

Smoothing of Correlation Functions in the Estimation of the Order of Influenza Infection between Adults and Children

Mizuho Takahashi,^a Tsuyoshi Kobari,^b Kazushige Ijuin,^c Kazuo Iwaki,^d Fumiyoshi Ishii,^e Rieko Matsuda,^f Yuzuru Hayashi,^{*,f} and Takehiko Yajima^g

^aNikko Citizen's Hospital, 1752–10 Arasawa, Kiyotaki, Nikko, Tochigi 321–1441, Japan, ^bKosumo Chouzai Yakkyoku Co., Ltd., 2–129–3 Miyahara, Kita, Saitama, Saitama 331–0812, Japan, ^cTanashi Yakuhin Co., Ltd., 4–25–5, Tanashi, Nishi-Tokyo, Tokyo 188–0011, Japan, ^dSchool of Pharmaceutical Sciences, Ohu University, 31–1 Misumido, Tomita-machi Koriyama, Fukushima 963–8611, Japan, ^eDepartment of Pharmaceutical Sciences, Meiji Pharmaceutical University, Noshio 2–522–1, Kiyose, Tokyo 204–8588, Japan, ^fNational Institute of Health Sciences, 1–18–1 Kami-Yoga, Setagaya, Tokyo 158–8501, Japan, and ^gFaculty of Pharmaceutical Sciences, Toho University, Miyama 2–2–1, Funabashi, Chiba 274–8510, Japan

(Received February 17, 2006; Accepted April 19, 2006; Published online April 20, 2006)

The cross-correlation functions between the time series of drug sales at community pharmacies have been used for the epidemiological understanding of influenza propagation process. The aim of this paper is to examine the methods of smoothing for determining the maximum position (time lag) of the cross-correlation function. The moving average method with a window of seven days is taken to eliminate the hebdomadal frequencies originating from the life style of people around the pharmacies, or rather, purchasers. Based on the smoothed data of influenza anti-viral agents for adults and children at pharmacies in Tokyo and its vicinity, this paper tries to give an answer to the question of which are earlier infected with influenza, adults or children, in the limited areas.

Key words — correlation, pharmacy, influenza, spectral analysis

INTRODUCTION

The spatial understanding of disease propagation process in society is the goal shared by epidemiology, medical geography and health research.^{1,2)} The focus of research is on person, time and space and the analytical tools include choropleth maps and geographic information systems.^{1–6)} The techniques of spectral analysis has also been used for common purposes, *e.g.*, power spectrum, auto-correlation function and cross-correlation function.^{7–14)}

The cross-correlation functions are concerned with the time lag of similar or correlated phenomena which occur in distant places or under different situations.¹⁵⁾ The time lag is the maximum position of the cross-correlation function and means that the correlation coefficient between the phenomena is the largest, if one of the phenomena is shifted by the time lag. As an example, the daily variations of the sales of an influenza anti-viral agent at distant pharmacies are observed to be similar in pattern, but one variation lags a number of days behind the other. In this situation, the cross-correlation function between the daily variations is available for assessing the propagation route and speed of influenza in the district.¹³⁾ The infection order of adults and children in limited areas can also be estimated from the sales pattern of influenza anti-viral agents at the same pharmacy.¹⁴⁾ This is feasible, because the formulations are different from adults (capsule) to children (dry syrup).

In the above-mentioned study, the interpretation of the cross-correlation functions is straightforward. Unfortunately, however, the determination of the time lags can be obscured by the hebdomadal frequencies which originate from the life style of people in recent human society.^{13,14)} For example, the prescription of the influenza anti-viral agent for children was observed to concentrate on Sundays and Mondays at a pharmacy, even if it is open throughout the year.¹²⁾

The purpose of the present paper is to examine a smoothing method (moving average method) for eliminating the hebdomadal frequencies from the cross-correlation functions of the pharmacy data. A case study is given including fifteen situations of distant pharmacies and different influenza seasons.

