감소에 직접적으로 영향을 줄 것으로 예상할 수 있으며, 실험 결과 그래프를 보면 peak volume이 크게 감소하였고, peak day 또한 지연되는 것으로 나타났다.

## 5. 결 론

본 연구에서는, 2005년 인구 주택 총 조사 자료(i.e. 전체 인구의 2%에 대한 성별, 연령 구조, 가구, 통근통학 여부 및 장소에 관한 정보와 읍/면/동 단위의 성별, 나이대별 인구 수)를 기반으로 우리나라의 대전광역시에 대한 인구 모델을 구축하였다. 이 인구 모델을 바탕으로, 개인별 이동 패턴에 따라 접촉 확률을 갖는 전염병 확산 시뮬레이션 모델을 구현하였다. 이 모델은 여러 대응 전략을 시험할 수 있는 모델이다. 모델에서 사용한 인구 모델 데이터 형식은, 우리나라 인구 조사 데이터 형식에 맞도록 설계되었으므로 국내 모든 지역에 대한 시뮬레이션 모델 구축이 가능하며, 또한 인구 구조를 이루는 집단(e.g. community, neighbor)의 구성원 수를 지역에 따라 조정하여 인구 밀도를 반영할 수 있도록 설계 되었다.

현재 외국 연구에서는 다양한 모사 수준을 가진 모델들이 개발되어 있으나 이들은 각 해당 국가 및 지역의 데이터와 환경 요소에 맞도록 제작된 모델로서, 우리나라에 적용하기 어려운 점이 있다. 국내의 경우, 기존 연구에서 사용된 수식 기반 전염병 전파 모델들이 있으나, 구체적인 전염병 대응 전략을 평가할 수 있는 상세한 수준이 반영된 모델은 찾아보기 어렵다. 이에 본 연구에서는, 전염병 시뮬레이션 연구에서 널리 활용되는 모델 중 하나인 FluTE에 기반하여 우리나라의 인구 모델을 반영한 개인 기반 전염병 전파 모델을 구축하였다. 본 연구에서 개발한 모델을 바탕으로 추가적인 검증과 보완을 통해 모델의 충실도를 높여 나간다면, 향후 전염병 전파 양상 이해

와 전염병 대응 전략의 효과 분석에 활용이 가능할 것으로 기대한다.

## 참고문헌

- Brailsford, S. C. (2007), Advances and challenges in healthcare simulation modeling : tutorial, In *Proceedings of the 39th conference on Winter simulation : 40 years! The best is yet to come*, 1436-1448, IEEE Press.
- Broeck, W. V., Gioannini, C., Gonçalves, B., Quaggiotto, M., Colizza, V., and Vespignani, A. (2011), The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale, *BMC infectious diseases*, **11**(1), 37.
- Chao, D. L., Halloran, M. E., Obenchain, V. J., and Longini Jr, I. M. (2010), FluTE, a publicly available stochastic influenza epidemic simulation model, *PLoS computational biology*, **6**(1), e1000656.
- Chen, J., Huang, F., Khan, M., Marathe, M., Stretz, P., and Xia, H. (2010), The effect of demographic and spatial variability on epidemics : A comparison between beijing, delhi, and los angeles, In *Critical Infrastructure (CRIS), 2010 5th International Conference on*, 1-8, IEEE.
- Del Valle, S. Y., Stroud, P. D., Smith, J. P., Mniszewski, S. M., Riese, J. M., Sydariak, S. J., and Kubicek, D. A. (2006), EpiSimS : epidemic simulation system, Los Alamos, NM : Los Alamos National Laboratory. Available at : <http://www.lanl.gov/programs/nisac/episims.shtml> (Accessed 30 March 2014).
- Department of Preventive Medicine and Public Health, Yonsei University College of Medicine, Department of Computational Science and Engineering, Yonsei University and Division of Public Health Crisis Responses, Infectious Disease Control Center, Korea Centers for Disease Control and Prevention (2010), Mathematical modeling of the pandemic novel influenza A(H1N1) virus and evaluation of epidemic response strategies in the Republic of Korea, *Public Health Weekly Report, KCDC*, **3**(19), 305-310.
- Dorjee, S., Poljak, Z., Revie, C. W., Bridgland, J., McNab, B., Leger, E., and Sanchez, J. (2013), A review of simulation modelling approaches used for the spread of zoonotic influenza viruses in animal and human populations, *Zoonoses and public health*, **60**(6), 383-411.
- Ekici, A., Keskinocak, P., and Swann, J. (2010), Modeling influenza pandemic, intervention strategies, and food distribution, *under review, Manufacturing and Service Operations Management*, 23-36.
- Epstein, J. M., Goedecke, D. M., Yu, F., Morris, R. J., Wagener, D. K., and Bobashev, G. V. (2007), Controlling pandemic flu : the value of international air travel restrictions, *PLoS One*, **2**(5), e401.
- Hethcote, H. W. (2000), The mathematics of infectious diseases, *SIAM review*, **42**(4), 599-653.
- Hwang, K.-S. and Lee, T. (2012), Identifying central nodes in pandemic disease spreading network, *Proceeding of Autumn Annual Conference of KHA*, Korean Institute of Industrial Engineers, 372-383.
- Hwang, K.-S. (2013), Study on the way of identifying central nodes using node property in epidemic disease spreading network, Master's Thesis, KAIST.
- Kang, S.-H., Kim, B., Jun, C.-H., Choi, B. K., and Lee, S.-H. (2013), Healthcare optimization : current status and vitalization suggestions, *Journal of the Korean Institute of Industrial Engineers*, **39**(4), 314-325.
- Kim, Y.-H., Park, J.-H., Kim, T.-H., and Moon, S.-A. (2003), A Study on the Damage forecast of Biological Terrorism : Focused on Smallpox,



- Journal of the military operations research society of Korea*, **29**(2), 26-44.
- Korea Centers for Disease Control and Prevention (2005), *Study on developing response strategy for new infectious disease by simulation*, Korea Centers for Disease Control and Prevention Research Report.
- Koo, J., Lee, G., Lee, J., Li, H., and Kim, B. (2012), Internet-based Generic Simulation Model for Outpatient Clinics, *Journal of The Korean Operations Research and Management Science Society*, **37** (4), 135-147.
- Lee, B. Y., Brown, S. T., Cooley, P., Potter, M. A., Wheaton, W. D., Voorhees, R. E., and Burke, D. S. (2010), Simulating school closure strategies to mitigate an influenza epidemic, *Journal of public health management and practice : JPHMP*, **16**(3), 252.
- Park, Y.-Y., Jung, Y.-Y., Won, D.-G., and Choi, N.-H. (2009), Modeling for Avian Influenza infection and spreading using system dynamics : Korean cases in 2007, *Korean System Dynamics proceeding*, 1-11.
- Smith, L., Beckman, R., and Baggerly, K. (1995), TRANSIMS : Transportation analysis and simulation system(No. LA-UR-95-1641), Los Alamos National Lab., NM(United States).
- Stein, M. L., Rudge, J. W., Coker, R., van der Weijden, C., Krumkamp, R., Hanvoravongchai, P., and Timen, A. (2012), Development of a resource modelling tool to support decision makers in pandemic influenza preparedness : The AsiaFluCap Simulator, *BMC public health*, **12**(1), 870.