

# Short Term Prediction of Infectious Diseases Patients from Prescription Surveillance

Tamie Sugawara<sup>1</sup>, Yasushi Ohkusa<sup>1\*</sup>, Hirokazu Kawanohara<sup>2</sup>, Miwako Kamei<sup>3</sup>

<sup>1</sup>National Institute of Infectious Diseases, Tokyo, Japan

<sup>2</sup>EM Systems Co. Ltd., Osaka, Japan

<sup>3</sup>School of Pharmacy, Nihon University, Chiba, Japan

Email: \*ohkusa@nih.go.jp

**How to cite this paper:** Sugawara, T., Ohkusa, Y., Kawanohara, H. and Kamei, M. (2018) Short Term Prediction of Infectious Diseases Patients from Prescription Surveillance. *Journal of Biosciences and Medicines*, 6, 61-68.

<https://doi.org/10.4236/jbm.2018.69007>

**Received:** August 22, 2018

**Accepted:** September 22, 2018

**Published:** September 25, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In Japan, the incidence of almost all common pediatric infectious diseases has been monitored, with 7 - 10 day delays, at medical institutions through the National Official Sentinel Surveillance of Infectious Diseases (NOSSID). On the other hand, based on prescriptions filled at external pharmacies, Prescription Surveillance (PS) collects information and provides estimated numbers of influenza, varicella, and gastroenteritis infectious (GI) patients to the public the following morning. For precise, and real-time estimation of incidences of common pediatric infectious diseases, we evaluated predictive power of PS for diseases other than influenza, varicella, and GI. Results demonstrated that PS information has sufficient predictive power for pharyngoconjunctival fever, group A streptococcal pharyngitis, exanthem subitum, and mumps, some predictive power for RS virus infection, erythema infectiosum and herpangina, but insufficient predictive power for hand, foot and mouth disease.

## Keywords

Prescription Surveillance, Pharmacy, Sentinel Surveillance, Pediatric Infectious Disease

## 1. Introduction

In Japan, the incidence of infectious disease has been monitored at medical institutions through official national surveillance of infectious diseases based on the Law for the Prevention of Infectious Diseases and Medical Care for Patients of Infections by Ministry of Health, Labour and Welfare (MHLW). Hereinafter, we designate this surveillance system as NOSSID. Cases of almost all common pediatric infectious diseases in NOSSID have been reported from only 3000 sen-

tinel medical institutions, collectively accounting for only about one-tenth of all pediatric medical institutions. Moreover, because the number of patients with each infectious disease per sentinel in the last week is published once a week (every Friday at noon), this report has been published with a 7 - 10 day delay. The sentinel medical institutions for these diseases have allocated pediatric hospitals and clinics. Therefore, these reports might not include information for adults and elderly patients or might heavily underreport that information.

Several studies have examined surveillance methods to detect infectious diseases early, without waiting for a diagnosis by a doctor. Prescription Surveillance (PS) [1] [2] [3] has been operated by the Japan Medical Association, Japan Pharmaceutical Association, School of Pharmacy, Nihon University, and EM Systems Co. Ltd. It reports the estimated number of patients based on information of prescriptions filled at external pharmacies. In December 2014, approximately 9950 pharmacies participated, collectively accounting for about 19.6% of all pharmacies.

This system mainly monitors prescriptions from hospital or clinics which use Application Service Provider for medical claim in pharmacies through safe internet connection. The collected data represents the number of prescriptions made, classified into therapeutic categories as described below. Therefore, the collected data include no personal information. Data related to the number of prescriptions are extracted automatically and analyzed daily.

In this system, the numbers of patients were estimated from the number of prescriptions for neuraminidase inhibitors, anti-varicella-herpes-zoster virus (VZV) drugs, antibiotic drugs, antipyretic analgesics, and multi-ingredient cold medications by prefecture each day. The numbers of patients with neuraminidase inhibitors or anti-VZV drugs were classified by three age groups: children younger than 15 years old, and adult younger than 64 years old, and elderly people over 65 years old. Moreover, antibiotics were classified into five types: penicillin, cephem, macrolide, new quinolone, and others [4]. These drugs were chosen to identify clusters of rash, fever, or digestive symptoms to detect bioterrorism attack, emerging/re-emerging diseases, or mass food poisoning. Particularly, because anti-VZV drug is a drug for varicella and zoster, a cluster of this drug in adult with no cluster in children or elderly people represents a signal of small pox [2]. The following morning, these data are presented on a web page (<http://prescription.orca.med.or.jp/syndromic/kanjyasuikei/>) the following morning.