MATERIALS AND METHODS

The information about the prescriptions of in-

*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; Fax: +81-3-3707-6950; E-mail: fumi@nihs.go.jp

Table 1. Time Lag of Sales of Tamiflu® Capsules and Tamiflu® Dry Syrup Resulting from Smoothed Data

Pharmacy No.	Location	Time lag		
		Smoothing		
		None	CF ^{a)}	TS ^{b)}
Period: 11/01/2002–10/31/2003				
1	Tanashi, Nishi-Tokyo, Tokyo	0	11	11
2	Tanashi, Nishi-Tokyo, Tokyo	13	16	15
Period: 11/01/2003–10/31/2004				
3	Shinmachi, Hanno, Saitama	0	2	1
4	Tanashi, Nishi-Tokyo, Tokyo	7	4	6
5	Tanashi, Nishi-Tokyo, Tokyo	3	3	2
6	Kohoku, Yokohama, Kanagawa	7	6	7
7	Nakagawa, Adachi, Tokyo	−7	1	0
8	Tana, Sagamihara, Kanagawa	0	1	2
9	Asakamachi, Koriyama, Fukushima	0	−2	−1
Period: 11/01/2004–10/31/2005				
10	Shinmachi, Hanno, Saitama	0	−11	−11
11	Tanashi, Nishi-Tokyo, Tokyo	0	−2	−2
12	Tanashi, Nishi-Tokyo, Tokyo	5	5	4
13	Kohoku, Yokohama, Kanagawa	−2	1	−2
14	Nakagawa, Adachi, Tokyo	0	−5	−6
15	Tana, Sagamihara, Kanagawa	0	−3	0

a) correlation function. b) time series.

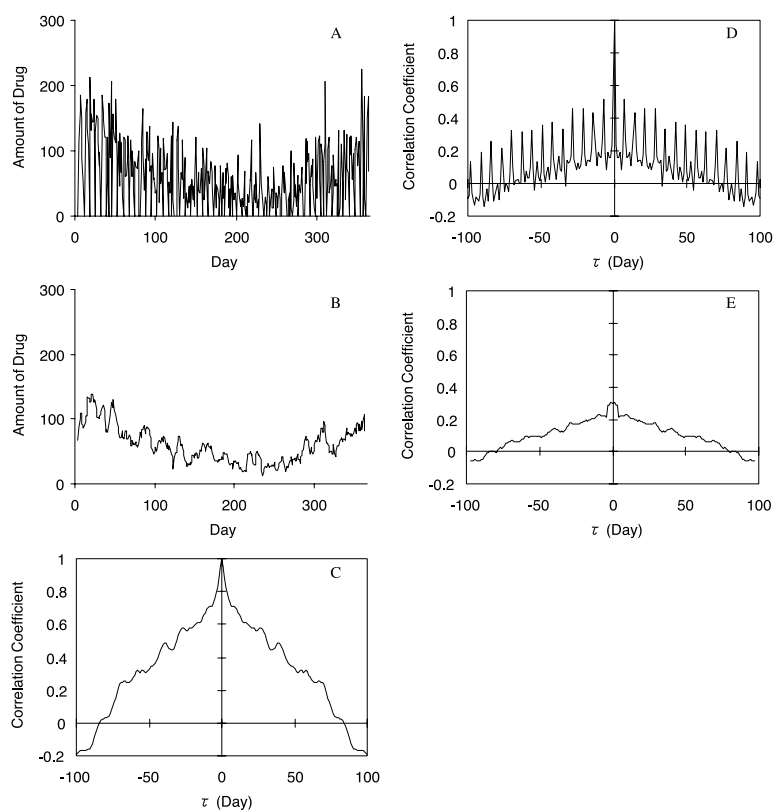


Fig. 1. Time Series, Auto-Correlation Functions (Weak Auto-Correlation) and Smoothed Data for the Sales of Neo-Amunoll® Powder at a Pharmacy

The period of the time series ranges from January 1st in 2004 to December 31st in 2004. The pharmacy is located in 2–129–3 Miyahara, Kita, Saitama, Saitama. The Y axes of A and B denote the weight. The meaning of figures A–E is the same as those of Fig. 4.

