Analysis of the Spread of Influenza From a Spatial Perspective

Shinichiro Iwata*1 and Ryo Nakamura2

1 Associate Professor, College of Industrial Technology, Nihon University, Japan
2 Keisei Construction Inc., Japan

Abstract

Current predictive models for influenza spread are based on an annual person trip data, which are difficult to interpret for a specific urban population. The present study tested whether total floor area according to building use (TFA) may constitute a more reliable alternative to derive a formula predicting the annual infection rate. This model was tested in 23 Tokyo wards based on spatial and epidemic data compiled from 2001 to 2011 using multiple regression analysis with TFA or person trip data as an explanatory variable and infection rate as objective variable. Furthermore, the annual epidemic patterns predicted by this model were validated against the influenza epidemic level maps available for 2004–2011. Results show that TFA is as effective as person trip data for the prediction of infection rate. A relationship was established between TFA and the number of infected individuals. The TFA model identified government offices and educational facilities as a major source of influenza spread to adjacent wards. The wards located away from the city center showed a higher infection rate than downtown Tokyo, despite their lower population density. In conclusion, this study suggests that TFA constitutes a more adequate variable for predictive models of influenza spread within an urban population.

Keywords: influenza spread; building use; multiple regression analysis; person trip; infection period

1. Introduction

Influenza has been a prevalent infectious disease in Japan every year. Data related to the spread of infection by physical contact and through air have been collected and analyzed, leading to the development of popular devices to eliminate the airborne virus. However, it is difficult to grasp the dynamics of the influenza epidemic in an urban environment.

In order to comprehend the spread of infection in an urban environment, it is necessary to consider the activities of the population from an architectural and spatial perspective. Behavioral models were developed on the basis of person trip data to predict the number of infected people in a year. However, many forms of influenza circulate around the world every year, making it difficult to grasp the dynamics of infection of an urban population based on trip data for one year. Therefore, new parameters must be identified to better monitor influenza epidemics over long periods.

The purpose of this study was to predict the rate of influenza infection based on urban environment usage over a decade, and to analyze the trends of infection spread from a spatial perspective.

2. Methods

This study was conducted in 23 Tokyo wards from 2001 to 2011. Authors hypothesized that the activities of people in an urban environment can be roughly estimated based on information concerning building use in the area. Infection rate was defined as the annual number of infected people over the total population in each ward. Authors tested the hypothesis that infection rate can be predicted by the total floor area usage, instead of person trip data, because the spatial data have been accumulated for years in all wards, and they are easier to quantify than person trip data.

First, authors derived formulas predicting the annual infection rate using the ward's data on the number of infected people and classified total floor area based on building use (TFA). Then, authors tested the validity of TFA against the formulas previously defined from person trip data. The data on the number of infected patients who received treatment from medical facilities from 2001 to 2011 were from the Tokyo Metropolitan Infectious Disease Report. The Tokyo Statistical Yearbook will be referenced for both statistical data and the 15 types of classification of building use.
Second, authors tested the accuracy of our prediction formula by exploring the relationship between the infection rates and building use as explanatory variables. Finally, using the history of the influenza epidemic level map (Note 1)\(^7\), authors conducted correlation analyses between building use ratios and the location or spread of infection. The influenza epidemic level maps from 2004 to 2011\(^4\) were used because they were the only data collected during 2001–2010.

In order to use disease data collected over several years, authors must consider that a number of individuals are counted repeatedly because an infection generally persists for several years. Fortunately, the infection cycle for influenza (from onset to end) is completed within a year. Nevertheless, the form of influenza can change from year to year. Therefore, authors analyzed each disease independently. Because the influenza epidemic continues throughout the year, authors counted the number of infected people during a "cycle" starting on the 25\(^{th}\) week of one year (week with the lowest number of recorded infections) and ending on the 24\(^{th}\) week of the following year.

### 3. Prediction of Infection Rate

#### 3.1 Prediction of Infection Rate Based on TFA

Multiple regression analyses were conducted on each ward for each year using a round-robin algorithm. Authors used each ward's infection rate as the objective variable and each ward's TFA as the explanatory variable. This analysis provided prediction formulas and combinations of explanatory variables with the highest coefficient of determination. To take into account the multicollinearity between the explanatory variables, the variance inflation factor (VIF) of the explanatory variable was obtained from the prediction formula. If VIF was >10, the explanation variable was excluded. Furthermore, a test for significance was also conducted for each explanatory variable. If the p value was >0.05, the corresponding explanatory variable was also excluded.

Table 1. shows the list of explanatory variables and the resulting standardized partial regression coefficient. When the value is positive, the explanatory variable shows that the infection rate increases with TFA, and vice versa. The coefficient of determination was >0.5 in 2001–2005 and 2010–2011, but <0.5 in 2006–2009. During the years with low coefficient of determination, a large number of explanatory variables were excluded based on the p value.

