Improvement of Influenza Incidence Estimation Using Auxiliary Information in Sentinel Surveillance in Japan

Miyuki Kawado¹, Shuji Hashimoto¹*, Akiko Ohta³, Mari S. Oba³, Kiyosu Taniguchi⁴, Tomimasa Sunagawa⁵, Tamano Matsu⁵, Masaki Nagai⁶ and Yoshitaka Murakami³

¹Department of Hygiene, Fujita Health University School of Medicine, Aichi, Japan
²Division of Public Health, Department of Social Medicine, Saitama Medical University Faculty of Medicine, Saitama, Japan
³Department of Medical Statistics, Toho University, Tokyo, Japan
⁴Department of Clinical Research, National Mie Hospital, Mie, Japan
⁵Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Tokyo, Japan
⁶Department of Social Medicine, Saitama Medical University, Saitama, Japan

Received: December 6, 2017 Revised: April 06, 2018 Accepted: April 13, 2018

Abstract:

Introduction:
Sentinel surveillance in Japan is used to estimate national influenza incidence under the assumption that Sentinel Medical Institutions (SMIs) are randomly selected. The current method might lead to overestimation when SMIs are recruited on a voluntary basis.

Aims & Objectives:
We aimed to improve influenza incidence estimation using auxiliary information without this assumption.

Materials and Method:
We used reports of influenza from SMIs in 2015, together with the number of all disease outpatients in September 2014 at all medical institutions from the Survey of Medical Institutions of Japan, as auxiliary information. The influenza incidence was estimated by the method using auxiliary information and the current method (without auxiliary information).

Result and Conclusion:
Influenza incidence rate per 1,000 population in 2015 estimated by using auxiliary information and by the current method was 63.7 (95% Confidence Interval (CI), 61.0-66.3) and 96.5 (95% CI, 93.0-100.0), respectively. The ratio of these estimates was 0.66. Our findings suggest that influenza incidence estimated by using the number of all disease outpatients as auxiliary information is more accurate than estimates by the current method.

Keywords: Surveillance, Infectious disease, Influenza, Epidemiology, Estimation, Methodology.

1. INTRODUCTION

Influenza sentinel surveillance has been established in many countries [1 - 5]. To estimate national incidence, the number of influenza patients reported by Sentinel Medical Institutions (SMIs) and the coverage of SMIs for the national population are often used [3]. In Japan, based on well-designed criteria for selecting SMIs in the guidelines for surveillance, influenza incidence is estimated under the assumption that SMIs are randomly selected [6, 7]. Previous
studies have indicated that if SMIs were recruited on a voluntary basis, medical institutions with more patient visits for influenza would likely be selected as SMIs [6, 8]; therefore, this method would lead to overestimation of influenza incidence. In Japan, the population coverage of each medical institution is unavailable owing to the free access healthcare system [6, 9]. However, data on the number of outpatients with diseases and injuries treated at each medical institution (all disease outpatients), available from the Survey of Medical Institutions conducted by the Ministry of Health, Labour and Welfare, would be useful as a substitute for the population coverage of SMIs when estimating national incidence [6, 10 - 12].

We attempted to improve influenza incidence estimation from sentinel surveillance data in Japan by using the number of all disease outpatients as auxiliary information and not using the assumption that SMIs were randomly selected.

2. MATERIALS AND METHODS

2.1. Surveillance for Influenza in Japan

The National Epidemiological Surveillance for Infectious Diseases (NESID) in Japan has been described elsewhere [5, 13]. It is organized by the Ministry of Health, Labour and Welfare, and involves a sentinel surveillance system for influenza. Prefecture governments select SMIs (about 3,000 pediatric and about 2,000 internal medicine SMIs) according to the Ministry guideline [14, 15]. The numbers of SMIs in the areas covered by health centers are approximately proportional to their population sizes. In reality, SMIs would rather select on a voluntary basis; although the Ministry guideline recommends random sampling of SMIs from all medical institutions with pediatric and/or internal medicine departments in the covered areas [6, 16, 17]. Each SMI reports the weekly number of influenza cases to the regional health center. Notification by health centers to the prefectural government and the Ministry is made through an online reporting platform.