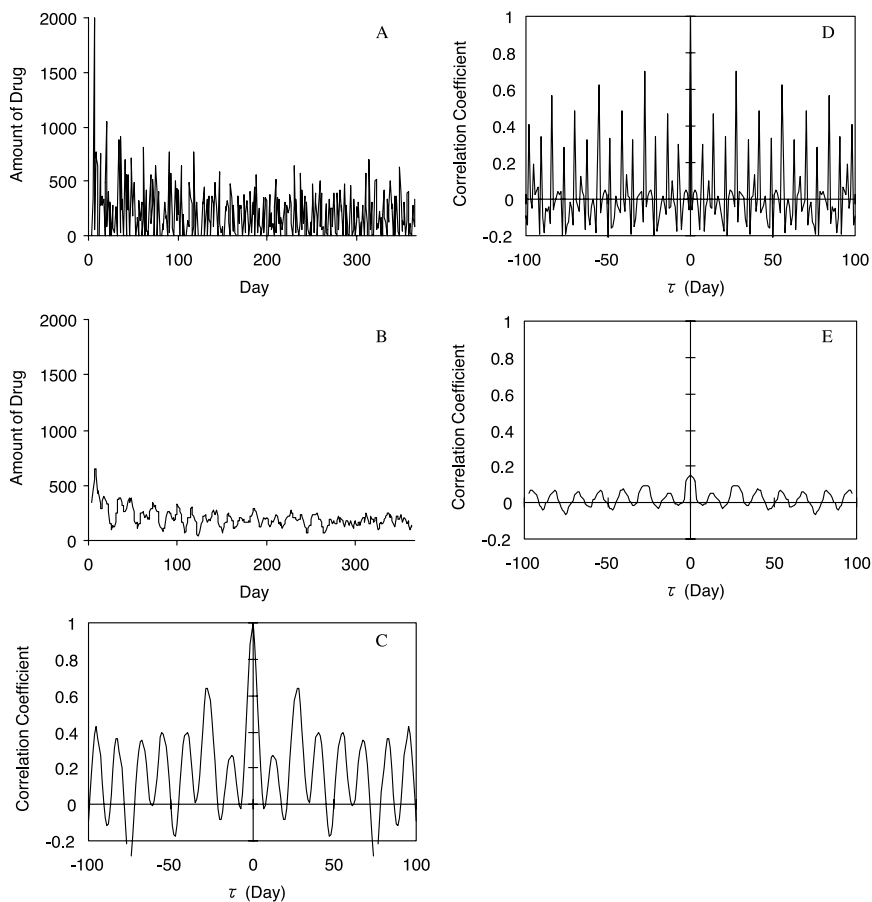


Fig. 2. Time Series, Auto-Correlation Functions (No Auto-Correlation) and Smoothed Data for the Sales of Daonil® at a Pharmacy

The period of the time series ranges from January 1st in 2004 to December 31st in 2004. The location of the pharmacy is described in the legend of Fig. 1. The Y axes of A and B denote the number of tablets. The meaning of figures A–E is the same as those of Fig. 4.

fluenza drugs was collected from the pharmacies listed in Table 1. Pharmacy (No. 1, 4, 11) is located near an emergency hospital, the pharmacy for Figs. 1 and 2 is near a hospital of geriatrics (but not emergency one) and the other pharmacies are near general clinics (but not emergency ones).

The cross-correlation function, $R(\tau)$, between time series, $A(t)$ and $B(t)$, describes the correlation coefficients between $A(t)$ and $B(t + \tau)$ as a function of τ .¹⁵⁾

$$R(\tau) = \frac{E[A(t)B(t + \tau)]}{\sqrt{E[A(t)^2]E[B(t + \tau)^2]}} \quad (1)$$

where $E[.]$ denotes the mean over time, t , of the random variables inside the square brackets and parameter τ is called lag. In this paper, if $B(t)$ lags behind $A(t)$, $R(\tau)$ has the maximum where $\tau > 0$. If not, $R(\tau)$ has the maximum where $\tau \leq 0$.

The moving average method¹⁶⁾ used here has a window of seven days and the value at the central point of the window is replaced by the calculated

average over the window. A strong hebdomadal frequency and its harmonics have been observed in the power spectral densities of drug sales time series.¹²⁾ Since a wide window such as 14 and 21 days is known to widen a peak, the window in this paper is fixed at 7 days.

Figure 3 shows two types of smoothing examined here. The difference is in the data to which the smoothing is applied: time series or cross-correlation function. The smoothing of the time series is common, but that of the cross-correlation function might seem to be unreasonable. As is mentioned below, a problem is found with the former, and the alternative is also taken.

