Quantitative Epidemiological Understanding of Influenza Propagation Process in Tokyo and Environs

Tsuyoshi Kobari,[a] Mizuho Takahashi,[b] Kazushige Ijuin,[c] Hisako Takeuchi,[d] Kazuo Iwaki,[e] Fumiyoshi Ishii,[f] Rieko Matsuda,[g] Yuzuru Hayashi,*[,g] and Takehiko Yajima[h]

[a]Kosumo Chouzai Yakkyoku Co,. Ltd., 2–129–3 Miyahara, Kita, Saitama, Saitama 331–0812, Japan, [b]Nikko Citizen's Hospital, 1752–10 Arasawa, Kiyotaki, Nikko, Tochigi 321–1441, Japan, [c]Tanashi Yakuhin Co,. Ltd., 4–25–5, Tanashi, Nishi-Tokyo, Tokyo 188–0011, Japan, [d]Triad Japan Co., Ltd., 2–18–10 Kitasato, Sagamihara, Kanagawa 228–0829, Japan, [e]School of Pharmaceutical Sciences, Ohu University, 31–1 Misumido, Tomitamachi Koriyama, Fukushima 963–8611, Japan, [f]Department of Pharmaceutical Sciences, Meiji Pharmaceutical University, Noshio 2–522–1, Kiyose, Tokyo 204–8588, Japan, [g]National Institute of Health Sciences, 1–18–1 Kami-Yoga, Setagaya, Tokyo 158–8501, Japan, and [h]Faculty of Pharmaceutical Sciences, Toho University, Miyama 2–2–1, Funabashi, Chiba 274–8510, Japan

(Received May 22, 2006; Accepted June 21, 2006; Published online June 22, 2006)

Here we show a quantitative, spatial interpretation of influenza propagation process in Tokyo and its environs in the season from November 1st, 2004 to October 31st, 2005. The time lags (day) of influenza propagation between distant sites are calculated by the cross-correlation functions of the daily variations in the amount of drug sale at community pharmacies. The influenza infection appears to have spread from the urban area of Tokyo to its suburbs in the season of 2004–05. From the time lags and distances of the pharmacy locations, the mean propagation speed is estimated to be 3.5 km/day.

Key words —— correlation, pharmacy, influenza, propagation

INTRODUCTION

The importance of the spatial structure of disease propagation in human and other populations has been recognized ever since John Snow identified the source of cholera outbreak by mapping the cases in 1854 for the city of London.[1–12] Ordinarily, the pattern of infection spread is visualized on a map and understood by techniques such as choropleth maps and geographic information systems.[1–7]

Recently, a method for estimating the route and speed of influenza propagation was proposed.[13–15] A cross-correlation functions, often encountered in spectral analysis, is used for this purpose. In general, the function gives the time lag of similar or correlated phenomena which occur in distant places.[16] A typical application is to a meteorological problem about the time needed for rainwater to move from a mountain to downstream lake. Because of the definite causality between the phenomena (precipitation and inflow volume), the cross-correlation function between them provides an estimate of the arrival time.

Similarly, the route and speed of influenza propagation were estimated by the cross-correlation function of the time series of drug sales at distant pharmacies under the assumption that the health conditions of people are reflected by the drug sales at a pharmacy of the area where they live.[13,14]

However, the previous study[13,14] collected the information from the minimum number of pharmacies (three) and the estimated propagation route and speed are no more than a mere application of the new proposition. Here, we use the information from fourteen pharmacies in Tokyo and its environs in the season of 2004–05 for the same purpose. The most prominent features of our study are the mathematical technique and sources of disease information, *i.e.*, the cross-correlation functions and pharmacies. Because of the nature of the information, our study does not violate the Act on the Protection of Personal Information.

MATERIALS AND METHODS

The information about the prescriptions of influenza drugs was offered by the pharmacies, a–n, listed in Table 1. Pharmacy i is located near an emergency hospital and the other pharmacies are near general clinics (but not emergency ones). The pharmacy data were available throughout the year, including those of weekends as well.