To overcome the shortcomings of NOSSID described previously, *i.e.* delay and underreporting, we checked the predictive power of PS for NOSSID. If it is confirmed as having sufficient power, then we can suggest use PS to recognize real-time situations. Especially, this study assesses the predictive capability of the incidence of common pediatric infectious diseases other than influenza, varicella, and gastroenteritis infectious (GI) from PS. Influenza, varicella, and GI were excluded because the numbers of patients with influenza, varicella, and GI had

been predicted precisely in real-time by PS; part of this information is announced to the public through the web [1] [2] [3] [5] [6]. If one can confirm the precision of the prediction, then the estimated number of patients from PS might be useful for providing real-time and precise information about the incidence of infectious diseases other than influenza, varicella, and GI. We examined the incidence of RS virus infection (RS), pharyngoconjunctival fever (PCF), group A streptococcal pharyngitis (A-SP), hand, foot and mouth disease (HFMD), erythema infectiosum (EI), exanthem subitum (ES), herpangina, and mumps. Even if no specific drug for a disease or one-to-one correspondence exists between a drug and disease, we examined the potential feasibility of the information in PS that predicted the number of patients with a type of infectious disease.

## 2. Materials and Methods

### 2.1. Data

PS estimates the numbers of patients by multiplying the reciprocal of the participation rate of pharmacies to PS, and the reciprocal of the proportion of external prescription in the prefecture to the total number of prescriptions in the prefecture. The numbers of patients who received neuraminidase inhibitors, anti-VZV drugs, antibiotic drugs, antipyretic analgesics, multi-ingredient cold medications, and antidiarrheal and intestinal drugs at an external pharmacy were recorded.

NOSSID provides the numbers of patients per sentinel per week as the incidence of each disease considered in this paper, except for RS. For RS, NOSSID provides only the total number of patients per week. This information has been published officially on every Friday noon, reflecting the situation of the prior week. For the procedures explained below, we use the latest information of NOSSID since Saturday. Therefore, on Saturday and Sunday, we can use the information of NOSSID for the prior week or of earlier weeks.

### 2.2. Analysis

We first examined the relation among the incidences from NOSSID and the information from PS. The incidences from NOSSID were regressed on information related to in-sample data. Specifically, we used weighted least-squares method, a commonly used method for multiple regressions, weighted by the incidence, as

$$d_{i,w(t)} = \alpha + \sum_{j=1}^{10} \beta_{i,j} x_{j,t} + \sum_{l=2}^{53} \delta_{i,l} y_{l(t)} + d_{i,w(t)-k(t)} + \zeta_i h_t + \psi_i n_t + \varepsilon_{i,t} \quad (i = 1, 2, \dots, 11) \quad (1)$$

where  $d_{i,w(t)}$  denotes the reported number of patients with infectious disease  $i$  (RS, PCF, A-SP, HFMD, EI, ES, herpangina, mumps, influenza, varicella, and GI) per sentinel for a week  $w$ , which includes day  $t$ , and where  $x_{j,t}$  denotes the number of patients who were prescribed drug  $j$  (neuraminidase inhibitors, anti-VZV drugs, penicillin, cephem, macrolide, new quinolone, other antibio-

tics, antipyretic analgesics, multi-ingredient cold medications, and anti-diarrheal/intestinal drugs) on day  $t$ . In this equation,  $y_{w(t)}$ ,  $h_p$  and  $n_t$  denote the dummy variables of epidemiological week, holiday, and a day following a holiday. Actually,  $k(t)$  is a function of day of the week representing a lag of NOSSID. We defined  $k(t) = 1$  for Saturday and Sunday, and  $k(t) = 2$  otherwise. We evaluated the contribution of information from PS using the  $F$  test under the null hypothesis that all coefficients of  $x_{j,t}$  are zero, which means no contribution of PS in Equation (1). We adopted 5% as significance level.