The influences of the moving average method are examined first on auto-correlation functions and second on cross-correlation functions. Finally, a case study is presented. The daily variation in the amount of drug sale at a pharmacy can be classified into three groups according to the strength of auto-correlation; strong, weak and none.¹²⁾ Figures 1, 2 and 4 give the

time series of weak, none and strong correlation, respectively.

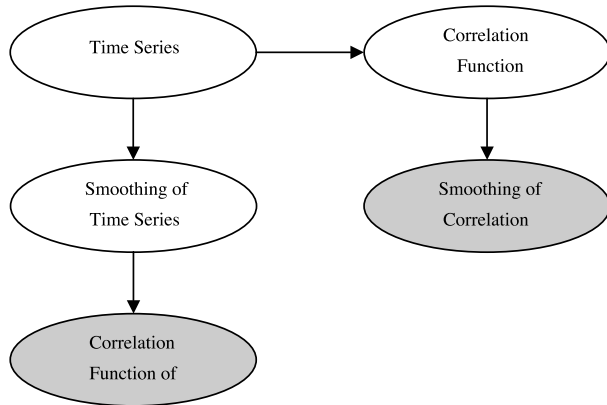


Fig. 3. Schematic Diagram of Two Smoothing Methods
The final data used for the analysis are indicated by the gray background.

RESULTS AND DISCUSSION

Figure 4A shows the daily variation in the sales of an influenza anti-viral agent (Tamiflu® Capsule 75, Chugai Pharmaceutical) at a community pharmacy and Fig. 4C and 4E show its auto-correlation functions made through the smoothing. The smoothing itself is applied to either the time series or auto-correlation function (for the smoothed data, see B and E, respectively). The original auto-correlation function without smoothing is shown in Fig. 4D. The arrangement of the figures is the same as that of Fig. 3.

In general, the auto-correlation function takes unity when $\tau = 0$ and approaches to zero when $\tau = \infty$.¹⁵⁾ The intact auto-correlation function (D) has these general properties. The smoothing (E) of the correlation function breaks the first property (unity when $\tau = 0$), but provides the reliable values when $\tau \neq 0$ (compare D with E). On the other hand, the first property is kept by the auto-correlation function (C) of the smoothed time series, but the correlation co-

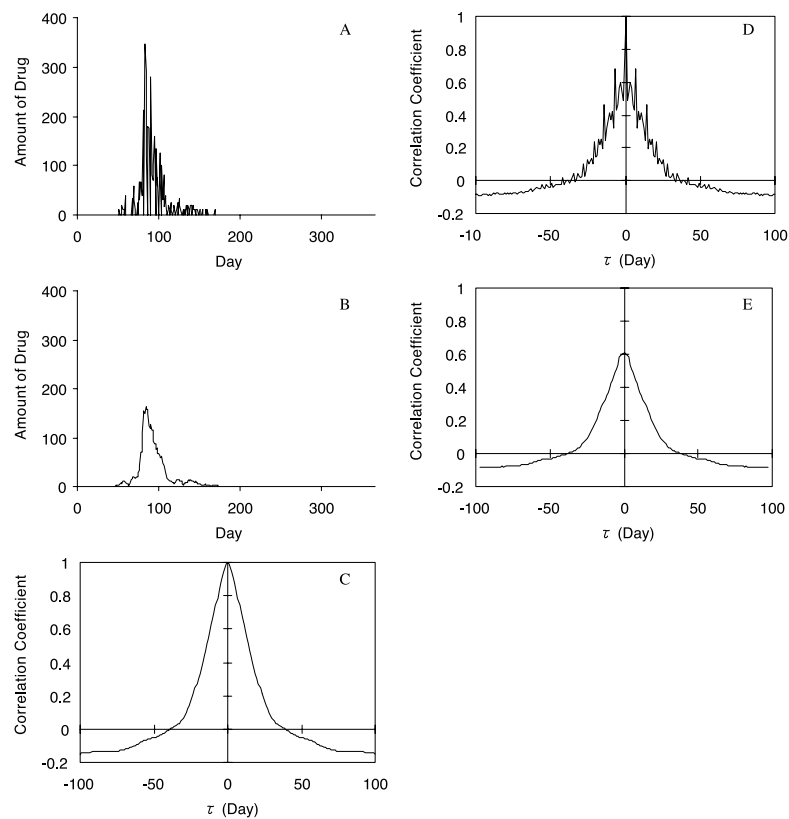


Fig. 4. Time Series, Auto-Correlation Functions (Strong Atuo-Correlation) and Smoothed Data for the Sales of Tamiflu® Capsule 75 at a Pharmacy