*To whom correspondence should be addressed: National Institute of Health Sciences, 1–18–1 Kami-Yoga, Setagaya, Tokyo 158–8501, Japan. Tel.: +81-3-3700-1141; E-mail: fumi@nihs.go.jp

Table 1. Time Lags (day) between the Sales of Tamiflu® Capsules at Pharmacies a–n

$A(t)$	$B(t)$														Mean	Relative
	a	b	c	d	e	f	g	h	i	j	k	l	m	n		Lag
a	0	−4	−3	−5	0	−2	−8	−8	−7	−9	−9	−7	−4	−3	−4.9	12.1
b	4	0	2	1	0	4	−3	−3	−5	−2	−8	−2	1	−6	−1.2	8.4
c	3	−2	0	−4	1	−1	−6	−5	−9	−7	−13	−4	−3	−1	−3.6	10.9
d	5	−1	4	0	8	0	−1	−3	−8	−3	−9	−5	−1	4	−0.7	7.9
e	0	0	−1	−8	0	6	−9	−7	−8	−14	−8	−7	−5	−5	−4.7	11.9
f	2	−4	1	0	−6	0	−4	−2	−6	−3	−10	−3	1	−4	−2.7	9.9
g	8	3	6	1	9	4	0	−1	−2	−5	−6	−3	2	3	1.4	5.9
h	8	3	5	3	7	2	1	0	−2	−1	−6	1	3	5	2.1	5.1
i	7	5	9	8	8	6	2	2	0	5	−1	2	7	1	4.4	2.9
j	9	2	7	3	14	3	5	1	−5	0	−3	1	2	7	3.3	3.9
k	9	8	13	9	8	10	6	6	1	3	0	7	8	13	7.2	0.0
l	7	2	4	5	7	3	3	−1	−2	−1	−7	0	4	4	2.0	5.2
m	4	−1	3	1	5	−1	−2	−3	−7	−2	−8	−4	0	3	−0.9	8.1
n	3	6	1	−4	5	4	−3	−5	−1	−7	−13	-4	−3	0	−1.5	8.7

The locations of pharmacies a–n are shown in Fig. 2. $A(t)$ and $B(t)$ denote the time series included in the cross-correlation function, $R(\tau)$: $R(\tau) = \frac{E[A(t)B(t+\tau)]}{\sqrt{E[A(t)^2]E[B(t)^2)]}}$ where E[.] denotes the mean over time, t, and τ is a parameter called lag.[15]

RESULTS AND DISCUSSION

Figures 1A and 1B show the daily variations in the sales amount of Tamiflu® capsules (Chugai Pharmaceutical, Tokyo, Japan) at the pharmacies which are about 25 km away (see i and m in Figure 2). The fluctuation which looks like spike noises with a hebdomadal cycle comes mainly from the life style of people in the society. However, these "gnoises" are eliminated from the cross-correlation function (C) by the moving average method with a window of seven days.[15]

The daily variations (A and B) are similar in pattern, but do not synchronize. In this case, the cross-correlation function is helpful to know the time lag between the phenomena.[13] Figure 1C is the plot of the correlation coefficients between time series, $A(t)$ and $B(t + \tau)$, as a function of τ (also see Table 1):[16] Here, $A(t)$ denotes the daily variation of Fig. 1A and $B(t)$ that of Fig. 1B. The cross-correlation function (C) has the maximum when $\tau = 7$ days. This means that the correlation is the strongest, when one phenomenon, $B(t + \tau)$, is shifted by seven days. Therefore, the maximum position of the plot means the time lag. Because of the causality of drug sales and disease, we can safely say that that the influenza infection around the pharmacy of Fig. 1B lags seven days behind that of Fig. 1A.

This paper takes fourteen pharmacies as a local network (for locations, see a–n of Fig. 2). Then, the total number of cross-correlation functions is sup-

