Brief Report

DOI: 10.5582/bst.2018.01279

Comparative study of preciseness in the regional variation of influenza in Japan among the National Official Sentinel Surveillance of Infectious Diseases and the National Database of Electronic Medical Claims

Yasushi Ohkusa[1], Tamie Sugawara[1], Kenzo Takahashi[2,*], Miwako Kamei[3]

[1] *National Institute of Infectious Diseases, Tokyo, Japan;*
[2] *Teikyo University Graduate School of Public Health, Tokyo, Japan;*
[3] *School of Pharmacy, Nihon University, Chiba, Japan.*

Summary **In Japan, national official surveillance for influenza has been performed at about 5,000 sentinel hospitals/clinics by the National Official Sentinel Surveillance of Infectious Diseases (NOSSID). Meanwhile, all electronic medical claims nationwide in the National Database of Electronic Medical Claims (NDBEMC) were recently disclosed by the Ministry of Health, Labour and Welfare of Japan. We compared the regional variation of influenza incidence among prefectures between the NOSSID and NDBEMC. The data were extracted from NOSSID and the NDBEMC for the 2010/2011 through 2013/2014 seasons. We compared the data of both datasets season by season by using Spearman's rank correlation in each season. Spearman's rank correlation values for the four seasons were 0.7823, 0.3907, 0.4961 and 0.4543, and their *p*-values were less than 0.00005, 0.0066, 0.0004 and 0.0013, respectively. Statistically, regional variation of influenza incidence in NOSSID is not imprecise, but its correlation with the NDBEMC dataset is not so high. It is important to note this fact when interpreting regional variation in NOSSID.**

Keywords: Influenza, National Database for Electronic Medical Claims, National Official Sentinel Surveillance, regional variation

1. Introduction

Official surveillance, especially for influenza, is sentinel surveillance in many countries such as the United States (*https://www.cdc.gov/flu/weekly/overview.htm*), Europe (*http://flunewseurope.org/System*) and Japan. Because sentinel surveillance is not comprehensive, its interpretation or preciseness should be evaluated more comprehensively and precisely, as so called "big data," and not as timely information. "Big data" has not been available in the past, and thus sentinel surveillance has been considered to be the gold standard without evaluation its preciseness. However, recently, we have had access to "big data" with which to evaluate sentinel surveillance.

**Address correspondence to:*
Dr. Kenzo Takahashi, Teikyo University Graduate School of Public Health, 2-11-1 Kaga, Itabashi-ku, Tokyo 173-8605, Japan.
E-mail: kenzo.takahashi.chgh@med.teikyo-u.ac.jp

Based on the Law Related to the Prevention of Infectious Diseases and Medical Care for Patients of Infections (The Infectious Diseases Control Law) in Japan, national official surveillance for influenza has been performed at about 5,000 sentinel hospitals/clinics, which represent about one tenth of the pediatric or internal medicine facilities nationwide. We call this system the National Official Sentinel Surveillance of Infectious Diseases (NOSSID) (*1,2*). As there has been no comparable surveillance so far, the preciseness of NOSSID cannot be evaluated. Recently, however, data of almost all electronic medical claims nationwide have been disclosed in the "National Database of Health Insurance Claims and Specific Health Checkups of Japan" (*3*) by the Ministry of Health, Labour and Welfare (MHLW) of Japan (*4-7*). We call this the National Database of Electronic Medical Claims (NDBEMC). This data covered 98.4% of all medical claims made in 2015 (*8*). All doctors must record a diagnosis for medical claims; therefore, this data must necessarily constitute the most reliable data source to date. Exceptionally, only small clinics operated by only one doctor older than 65 years of age were exempted from providing non-electronic medical claim information as an interim measure.

In Japan, almost all influenza patients receive a rapid diagnostic test. Therefore, influenza patients in Japan are not patients with influenza-like-illness; rather, they are patients actually diagnosed as having influenza by a rapid diagnostic test. Information in both NOSSID and the NDBEMC is reported by the diagnosing doctor. For NOSSID, doctors have to report their diagnostic information to the public health center weekly. NDBEMC contains the information they send monthly to the health insurer when submitting electronic medical claims, which includes the diagnosis.

Our previous study evaluated the number of influenza patients from NOSSID nationwide (*9*). In the present study, we focus on the regional variation of influenza incidence among prefectures in NOSSID in comparison with that in the NDBEMC. So far, as there is no more comprehensive information available than that from sentinel surveillance, we have to believe that the sentinel surveillance data is accurate. For example, the MHLW or mass media have sometimes reported the highest influenza incidence in a prefecture based on sentinel surveillance. However, we questioned whether the information was true or credible. This study attempted to answer this question by using NDBEMC, which has the most comprehensive data, to evaluate the preciseness of NOSSID by comparing the data of NOSSID with that of NDBEMC.

2. Materials and Methods

2.1. *Data and study period*

NOSSID for influenza has about 5,000 sentinels, which represent almost one tenth of all pediatric or internal medicine clinics/hospitals nationwide. Among these 5,000 sentinels, 3,000 are pediatric clinics/hospitals and 2,000 are clinics/hospitals for adult patients. They report the number of influenza patients per sentinel weekly by prefecture. Age, sex, and other characteristic of the patients were not available. We used the published data of the number of reported patients by prefecture weekly.