For the location and period, see No. 8 of Table 1. A: the daily variation in the number of capsules supplied; B: the smoothed data of A; C: the auto-correlation function of the smoothed time series (B); D: the auto-correlation function of the original time series (A); E: the smoothed data of the intact auto-correlation function (D).

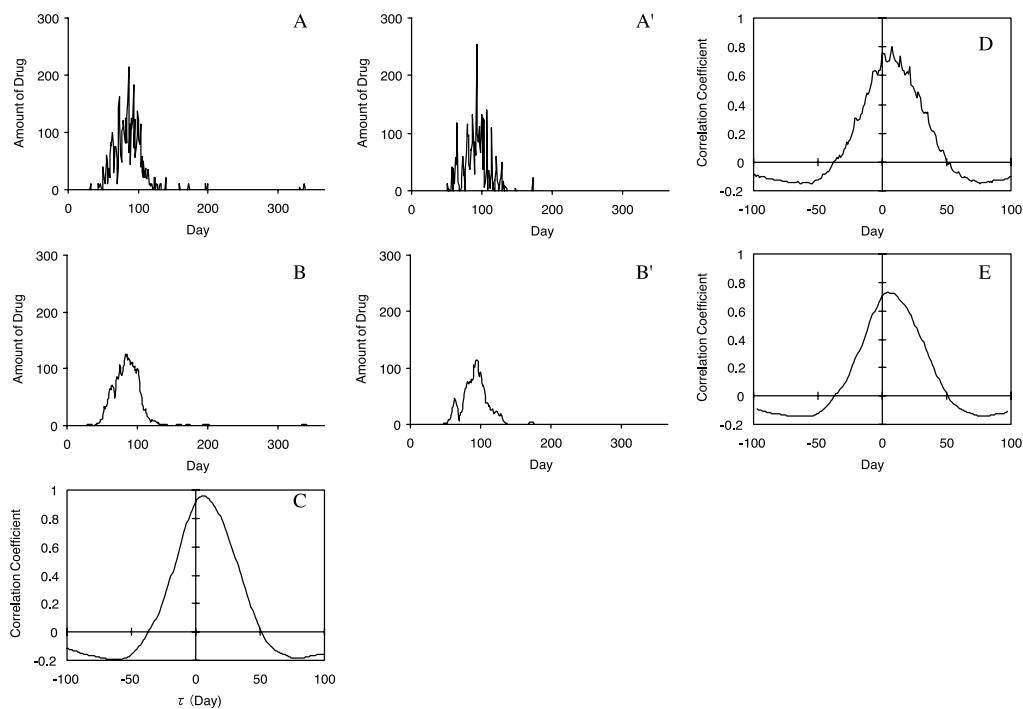


Fig. 5. Time Series, Cross-Correlation Functions and Smoothed Data for the Sales of Tamiflu® Capsules and Tamiflu® Dry Syrup at a Pharmacy

For the location and period, see No. 4 of Table 1. A: the daily variation in the number of capsules supplied; A': the daily variation in the weight of dry syrup supplied; B: the smoothed data of A; B': the smoothed data of A'; C: the cross-correlation function of the smoothed time series (B and B'); D: the cross-correlation function of the original time series (A and A'); E: the smoothed data of the intact cross-correlation function (D).

efficients are highly enhanced especially around $\tau = 0$ (compare C with D).

Figure 1 shows the time series (A) of a common cold drug (Neo-Amunoll® Powder, Sanwa Kagaku Kenkyusho) and its auto-correlation functions with the smoothing carried out before and after the auto-correlation (C and E). Although the shape of the auto-correlation function (D) is different from that of Fig. 4, the effects of the smoothing are similar.