2.2. Data

We analyzed reports of influenza in 2015 from SMIs included in the NESID in collaboration with the National Institute of Infectious Diseases of Japan, which routinely summarizes surveillance data. With permission from the Ministry, we also collected data of all disease outpatients in September 2014 from each hospital pediatric and internal medicine department and from each clinic of the Survey of Medical Institutions of Japan [10, 11, 13]. This survey is conducted every 3 years to determine the status of medical institutions (including all disease outpatients only in September) in Japan. The data of about 94% of SMIs are linked with the number of all disease outpatients by the name and address of the medical institution, and these data are available for estimating influenza incidence. Because these statistical data do not include personally identifiable information, no ethical review was needed.

2.3. Estimation of Influenza Incidence

Influenza incidence was estimated in the present study by two methods: without the use of auxiliary information (the current method), and with the number of all disease outpatients as auxiliary information [6, 7].

The current method of estimating influenza incidence, without the use of auxiliary information, has been previously described [6, 7]. Influenza incidence in each prefecture and the type of medical institution is estimated as the number of influenza patients at SMIs divided by the proportion of SMIs to all medical institutions. Total incidence is estimated as the total of influenza incidence estimates for all prefectures and type of medical institution. The four types of medical institution are as follows: 1) hospital pediatrics department, 2) clinic with pediatrics as the main department, 3) clinic with internal medicine and with pediatrics not as the main department, and 4) hospital internal medicine department and clinic with internal medicine but no pediatrics department.

To estimate influenza incidence by the new method, we used the ratio estimation method with the number of all disease outpatients as auxiliary information [18, 19]. The incidence of influenza in each prefecture and the type of medical institution was estimated as the number of influenza patients at SMIs divided by the ratio of the sum of SMIs to all medical institutions. Total incidence is estimated as the total of influenza incidence estimates for all prefectures and type of medical institution. The four types of medical institution are as follows: 1) hospital pediatrics department, 2) clinic with pediatrics as the main department, 3) clinic with internal medicine and with pediatrics not as the main department, and 4) hospital internal medicine department and clinic with internal medicine but no pediatrics department.

To estimate influenza incidence by the new method, we used the ratio estimation method with the number of all disease outpatients as auxiliary information [18, 19]. The incidence of influenza in each prefecture and the type of medical institution was estimated as the number of influenza patients at SMIs divided by the ratio of the sum of SMIs to all medical institutions. Total incidence is estimated as the total of influenza incidence estimates for all prefectures and type of medical institution. The four types of medical institution are as follows: 1) hospital pediatrics department, 2) clinic with pediatrics as the main department, 3) clinic with internal medicine and with pediatrics not as the main department, and 4) hospital internal medicine department and clinic with internal medicine but no pediatrics department. The reason for this is because the number of all disease outpatients...
obtained in the survey is by hospital department but is only by the institution in clinics [10, 11]. The appendix describes both methods in detail.

3. RESULTS

Table 1 shows the mean number of all disease outpatients in September 2014 at all medical institutions and SMIs. There were a total of 4,619 SMIs (6.9% of all medical institutions). Mean numbers of all disease outpatients at each type of medical institution were higher in SMIs than in all medical institutions. The ratios of mean numbers of all disease outpatients ranged from 1.1 to 1.6 at the five types of medical institution.

Table 1. Mean number of all disease outpatients in September 2014 at all and sentinel medical institutions.

<table>
<thead>
<tr>
<th>Type of Medical Institution</th>
<th>Hospital Pediatrics Department</th>
<th>Clinic with Pediatrics as the Main Department</th>
<th>Clinic with Internal Medicine and with Pediatrics not as the Main Department</th>
<th>Hospital Internal Medicine Department</th>
<th>Clinic with Internal Medicine but no Pediatrics Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Medical Institutions</td>
<td>2,328</td>
<td>5,840</td>
<td>13,616</td>
<td>7,025</td>
<td>37,654</td>
</tr>
<tr>
<td>Mean number of all disease outpatients in September 2014</td>
<td>667</td>
<td>1,068</td>
<td>1,059</td>
<td>1,333</td>
<td>905</td>
</tr>
<tr>
<td>Sentinel Medical Institutions (SMIs)</td>
<td>631</td>
<td>1,882</td>
<td>975</td>
<td>505</td>
<td>627</td>
</tr>
<tr>
<td>(%)</td>
<td>(27.1)</td>
<td>(32.2)</td>
<td>(7.2)</td>
<td>(7.2)</td>
<td>(1.7)</td>
</tr>
<tr>
<td>Mean number of all disease outpatients in September 2014</td>
<td>949</td>
<td>1,176</td>
<td>1,553</td>
<td>2,171</td>
<td>1,249</td>
</tr>
<tr>
<td>Ratio¹</td>
<td>1.4</td>
<td>1.1</td>
<td>1.5</td>
<td>1.6</td>
<td>1.4</td>
</tr>
</tbody>
</table>

1) Proportion of SMIs to all medical institutions. 2) Ratio of mean number of all disease outpatients in SMIs to that in all medical institutions.