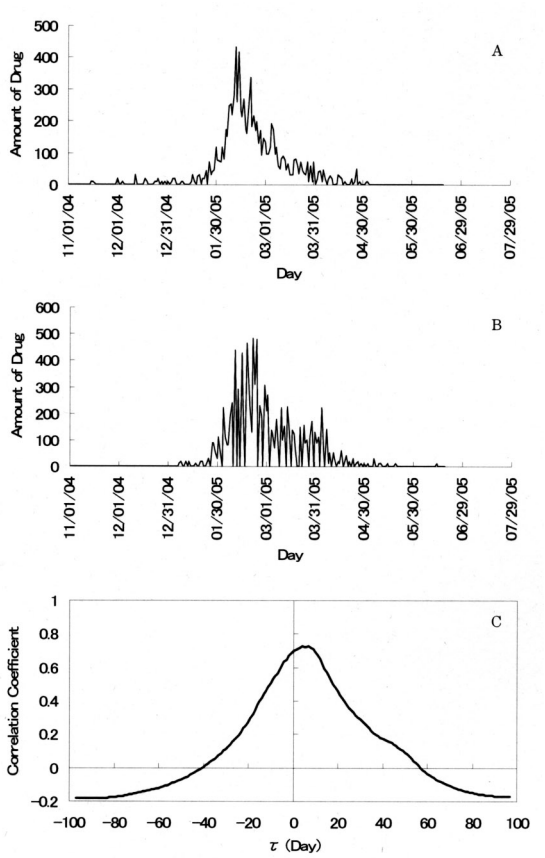

Fig. 1. Daily Variations in the Sales of an Influenza Anti-Viral Agent (Tamiflu® Capsule) at Pharmacies (A and B) and Cross-Correlation Function (C) of the Variations in the Season from November 1st, 2004 to October 31st, 2005

(A) pharmacy i; (B) pharmacy m (the locations are shown in Fig. 2). The Y-axis denotes the number of capsules dispensed at the pharmacies. The correlation function (C) is smoothed by the moving average method with a window of seven days.

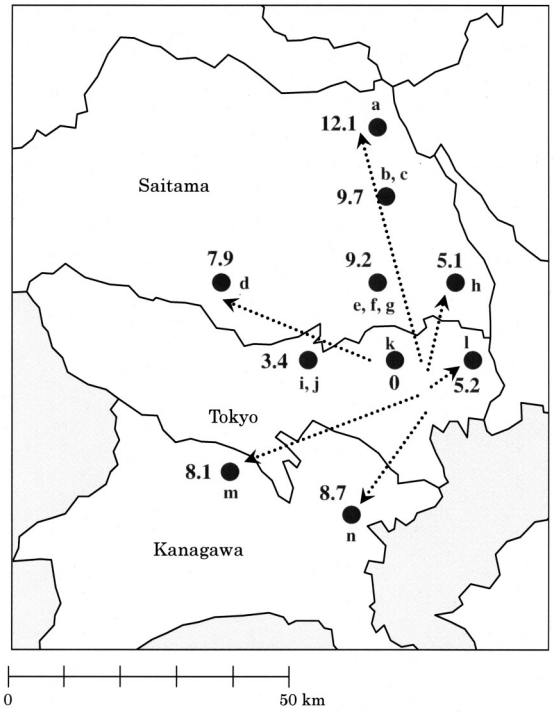

Fig. 2. Influenza Propagation Routes with Delay Times (day) from Pharmacy k in Tokyo and Environs in 2004–05 Season

The sites of pharmacies a–n are marked with circles. The delay times indicated are the values of Relative Lag of Table 1. The time delays of close pharmacies are averaged. The propagation routes are indicated by the arrows.

posed to be the combinations of 14 things taken 2 at a time ($_{14}C_2 = 91$). If 91 lines marked with the time lags were drawn among the pharmacy sites, the map would have been a mess. To circumvent this problem, we determine the order of infection periods among the sites.

Table 1 lists the time lags from the permutations of data ($_{14}P_2 = 182$). The result of Fig. 1C is in (i, m). The diagonal elements of the matrix is zero, since they are derived from the "auto-correlation" functions. The second rightmost column includes the mean of the time lags between a fixed pharmacy [$A(t)$] and each of the others [$B(t)$]. These mean lags can be regarded as the relative order of infection, but statistical verification is inevitable for this interpretation as mentioned below.

We can put the infection order in other words. Pharmacies a–n are ranked with respect to the time lags as shown by the 14×14 matrix of Table 1. The concordance of the different ranking results (each row of the matrix) can be tested by the Kendall's coefficient of concordance. The coefficient calculated (= 0.74) is close to the perfect concordance (= 1). Therefore, the ranking is statistically significant and the averaged lags are acceptable.

By the definition of this paper, the highest value of the time lag corresponds to the earliest infection. Pharmacy k has the highest mean lag (= 7.2 days). Then, the delay of infection at another pharmacy can be described relatively to the mean lag of pharmacy k (see Relative Lag of Table 1 and for proof, see paragraph appendix).