The prediction of influenza, varicella, and GI was confirmed previously [1] [2] [3] [5] [6]. Therefore, we used the prediction precision for these diseases as the benchmark to evaluate the other considered disease precision.

We used correlation coefficients and discrepancy rates among predictions from PS information and NOSSID observations. The discrepancy rate of infectious disease  $i$ ,  $r_p$ , is defined as

$$r_i = \frac{|\hat{d}_{i,w(t)} - d_{i,w(t)}|}{\sum d_{i,w(t)}} \times 100 \quad (2)$$

where  $d_{i,w(t)}$  and  $\hat{d}_{i,w(t)}$  respectively represent the reported numbers of patients with infectious disease  $i$  on day  $t$  from NOSSID and its prediction from Equation (1) using available information at day  $t$ . They stand for the weighted average of the absolute value of prediction error weighted by the reported number from NOSSID.

The study period was March 16, 2011 through December 31, 2016. Especially, we verified the predictive power of this model in 2016 prospectively, which indicates the use of available data only at day  $t$  for estimation and prediction. Study area was the entirety of Japan.

### 2.3. Ethics

Data from PS were aggregated and de-linked from personal information related to patients, medical institutions, and pharmacies: these are anonymous data. The reported numbers of patients from NOSSID are published data. Therefore, no ethical issues are posed by using these data for this study.

## 3. Results

**Table 1** presents  $F$  statistics for the information from PS in regression analyses using the data obtained up through the end of 2015. From these regression analyses, all  $F$  statistics were found to be significant. Additionally, Table presents discrepancy rates and correlation coefficients among the reported numbers of patients from NOSSID and prediction from PS in 2016, prospectively operated: we used only available data in each day in 2016. In this evaluation, the only disease for which the discrepancy rate was lower than 10% was ES. The correlation coefficients of those diseases were 0.9088 - 0.9614. For the benchmark diseases, the discrepancy rates and correlation coefficients were 15.82% - 28.03% and

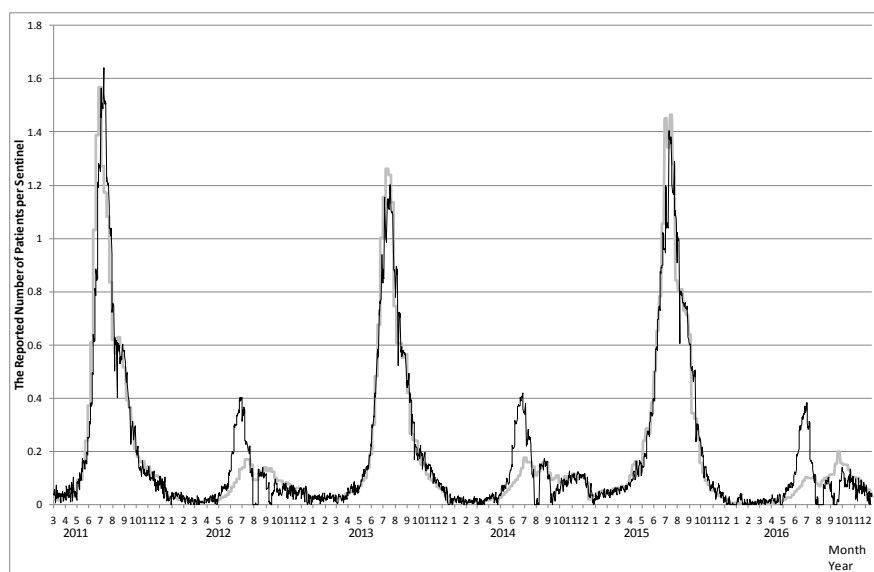
0.9230 - 0.9667.

**Figure 1** portrays the reported number of HFMD patients from NOSSID and the prediction from PS. Until 2016, the gray line represents the prediction from PS using Equation (1) with all information until the end of 2015. However, in 2016, the prediction was calculated prospectively, requiring the use only of available data each day. The two lines in Figure, even in 2016, were quite similar, although the discrepancy rate of HFMD was the worst in shown in Table.

**Table 1.** *F* statistics for the information from PS until 2016, with discrepancy rates and correlation coefficients for predictions and observations in 2016.