Fig. (1) shows weekly influenza incidence rate in 2015 estimated using auxiliary information and the current method. Influenza incidence rate estimated by using auxiliary information was lower in each week than that by the current method. The ratios of estimates ranged from 0.45 to 0.72 during the period between weeks 1 and 16, with estimated weekly influenza incidence rate of more than 1 per 1,000 population.

Fig. (1). Weekly influenza incidence rate in 2015 estimated by methods with and without use of auxiliary information.

Table 2 shows influenza incidence in 2015 estimated using auxiliary information and the current method; estimated
influenza incidence rate per 1,000 population by these two methods was 63.7 (95% Confidence Interval (CI), 61.0-66.3) and 96.5 (95% CI, 93.0-100.0), respectively. The ratio of those estimates was 0.66. The incidence by the new and by the current method was 8,090,000 (95% CI, 7,750,000-8,430,000) and 12,270,000 (95% CI, 11,830,000-12,710,000), respectively.

Table 2. Influenza incidence in 2015 estimated by current method (without auxiliary information) and using auxiliary information.

<table>
<thead>
<tr>
<th>Influenza Incidence in 2015</th>
<th>Number (Thousands)</th>
<th>Rate per 1,000 Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
</tr>
<tr>
<td>Method using auxiliary information</td>
<td>8,090</td>
<td>7,750 - 8,430</td>
</tr>
<tr>
<td>Current method</td>
<td>12,270</td>
<td>11,830 - 12,710</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.66</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval. 1) Ratio of influenza incidence estimated by use of auxiliary information to that estimated by current method.

4. DISCUSSION

Our study showed that the average number of all disease outpatients at SMIs was higher than that of all medical institutions, for all types of medical institution. These data indicated that the mean size of the population who made influenza visits was larger at SMIs than at all medical institutions. From these data, we hypothesized that the assumption that SMIs were chosen by random sampling from all medical institutions was invalid [6, 7]. As mentioned in previous studies [6 - 8, 17], use of the current method under this assumption could lead to overestimation of influenza incidence. However, the method using auxiliary information does not use this assumption.

Based on the other information, we discuss the accuracy of influenza incidence estimated by the current method and the method using auxiliary information. Table 3 shows influenza incidence estimated by the current method, the numbers of influenza patients from the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB), and the NDB coverages for all medical claims for three influenza seasons; this information was obtained from a previous study [8] and a report of the Health Insurance Claims Review and Reimbursement Services of Japan [9]. The universal insurance system in Japan covers most Japanese citizens. An NDB system to collect data for all medical claims in electronic format from the insurance system has recently been established by the Ministry of Health, Labour and Welfare of Japan [20, 21]. The coverage of the NDB for all medical claims (including medical claims in non-electronic format) has increased rapidly, from 50.2% in 2008 to 93.1 in 2010 and 95.5% in 2012 [9, 20]. Assuming that the actual influenza incidence is equal to the number of influenza patients from the NDB divided by the coverage, the magnitude of overestimation of influenza incidence by the current method is calculated to be 1.76, 1.71, and 1.55 in the 2010/2011, 2011/2012, and 2012/2013 seasons, respectively [8]. Considering that influenza patient visits are more likely to be at clinics, completion of the electronic medical claims system has been delayed at clinics, and that the NDB coverage for all medical claims has rapidly increased, the magnitude of overestimation of influenza incidence by the current method in 2012/2013 (1.55 times) might be more accurate than in 2010/2011 and 2011/2012 [20].

Table 3. Influenza incidence estimates by current method, influenza patients from NDB, and NDB coverages for all medical claims in 2010/2011, 2011/2012 and 2012/2013 seasons.