Regarding the different types of influenza, a new type (A/H1pdm) dominated during 2008–2010, and the coefficient of determination was low in two of these three years. Although we have determined a correlation between the coefficient of determination and the influenza type, it is not enough to conclude based only on the influenza type each year. The analysis should consider the number of infected individuals by each influenza type each year.

Concentrating on the seven years with high coefficient of determination (2001–2005 and 2010–2011), authors determined the number of times that building use was cited as the explanatory variable, and noted the directionality of the standardized partial regression coefficient. A positive correlation was cited four times for "government office facilities," three times for "complex of residence and commerce," and twice for "education and cultural facilities." Although these facilities are used by an unspecified number of people, similar to "office buildings" and "exclusive commercial facilities," the difference is that the primary users of the former facilities are local residents. In contrast, negative correlations were cited...
three times for "complexes of residence and factory" and twice for "exclusive commercial facilities," "detached houses," "apartments," and "accommodation and entertainment facilities." Therefore, more types of buildings were cited more than once for negative than positive correlations. The difference arises because of the inclusion of residential facilities. "Office buildings" was cited with a negative correlation in 2003 and a positive correlation in 2004.

3.2 Comparing the TFA and Person Trip Formulas

Person trip research documents the movements of individuals in certain regions once every 10 years to capture the actual condition of transportation. The movement of an individual in a day is called a "trip," defined by a "departure place," a "destination," a "purpose," a "means," and a "time" of transportation. In this study, the total number of individuals who moved to each destination is calculated for each ward and defined as the number of trips. Destinations were used as explanatory variables. It was assumed that their activities in an urban environment would not completely change in 10 years. Therefore, authors used the latest person trip data researched in 2008 to predict the infection rate.

Table 2. shows the list of explanatory variables and the resulting standardized partial regression coefficients for 2001–2010. Authors detected a similar trend as in Table 1. in terms of directionality, despite the different coefficients, which supports the relationship between person trip and infection rate. Aside from the years 2001 and 2006, the values obtained with the TFA formula are very similar to those obtained with the person trip formula (Fig.1.). Altogether, these data suggest that the activities of the urban population have a strong impact on influenza spread in most years.

The TFA and person trip data cannot be compared in terms of explanatory variables because Table 1. and Table 2. present different lists of items. Therefore, authors attempted to compare the items common to both tables. "Government office facilities" and "religious facilities" were cited only with positive correlations, whereas "educational facilities" was cited only with a negative correlation in Table 2. Even though "government office facilities" was similarly cited only as the positive correlation in Table 1., TFA of the government offices in a ward is expected to correlate with the number of individuals who use the government office in the ward.

In contrast, while "cultural and religious facilities" and "educational facilities" in Table 2. belong to the "educational and cultural facilities" category in Table 1., the former was always cited as a positive correlation and the latter as a negative correlation. This analysis
suggests a trade-off between the TFA of education facilities in a ward and the number of individuals who commute to education facilities in the same ward. It is expected that the more schools there are in a ward, the less people commute to the school in the wards and infection rate decreases.
"Office and bank buildings" was cited five times, with positive correlations for three years and negative correlations for the other two. The positive coefficients were extremely high compared with the negative ones, suggesting a strong influence of the movement of office workers on influenza spread.

4. Spatial Analysis Using an Influenza Level Map

4.1 Changes in the Number of Infected Individuals and Infected Wards

The purpose of this section was to determine why the infection rate could be predicted in some years, and not in the others, by analyzing the infection process spatially from the outbreak to the end. Fig.2 shows the influenza epidemic level maps for 2005–2012 with each infected ward where alarms or cautions were issued (Note 2) and how they evolved from the onset to the end. Table 3. shows how the number of infected wards was counted each week using the influenza epidemic level maps as an example. The changes in the number of infected wards and infected individuals were associated with the coefficient of determination.

Fig.3. shows the changes in the number of infected individuals from 2001 to 2011. Classification of the changes in number of infected individuals based on the coefficient of determination generated a single peak during 2001–2005 with a coefficient of determination > 0.7. The peak was consistently positioned around week 4–14. In 2011, there was a remarkable increase in the number of infected individuals grouped in a single peak around week 4–14. The coefficient of determination was relatively high (>0.5). In 2007–2009, a bimodal distribution with a smaller secondary peak is seen. The duration of infection (15 weeks) was longer than that in other years, and the coefficient of determination was < 0.5. In 2006, while the distribution resembled that of 2001–2005, the coefficient of determination was low (0.371). In 2010, there was an irregular transition out of the peak with a coefficient of determination > 0.7. Therefore, the numbers of infected individuals was represented by a simple peak in 9 of 11 years, and the coefficient of determination was high, or the number of infected individuals was not represented as a simple peak and the coefficient of determination was low. Regarding 2006 and 2010, further analysis is necessary and epidemic of a new type of influenza could have influenced 2010.

Fig.4. depicts the changes in the number of infected wards from 2004 to 2011. The relationship with the coefficient of determination is fundamentally similar to that obtained with the number of infected individuals (Fig.3.). Thus, it is clear that the change in the number of infected individuals is linked to the increase or decrease in spatial areas where infection exists. In 2006, the peak was approximately 4 weeks later than in 2004 and 2005; however, the reason remains unclear.