The delay times of local infection (Relative Lag), mapped in Fig. 2, increase with increasing distance from the heart of Tokyo. No exceptions are observed in this study. Although the spatial order of the infection does not directly indicate the propagation route, it is probable that the influenza infection spread from the urban area of Tokyo to its suburbs through the public transport systems in the season of 2004–05. In the area, the transport facilities are stretched in a radial pattern and are intertwined complicatedly to enhance the carrying power. The sites of the pharmacies (● of Fig. 2) can easily be reached from the center of Tokyo through the railroad systems and are within the sphere of commutation. The arrows in Fig. 2 denote, though roughly, both the disease propagation routes and rail lines. A similar spatial pattern of influenza from an urban area to suburbs was suggested previously.[4)]

The scale of the pharmacy network used is not large enough to discover infection foci from which the epidemic starts to propagate. However, pharmacy k can be considered to be near a focus. The existence of another infection focus cannot be indicated from the data of this study. The disease propagation speed along the route from pharmacy k to another can be estimated: a, 3.5; b and c, 3.1; d, 4.4; e, f and g, 1.6; h, 3.5; i and j, 4.6; l, 2.7; m, 4.4; n, 3.3 km/day. The average is 3.5 km/day. The ratio of the highest speed to lowest speed (= 4.6/1.6) is so small that it corroborates the reliability of our estimation. Pharmacies, d, i, j, and k are located along the railroads (Seibu Ikebukuro line and Seibu Shinjuku line) running parallel from the heart of Tokyo and interestingly, the estimates are almost the same (4.4 and 4.6 km/day). Similarly, the sites (a, b, c, e, f, and g) which are along the same railroad (Utsunomiya line) are characterized by the comparable infection speeds (3.5, 3.1, and 1.6 km/day).

The infection speeds calculated with a pharmacy other than k as an infection focus are inconsistent. For example, the propagation speed between pharmacies a and n (= 20.7 km/day) is much higher than the average (= 3.5 km/day), suggesting that these

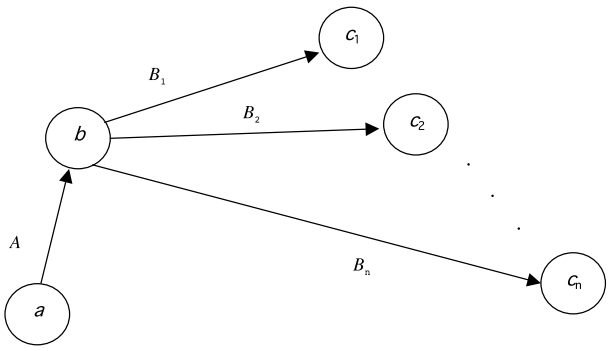

Fig. 3. Time Lags, A, B_1, ..., B_n, between Pairs of Sites, a, b, c_1, ..., c_n

sites are both far away from an infection focus.

The infection routes (arrows of Fig. 2) appear to be independent of each other except the routes to pharmacy a and to h. Between pharmacies a and d, the propagation speed is calculated to be 9.3 km/day; d and m, 169; m and n, 38.5; n and l, 10.1; l and h, 146; h and a, 4.4 km/day. These speeds across the routes of Fig. 2 are much higher than the average (3.5 km/day). Therefore, the disease propagation across the proposed routes is unlikely. The infection route to site a is indistinguishable from the route to site h, although they are based on different transport systems.

We stress here that even a small-scale network of pharmacies can provide the useful information about the route and speed of disease propagation in society. There remain a variety of things to be tackled, since the occurrence of disease is not just a matter of individuals but is the result of complex interactions among multifactorial etiologies such as humans, society, environment, medicine, etc. A large number of information sources including pharmacies, hospitals, clinics, chain drugstores and convenience stores will play an important role in constructing nation-wide vigilance of people's health.

In appendix, it shows that the difference between the mean values of Table 1 gives the actual time lag between two sites [*e.g.*, the lag between pharmacies a and k is that 7.2 – (–4.9) = 12.1]. We discuss the reason why the diagonal elements of the matrix (= 0, see Table 1) should be included in the mean of the time lags. An alternative is to ignore the diagonal elements, but this average is not useful.