<table>
<thead>
<tr>
<th>Season</th>
<th>Influenza Incidence Estimated by Current Method, without Auxiliary Information (Thousands)</th>
<th>Number of Influenza Patients from NDB (Thousands)</th>
<th>Coverage of NDB for all Medical Claims (%)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010/2011</td>
<td>13,880</td>
<td>7,356</td>
<td>93.1</td>
<td>1.76</td>
</tr>
<tr>
<td>2011/2012</td>
<td>16,800</td>
<td>9,273</td>
<td>94.5</td>
<td>1.71</td>
</tr>
<tr>
<td>2012/2013</td>
<td>11,800</td>
<td>7,287</td>
<td>95.5</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Abbreviation: NDB, National Database of Health Insurance Claims and Specific Health Checkups of Japan.
1) Seasons for influenza incidence estimation were between week 36 in 2010 and week 35 in 2011; between week 36 in 2011 and week 35 in 2012; and between week 36 in 2012 and week 13 in 2013. Seasons for the number of influenza patients were between September 2010 and August 2011; September 2011 and August 2012; and September 2012 and March 2013. Seasons for NDB coverage were March 2011, March 2012, and March 2013.
2) Influenza incidence estimates and numbers of influenza patients have been published previously [8].
3) NDB coverages were obtained from a published report [9].
4) The ratio was the influenza incidence estimate divided by the number of influenza patients and the NDB coverage.
The present study showed that influenza incidence in 2015 estimated by using auxiliary information was lower than that by the current method, without using auxiliary information; the ratio of those estimates was 0.66. Assuming that the actual influenza incidence was equal to the number of influenza patients from the NDB divided by the coverage, the magnitude of overestimation of influenza incidence by the new method (using auxiliary information) would be calculated as the ratio of 0.66 multiplied by the magnitudes of overestimation by the current method (1.76, 1.71, and 1.55), yielding 1.16, 1.13, and 1.02 in 2010/2011, 2011/2012, and 2012/2013, respectively. Although the ratio of 0.66 based on the influenza incidence in 2015 might not directly apply in other years, the new method using auxiliary information would greatly reduce the overestimation of influenza incidence compared with that of the current method, and could lead to estimates that are close to the actual influenza incidence. Therefore, it would be appropriate to change the current method to the method that includes auxiliary information to estimate influenza incidence in the NESID.

To improve the current method of influenza incidence estimation, we used the ratio estimation method with all disease outpatients as auxiliary information [18, 19]. The reasons were as follows. The population coverage of each medical institution is unavailable in Japan owing to the free access healthcare system [6, 19]. To improve estimation of the number of patients in sample surveys, ratio estimation is used relatively frequently. The Patient Survey with a sample of medical institutions throughout Japan conducted by the Ministry of Health, Labour and Welfare can serve as an example. In that survey, the number of patients with diseases and injuries nationally is estimated by ratio estimation, with all disease inpatients and outpatients used as auxiliary information [11, 22]. A previous study proposed ratio estimation using local general practitioner density (the number of general practitioners per the population in a local area) as auxiliary information, to improve influenza incidence estimated without use of auxiliary information in sentinel surveillance in France [19].

Ratio estimation requires auxiliary information about the entire population [18]. We used the number of all disease outpatients in September as auxiliary information. Many diseases, including pediatric diseases, have seasonal variation, and the number of all disease outpatients varies with season. The number of all disease outpatients in September might not be the best auxiliary information to use for estimating influenza incidence with great seasonal variation. However, if we are unaware of a better alternative that can be regularly obtained in Japan. The number of all disease outpatients is provided by the Survey of Medical Institutions at triennial intervals and is used as auxiliary information to estimate the number of patients with diseases and injuries nationally in the Patient Survey of Japan [10, 11, 22]. Information on the number of general practitioners per the population in a local area as auxiliary information, to improve influenza incidence estimated without use of auxiliary information in sentinel surveillance in Japan because the numbers of SMIs in local areas are approximately proportional to their population sizes [15, 19].

As mentioned above, four types of medical institution are used as the strata in the current method of the NESID whereas five types of the institution were used in the new method of the present study. The influenza incidence in 2015 estimated by the current method was 10,920,000 using the strata with five types of medical institution, which was 0.89 times that in the strata with four types (Table 2). The difference in these estimates using strata with four or five types of the medical institution would be associated with the distribution of influenza patients in medical institutions and SMIs in the strata of hospital internal medicine department and clinic with internal medicine but no pediatrics department. The reasons for this difference are unknown.