Figure 2 takes an antidiabetic agent (Daonil®, Aventis Pharma) as an example of no auto-correlation. As a whole, the smoothing effects are similar to the above examples. However, the hebdomadal frequencies look more conspicuous than necessary (C).

Figure 5 shows the application of the moving average method to the infection order of influenza between adults and children in the limited area. The intact cross-correlation function of the time series of influenza anti-viral agents for adults (Tamiflu® Capsule 75) and children (Tamiflu® Dry Syrup 3%, Chugai Pharmaceutical) at a pharmacy provides a time lag of 7 days (D). As for the drug supply, the children lag 7 days behind the adults there. However, this time lag can be obscured by the hebdo-

mal frequencies occurring in the time series (A and A'). If the cross-correlation function is calculated from the smoothed time series, the time lag is observed to be 6 days (C). On the other hand, if the smoothing is applied to the cross-correlation function (D), the smoothed function (E) gives the time lag of 4 days (also see No. 4 of Table 1).

The above methods of smoothing are compared further at different pharmacies and in different influenza seasons (Table 1). Seven community pharmacies and three seasons are included in the analysis. The time lags obtained from the smoothing applied to the time series or cross-correlation function are in good agreement with each other in every situation. They are almost consistent with the results from the cross-correlation functions without smoothing, although three are four exceptions (No. 1, 7, 10 and 14).

The exceptions are helpful to examine the effects of the smoothing (see Fig. 6). In the intact cross-correlation functions, the largest correlation coefficients (at $\tau = 0$ for A, -7 for B, 0 for C and 0 for D) can be considered to come from the hebdomadal cycles. These accidentally largest coefficients seem to be successfully corrected by the moving average

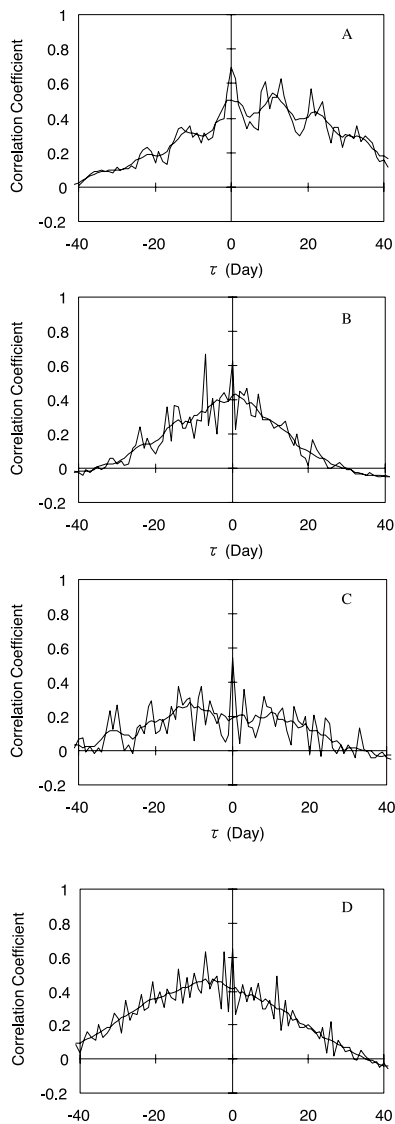


Fig. 6. Intact and Smoothed Cross-Correlation Functions for the Sales of Tamiflu® Capsules and Tamiflu® Dry Syrup at Pharmacies

Fine lines: intact cross-correlation functions; thick lines: smoothed cross-correlation functions. The pharmacies and periods are described in Table 1: A: No. 1; B: No. 7; C: No. 10; D: No. 14. The smoothing is applied to the intact correlation functions.

method.

Hereafter, the order of influenza infection between adults and children is considered with reference to the smoothed data. In most cases of Table 1 (see the columns of CF and TS), the infection of the adults precede that of the children. By definition, this order is derived from the positive signs of time lags (see MATERIALS AND METHODS). At the pharmacy located at Hanno (No. 10), however, the extremely reverse order is observed ($\tau = -11$ days). In this case, the largest correlation coefficient is 0.25

for the smoothed cross-correlation function (see Fig. 6C). This value ($= 0.25$) is too small to be significant and the result of the pharmacy will be ignored in the discussion below.