	<i>F</i> statistics	Discrepancy rate (%)	Correlation coefficient
RS virus infection	12.07	24.40	0.9403
Pharyngoconjunctival fever	17.79	12.11	0.9540
Group A streptococcal pharyngitis	16.63	10.84	0.9370
Hand, foot and mouth disease	3.29	84.04	0.9393
Erythema infectiosum	4.86	21.42	0.9088
Exanthem subitum	8.96	9.30	0.9278
Herpangina	5.87	15.94	0.9614
Mumps	4.15	10.68	0.9514
Influenza	105.48	23.67	0.9230
Varicella	19.85	28.03	0.9667
Gastroenteritis infectious	60.94	15.82	0.9432

Note: *F* test was applied to information obtained from PS using the data until the end of 2015. All *F* statistics are significant. The discrepancy rate and correlation coefficient were calculated using data only in 2016 with prospectively operated which means used only available data in each day in 2016.



**Figure 1.** Observation from NOSSID per sentinel per week and the prediction from PS: Example of hand, foot and mouth disease. Note: The black line represents the reported number from NOSSID. The gray line represents the prediction from PS by Equation (1) using all information until the end of 2015. In 2016, the prediction was calculated prospectively, which means to use only available data each day.

## 4. Discussion

This paper presented an examination and demonstrated that information from PS can completely predict the number of patients with common pediatric infectious diseases. The prediction results are obtained independently of prescription of a specific drug for a disease or one-to-one correspondence among drugs and diseases. Therefore, we cannot decompose the predictive power to some drugs. It might not be meaningful to measure the respective contributions of drugs of various types. Moreover, the estimated coefficient of each explanatory variable in multiple regression signifies that the marginal impact of the dependent variable controlled the effect of the other explanatory variables. Therefore, intuitively, it is apparently difficult to interpret these coefficients sometimes. However, result demonstrated that *F* statistics for the information from PS in regressions were significant. These results demonstrate that these regression equations were valid. Consequently, we inferred that each explanatory variable of the number of patients from PS was effective in these regressions.

Earlier studies verified the predictive power of PS for influenza, varicella, and GI [1] [2] [3] [5] [6]. For these diseases, most patients in Japan were respectively prescribed disease-specific drugs, such as neuraminidase inhibitors, anti-VZV drugs, and antidiarrheal and intestinal drugs. Therefore, we used the results of these diseases as the benchmark when evaluating the estimation and prediction results for the considered diseases.

This study was undertaken to verify the predictive power of PS for the considered diseases. Therefore, we mainly evaluated the discrepancy rates. The correlation coefficients cannot reflect consistent but parallel differences among two sets of data. Therefore, for prediction, the absolute level is important as well as the changing in variables. The discrepancy rates of the respective benchmark diseases were 23.67%, 28.03%, and 15.82%, respectively, for influenza, varicella, and GI. Therefore, we can infer that the information from PS has sufficient predictive power for these diseases if the discrepancy rate is lower than 15.82%. Conversely, if the discrepancy rate is higher than 28.03% for some disease, then we can infer that the information from PS has no predictive power. In a moderate case in which the discrepancy rate is higher than 15.82% but less than 28.03%, we inferred that PS has some predictive power, but also that its estimates are not so precise for these diseases.

The discrepancy rates for PCF, A-SP, ES, and the mumps were lower than 15.82%. The discrepancy rates of RS, EI, and herpangina were not lower than 15.82%, but lower than 28.03%. However, that of HFMD was higher than 28.03%. Therefore, we conclude that the information from PS has sufficient predictive power for PCF, A-SP, ES, and mumps. Unfortunately, the information from PS did not have sufficient predictive power for HFMD. The epidemiology of HFMD is well known to be quite irregular season-by-season. Even during the study period, high seasons occurred in 2011, 2013, and 2015, with intervening low seasons in 2012, 2014, and 2016. Therefore, prediction of the incidence of

HFMD appears to be much more difficult than that of other diseases considered herein.

In Figure, we showed only the case of HFMD. As shown in Table, the prediction power for HFMD was the worst, although the two lines in Figure were apparently very similar. The figures for other diseases than HFMD should show two more closed two lines. Consequently, it may be difficult to recognize the difference between the two lines: therefore, it might not be informative. For that reason, we omit figures for other diseases.