4.2 Infection Period and Use in Ward

For each of the 23 wards, authors calculated the duration ratio as the caution level over the alarm level from the onset of influenza until the end (hereafter referred to as the ratio of infection duration). Table 4. shows the ratio of infection duration in each ward. Using the average ratio of infection duration as a criterion, authors identified the seven wards with the shortest duration of infection (Group S), and the seven wards with the longest duration of infection (Group L),
then authors compared the building use and location of Group S and Group L. Group L wards included "Edogawa," "Katsushika," "Koto," "Nerima," "Arakawa," "Suginami," and "Ota." These wards, located towards the periphery of Tokyo (Fig.5.), have a higher ratio of building uses relative to living environment (Fig.6.). Group S wards included "Meguro," "Chiyoda," "Shibuya," "Itabashi," "Toshima," "Minato," and "Chuo." These wards have a higher proportion of government office facilities and office buildings (Fig.6.), including the major Tokyo business district and the downtown area.

Time-course analysis of the infected wards revealed that Group S wards became infected after the Group L wards. Furthermore, infection in the Group S wards ended relatively early after the peak, rarely reaching the alarm level. Because the city center has more individuals going in and out and a higher population density, it is presumed that these wards would constitute the main route of infection and the risk of infection would be higher. However, this presumption was proved incorrect when authors found that the wards located away from the city center had a higher rate of infection from onset to end.

5. Conclusions
The present study established a strong relationship between the number of infected individuals and TFA using data compiled over a period of 11 years. Authors have shown that TFA data can be used instead of person trip data when the risk of infection is calculated based on population movements. In addition, we showed that the accuracy of the prediction formula was high in 9 of 11 years, when the changes in the number of infected individuals or wards were represented by a single peak. The predictability of the infection rates, using TFA as explanatory variable, allowed us to determine the impact of specific building use on the infection rate. It is clear that "government office facilities" and "educational facilities," which are frequently used by local residents, have the highest correlativity with the infection rate. The lowest values were ascribed to "exclusive commercial facilities," which are frequently used by individuals from outside the ward. This study has shown that we need to take preventive actions in the local public facilities that significantly contribute to the spread of infection to other wards.

In addition, we confirmed that the predicting formula based on the person trip data supports correlativity between the number of transference of office workers and the infection rate. It is inferred that there is no correlativity between the infection rate and the space which office workers frequently use and it is not listed in Table 1. because there is no distinguishing correlativity between the TFA of "office buildings" and infection rate. Commuter trains meet this condition and may effectively spread the infection.

The infection period of residential areas located in the periphery of Tokyo was longer than that in the downtown area, despite a lower population density, because influenza epidemics were always detected first in the residential area and a patient's sphere of activity is restricted to their neighborhood after contracting the infection. These data suggest that infection control measures should prioritize the residential areas because they constitute a major route of infection.

New studies to identify the parameters that would explain the years that the TFA formula could not predict the infection rate are currently underway.

References
1) Sakai, T. (2003) Use of Geographic Information System for Determining the Epidemiology of Influenza Viruses Infections. Niigata medical journal, 117(11), pp.626-635.
2) Ejima, K. Suzuki, H. and Alhara, K. (2010) A mathematical model of the spread of swine flu based on trip data in Tokyo urban area. Institute of Transportation Economics, Transportation & Economy, 70(1), pp.54-62.
3) Christos, N. Luis, C. F. Marta, C. G. and Ruben, J. (2012) A Metric of Influential Spreading during Contagion Dynamics through the Air Transportation Network. PLoS ONE, e40961.
4) Tokyo Metropolitan Institute of Public Health (2002-2012) Tokyo Metropolitan Infectious Disease Report.
5) Statistics Division, Bureau of General Affairs, Tokyo Metropolitan Government (2002-2012) Tokyo Statistical Yearbook.
6) Bureau of urban development Tokyo metropolitan government (2008) Person trip research.
7) The National Institute of Infectious Diseases (2004-2011) Weekly reports of influenza virus isolation/detection.
8) Nakamura, R. and Iwata, S. (2012) 9th International Symposium on Architectural Interchanges in Asia. Prediction of the spread of influenza infection based on gross floor space by usage.

Notes
1) The influenza epidemic level represents the ratio of the number of health centers that exceeded the caution and alarm level in each prefecture over the total number of health centers. Prefectures are shown in three shades of yellow when one or more of their health centers exceeded the caution level. Prefectures are shown in three shades of red when one or more of their health centers exceeded the alarm level. In Tokyo, the information is given through the city unit, and within the 23 wards where there is one health center per ward.
2) A caution level is reached when the number of infected individuals per week in a ward exceeds the reference value of 10. The alarm level is reached when the number of infected individuals per week in a ward exceeds the reference value of 30. Once an alarm level is reached, the alarm level continues until the number of infected individuals per week becomes less than the reference value of 10.