This study has some limitations. We estimated the influenza incidence in Japan for 2015 only. The influenza incidence rate has varied widely by year and by area [7, 23]. The dominant virological type of seasonal influenza has changed by year; in the 2014/2015 season, influenza type A(H3) was dominant in Japan [24]. Although the accuracy of influenza incidences in Japan estimated by the two methods might not have been extremely affected by the year or virological type of seasonal influenza, it is important to examine the estimates for several years [17]. The numbers of all disease outpatients in September 2014, not in 2015 were used. About 94% of SMIs linked their data of numbers of all disease outpatients by the medical institution name and address, to estimate influenza incidence. Our influenza incidence estimate in 2015 was 0.97 times that based on data of all SMIs (12,590,000 reported in the NESID). We did not directly evaluate the accuracy of influenza incidence estimated by the method using auxiliary information. To evaluate the accuracy, we used numbers of influenza patients in the 2010/2011, 2011/2012 and 2012/2013 seasons from the NDB that were published in a previous study [8]. The NDB had been made available on a trial basis for research until March 2013, but it was unavailable in 2015 [8, 20]. Direct evaluation of the accuracy of influenza incidence estimated by the method using auxiliary information based on actual incidences is very important and requires further investigation in the future.
CONCLUSION

Our findings suggest that influenza incidence estimated by using the number of all disease outpatients as auxiliary information would be more accurate than estimates made without using such information.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

This study was supported by a Research grant on Emerging and Re-emerging Infectious Diseases of the Ministry of Health, Labour and Welfare of Japan [grant number H27-Shinkougyousei-Shitei-001].

APPENDIX

Methods of Estimating Influenza Incidence

The methods of estimating influenza incidence (current method and by including auxiliary information) are as follows [6, 7]. Let \(n, v\) and \(\xi\) be the number of SMIs, the number of all medical institutions, and the total number of all disease outpatients in all medical institutions, respectively. Let \(Y_i\) and \(x_i\) be influenza incidence and the number of all disease outpatients in the \(i\)-th SMI, respectively, for \(i = 1, 2, \ldots, n\). Let \(\bar{Y}, s_y, \) and \(\bar{x}\) be the mean and standard deviation of \(\{Y_i\}\), and the mean of \(\{x_i\}\), respectively.

Let \(\alpha\) be the total influenza incidence in all medical institutions. By the current method, the estimate of \(\alpha\) is given to be \(\hat{\alpha} = v \times \bar{Y} / n\); i.e., the incidence is estimated as the total incidence in SMIs \((n \times \bar{Y})\) divided by the proportion of SMIs among all medical institutions \((n/v)\). The approximate confidence interval for \(\alpha\) is given to be \((\hat{\alpha} - 1.96 \times s_y, \hat{\alpha} + 1.96 \times s_y)\) where \(s_y^2\) is an estimate of variance of \(\hat{\alpha}_y\) and is given to be \((v \times s_y^2)((n - 1)/n)(v/(v - 1))(1/n - 1/v)\).

By the method using auxiliary information, the estimate of \(\alpha\) is given to be \(\hat{\alpha}_\ell = \xi \times R\) where \(R = \bar{Y} / \bar{x}\); i.e., the incidence is estimated as the total incidence in SMIs \((n \times \bar{Y})\) divided by the ratio of the sum of the numbers of all disease outpatients in SMIs to the sum of those in all medical institutions \((n \times x/\xi)\). The approximate confidence interval for \(\alpha\) is given to be \((\hat{\alpha}_\ell - 1.96 \times s_\ell, \hat{\alpha}_\ell + 1.96 \times s_\ell)\) where \(s_\ell^2\) is an estimate of variance of \(\hat{\alpha}_\ell\) and is given to be \(v^2 \left(\sum (Y_i - R \times x_i)^2/(n - 1)\right) (1/n - 1/v)\) [16].

Consider that the incidences in some strata, such as type of medical institution and prefecture, are estimated using the above methods. Let \(k\) be the number of strata, \(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_k\) the estimated incidences in the strata, and \(s_1^2, s_2^2, \ldots, s_k^2\) their estimated variances by the either method (with or without auxiliary information). The approximate confidence interval for the total incidence is given as \((\hat{\alpha}_\ell - 1.96 \times s_\ell, \hat{\alpha}_\ell + 1.96 \times s_\ell)\) and \(\hat{\alpha}_\ell = \hat{\alpha}_1 + \hat{\alpha}_2 + \ldots + \hat{\alpha}_k\) and \(s_\ell^2 = s_1^2 + s_2^2 + \ldots + s_k^2\).

REFERENCES


© 2018 Kawado et al.
This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: https://creativecommons.org/licenses/by/4.0/legalcode. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.