No similar study has been reported in the relevant literature because no nationwide real-time surveillance system is comparable in preciseness to national official surveillances in any countries. An exceptional study had been conducted in Taiwan [7], which used information from emergency department in nationwide and daily. However this Taiwan study did not compare it to national official surveillance and did not show prediction power in information from the emergency department.

This study had some limitations. First, we showed some prediction power of PS for NOSSID, but we confirmed it only at nationwide level. However we need information at a more local level, such as that of a prefecture or city. In general, prediction power in PS might decrease at the local level because the number of pharmacies participating in PS should be far smaller at a local level. Moreover, distribution of the number of pharmacies participated in PS probably reflects population distribution precisely. It might cause a variation in prediction power in PS by local area. Therefore, confirmation of prediction power in PS by prefecture or cities is anticipated as the next challenge.

Another limitation is that, although we proved some prediction power in PS in 7 - 10 days, we did not demonstrate the usefulness of such a short-term prediction. We must confirm its usefulness through operation of this system prospectively and by publishing it on a timely basis through a web site.

## 5. Conclusion

We examined the predictive power of PS information for common pediatric infectious diseases aside from influenza, varicella, and GI. Results demonstrated that the information from PS has sufficient predictive power for PCF, A-SP, ES, and mumps, with some predictive power for RS, EI, and herpangina. However, insufficient predictive power was shown for HFMD. For these diseases, we plan to provide incidence information to the public rapidly via the internet.

## Conflicts of Interest

The authors have no conflict of interest to declare.

## References

- [1] Sugawara, T., Ohkusa, Y., Ibuka, Y., Kawanojara, H., Taniguchi, K. and Okabe, N. (2012) Real-Time Prescription Surveillance and Its Application to Monitoring Sea-

sonal Influenza Activity in Japan. *Journal of Medical Internet Research*, **14**, e14. <https://doi.org/10.2196/jmir.1881>

- [2] Sugawara, T., Ohkusa, Y., Kawanohara, H., Taniguchi, K. and Okabe, N. (2011) Chickenpox Case Estimation in Acyclovir Pharmacy Survey and Early Bioterrorism Detection. *Kansenshogaku Zasshi*, **85**, 632-637. (In Japanese) <https://doi.org/10.11150/kansenshogakuzasshi.85.632>
- [3] Nakamura, Y., Sugawara, T., Kawanohara, H., Ohkusa, Y., Kamei, M. and Oishi, K. (2015) Evaluation of the Estimated Number of Influenza Patients from National Sentinel Surveillance Using National Database of Electronic Medical Claims. *Japanese Journal of Infectious Diseases*, **68**, 27-29. <https://doi.org/10.7883/yoken.JJID.2014.092>
- [4] Sugawara, T., Ohkusa, Y., Gu, Y., Kawanohara, H., Taniguchi, K. and Okabe, N. (2012) Estimation of Amount of Antimicrobials Used by Pharmacy Surveillance. *Japanese Journal of Infection Prevention and Control*, **27**, 195-198. <https://doi.org/10.4058/jsei.27.195>
- [5] Nakamura, Y., Kawanohara, H. and Kamei, M. (2015) Evaluation of the Number of Varicella Patients Estimated by Prescription Surveillance. *Kansenshogaku Zasshi*, **89**, 23-29. <https://doi.org/10.11150/kansenshogakuzasshi.89.23>
- [6] Nakamura, Y., Kawanohara, H. and Kamei, M. Association between Diarrheal Infectious Diseases and Antidiarrheal Drugs by Prescription Surveillance. *Kansenshogaku Zasshi*, in press.
- [7] Wu, T.S., Shih, F.Y., Yen, M.Y., Wu, J.S., Lu, S.W., Chang, K.C., Hsiung, C., Chou, J.H., Chu, Y.T., Chang, H., Chiu, C.H., Tsui, F.C., Wagner, M.M., Su, I.J. and King, C.C. (2008) Establishing a Nationwide Emergency Department-Based Syndromic Surveillance System for Better Public Health Responses in Taiwan. *BMC Public Health*, **8**, 18. <https://doi.org/10.1186/1471-2458-8-18>