As for the above pharmacy at Hanno (Fig. 6C), the largest correlation coefficient obtained from the smoothing of the time series is found at the same lag ($\tau = -11$ days, see Table 1), but its value is much larger ($= 0.79$, not shown) than that of the other smoothing (≈ 0.25 , see Fig. 6C). Therefore, the only use of the smoothing of the time series is not adequate, and the smoothing of the correlation function should also be referred to.

Most cases examined here indicate the earlier infection of adults than children. However, the order of influenza infection is not an easy question. At the pharmacies at Hanno and Tanashi, the earlier infection of the adults is observed in the season of 2003–2004 (No. 3 and 4 of Table 1), but the reverse results are obtained in the other season of 2004–2005 (No. 10 and 11). The nearby pharmacies which are about 100-meter away do not necessarily yield the same order of infection (No. 4, 5, 11 and 12).

In conclusion, the different methods of smoothing have been examined for determining the time lag (the position of the maximum) of the cross-correlation function of the time series of drug sales at community pharmacies. If the smoothing is applied to the cross-correlation function itself, the result is satisfactory as a whole. However, the smoothing of the time series before the cross-correlation function gives an enhanced correlation coefficient. Either method is useful for the determination of the time lag. However, if the value of the correlation coefficient is important, the smoothing of the correlation function is preferable.

At many pharmacies located in Tokyo and its vicinity, the earlier infection of adults than that of children was observed for different seasons of infection, although there are exceptions. The longer action radius of the adults will be one of the reasons for this result. The infection order of influenza between the adults or children is not an easy question, but the amount of pharmacy information used in this study is not sufficient. A large-scale network of community and hospital pharmacies is essential for the further study on the time-and-space pattern of influenza propagation in human society.

Acknowledgements The authors would like to acknowledge Dr. Takeuchi, Hisako at Kamome Yakkyoku, Kitasato Kenkoukan (Triad Japan Co.,

Ltd.) and Mr. Yamaguchi, Jin (Yamaguchi Yakuhin Co., Ltd.) for their offering the pharmacy data.

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) Suzuki, H., Saito, R., Shoubugawa, Y. and Sakai, T. (2005) Epidemiology of influenza. *Virus Report*, **2**, 81–87.
- 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 Ryouiki*, **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., Naohito, T. and Taniguchi, K. (2004) Geographic and temporal trends in influenzalike illness, Japan, 1992–1999. *Emerg. Infect. Dis.*, **10**, 1822–1826.
- 7) Sumi, A., Ohtomo, N., Tanaka, Y., Sawamura, S., Olsen, L. F. and Kobayashi, N. (2003) Prediction analysis for measles epidemics. *Jpn. J. Appl. Phys.*, **42**, 7611–7620.
- 8) Sugihara, G. and May, R. M. (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* (London), **344**, 734–741.
- 9) Jose, M. V. and Bishop, R. F. (2003) Scaling properties and symmetrical patterns in the epidemiology of rotavirus infection. *Philos. Trans. R. Soc. Lond., B Biol. Sci.*, **358**, 1625–1641.
- 10) Segal, A. O., Crighton, E. J., Moineddin, R., Mamdani, M. and Upshur, R. E. (2005) Croup hospitalizations in Ontario: A 14-year time-series analysis. *Pediatrics*, **116**, 51–55.
- 11) Sumi, A., Olsen, L. F., Ohtomo, N., Tanaka, Y. and Sawamura, S. (2003) Spectral study of measles epidemics: The dependence of spectral gradient on the population size of the community. *Jpn. J. Appl. Phys.*, **42**, 721–733.
- 12) Ijuin, K., Hatanaka, N., Segawa, K., Nakano, T., Nakata, K., Tohara, A., Sato, M. and Hayashi, Y. (2006) Stochastic properties of prescriptions for infectious and non-infectious diseases dispensed by a city pharmacy. *Jpn. J. Pharm. Health Care Sci.*, **32**, 51–54.
- 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) Hino, M. (1982) *Spectral Analysis (Supekutoru Kaiseki)*, Asakura Shoten, Tokyo.
- 16) Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y. and Kaufman, L. (1988) *Chemometrics: a textbook*, Elsevier Science Publishing Company, New York.