The number of sites is $n + 2$: a, b, c_1, ..., c_n (see Fig. 3). It is assumed that the time lag (A, B_1, ..., B_n) between two sites does not depend on the route from one site to the other. For example, the time lag of the direct route from a to c_1 is equal to the time lag *via* b ($= A + B_1$). We consider the average of all the time lags from a site to the others, while the time lag between the same sites is included as zero. The mean at cite a is:

$$\begin{aligned} L_a &= \frac{0 + A + (A + B_1) + (A + B_2) \cdots + (A + B_n)}{n + 2} \\ &= \frac{(n + 1)A + B_1 + B_2 \cdots + B_n}{n + 2} \end{aligned}$$

The first term of the denominator, 0, means the time lag between a and a. The mean at cite b is:

$$\begin{aligned} L_b &= \frac{0 - A + B_1 + B_2 + \cdots + B_n}{n + 2} \\ &= \frac{-A + B_1 + B_2 \cdots + B_n}{n + 2} \end{aligned}$$

The difference between the mean lags is:

$$L_a - L_b = A$$

The denominator of the mean lags (L_a and L_b) should be $n + 2$ in order that the difference ($L_a - L_b$) describes the time lag (A). If the time lag between a and a does not participate in the average, the total number of participants is $n + 1$ and then $L_a - L_b = (n + 2)A/(n + 1)$, which makes no sense.

REFERENCES

1) Glass, G. E. (2000) Update: Spatial aspects of epidemiology: The interface with medical geography. *Epidemiol. Rev.*, **22**, 136–139.
2) Moore, D. A. and Carpenter, T. E. (1999) Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol. Rev.*, **21**, 143–161.
3) Torok, T. J., Kilgore, P. E., Clarke, M. J., Holman, R. C., Bresee, J. S. and Glass, R. I. (1997) Visualizing geographic and temporal trends in rotavirus activity in the United States, 1991 to 1996. *Pediatr. Infect. Dis. J.*, **16**, 941–946.
4) Suzuki, H., Sakai, T., Saito, R., Komata, O. and Sato, I. (2002) Use of geographic information systems for determining the epidemiology of influenza virus infections. *Kagaku Ryoho no Ryoiki*, **18**, 1801–1807.
5) Carrat, F. and Valleron, A. J. (1992) Epidemiologic mapping using the "Kriging" method: Application to an influenza-like illness epidemic in France. *Am. J. Epidemiol.*, **135**, 1293–1300.
6) Sakai, T., Suzuki, H., Sasaki, A., Saito, R., Tanabe, N. and Taniguchi, K. (2004) Geographic and temporal trends in influenzalike illness, Japan, 1992–

1999. *Emerg. Infect. Dis.*, **10**, 1822–1826.
7) Ferguson, N. M., Donnelly C. A. and Anderson R. M. (2001) The foot-and mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science*, **292**, 1155–1160.
8) May, R. (2000) Simple rules with complex dynamics. *Science*, **28**, 601–602.
9) Lloyd, A. L. and May, R. M. (2001) How viruses spread among computers and people. *Science*, **292**, 1316–1317.
10) Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of "samll-world" networks. *Nature* (London), **393**, 440–442.
11) Morens, D. (2001) Certain diseases, uncertain explanations. *Science*, **294**, 1658–1659.
12) Medley, G. F. (2001) Predicting the unpredictable. *Science*, **294**, 1663–1664.
13) Ijuin, K., Matsuda, R. and Hayashi, Y. (2006) A method for estimating infection route and speed of influenza. *Yakugaku Zasshi*, **126**, 161–165.
14) Ijuin, K., Matsuda, R. and Hayashi, Y. (2006) A method for estimating the order of influenza infection between adults and children. *Yakugaku Zasshi*, **126**, 311–314.
15) Takahashi, M., Kobari, T., Ijuin, K., Iwaki, K., Ishii, F., Matsuda, R., Hayashi, Y. and Yajima, T. (2006) Smoothing of correlation functions in the estimation of the order of influenza infection between adults and children. *J. Health Sci.*, **52**, 436–442.
16) Hino, M. (1982) *Spectral Analysis (Supekutoru Kaiseki)*, Asakura Shoten, Tokyo.