The NDBEMC can be used to count patients diagnosed as having influenza. We used the reported number of patients who had been diagnosed as having influenza, excluding suspected cases but including complex cases, by prefecture and by month in the NDBEMC. Age, sex, and other characteristic of the patients were also not available in this application. In May 2015, approximately 98.4% of all electronic medical claims were covered (*8*). Data from the NDBEMC were available from April 2009 through March 2015. We obtained permission to use this data for research purposes.

The influenza season was defined as the period spanning September to August of the following year or, equivalently, from the 36th epidemiological week to the 35th epidemiological week of the following year. The study period was from the 2010/2011 season through the 2013/2014 season.

2.2. *Statistical analysis*

Data from NOSSID were aggregated in the season by prefecture and divided by the number of sentinels in each prefecture. Data from the NDBEMC were aggregated in the season by prefecture and divided by the population in each prefecture. The mid-year population during each season was adopted as the population for each prefecture. Since the data distribution was not linear, thus we decided to apply non-parametric method. We compared the rank of influenza incidence by prefecture in both datasets season by season for the study period so as to solve objective in the present study. We used Spearman's rank order correlation among prefecture in the incidence of influenza in each season. We adopted 5% as a significant level.

2.3. *Ethical considerations*

This study used only anonymous data that had been de-linked from individual patient information. Therefore, ethical issues related to medical institutions and pharmacies do not pertain in this study. The NOSSID data were open to all. Author MK was approved to use the NDBEMC data by the MHLW on July 27, 2016 (Research project: Estimation of the number of patients of infectious diseases).

3. Results and Discussion

Figure 1 shows the regional variations in the incidence rate in the NDBEMC over the four seasons, and Figure 2 shows the regional variations in the total number of patients per sentinel in NOSSID. These figures show that the influenza incidence in the 2013/2014 season, which is shown as a black line in the graphs, was higher than that for other three seasons in almost all prefectures in the NDBEMC, whereas in NOSSID, there was no clearly dominating season in almost all prefectures.

Table 1 shows the prefecture rankings for the number of influenza cases in both datasets. We can see some discrepancies among the NDBEMC and NOSSID. For example, Okinawa was on top for three seasons in NOSSID and in second place in the 2012/2013 season. It was also second in the NDBEMC for three seasons, but it was 43rd, which was 5th from the bottom, in the 2012/2013 season. Conversely, the top prefectures in the NDBEMC did not appear in the top ten of NOSSID in the same season except for Hiroshima in the 2012/2013

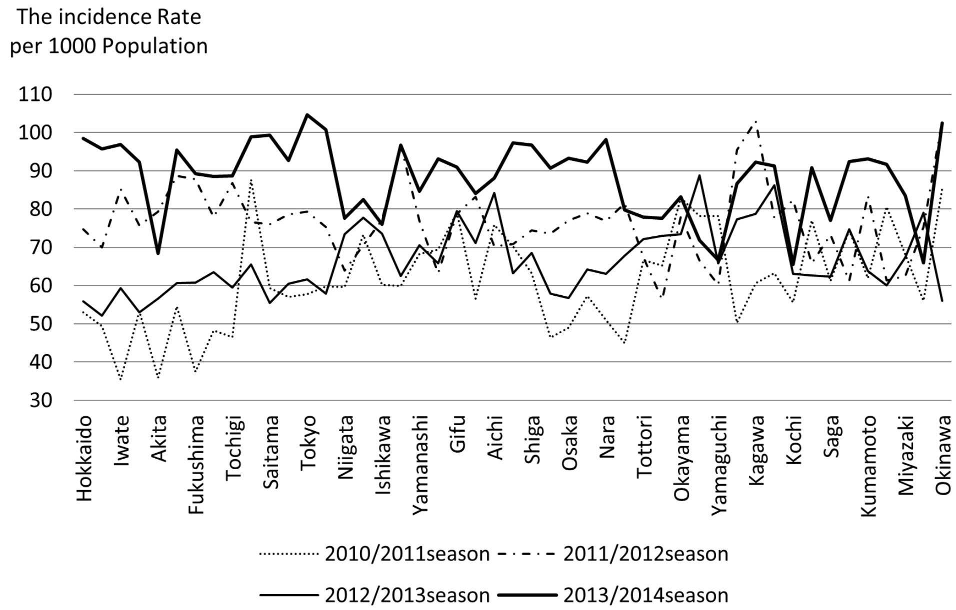

Figure 1. Incidence rate per 1000 population in the NDBEMC by prefecture for the four seasons. *Note*: The NDBEMC is the National Database of Electronic Medical Claims. This figure shows the cumulative incidence, which is the number of newly diagnosed influenza patients from the 2010/2011 season through the 2013/2014 season in the NDBEMC divided by the population by prefecture. The influenza season is defined as September to August of the following year. The bold black line shows the situation in the 2013/2014 season, the thin black line shows the 2012/2013 season, the gray line shows the 2011/2012 season, and the dashed line shows the 2010/2011 season.

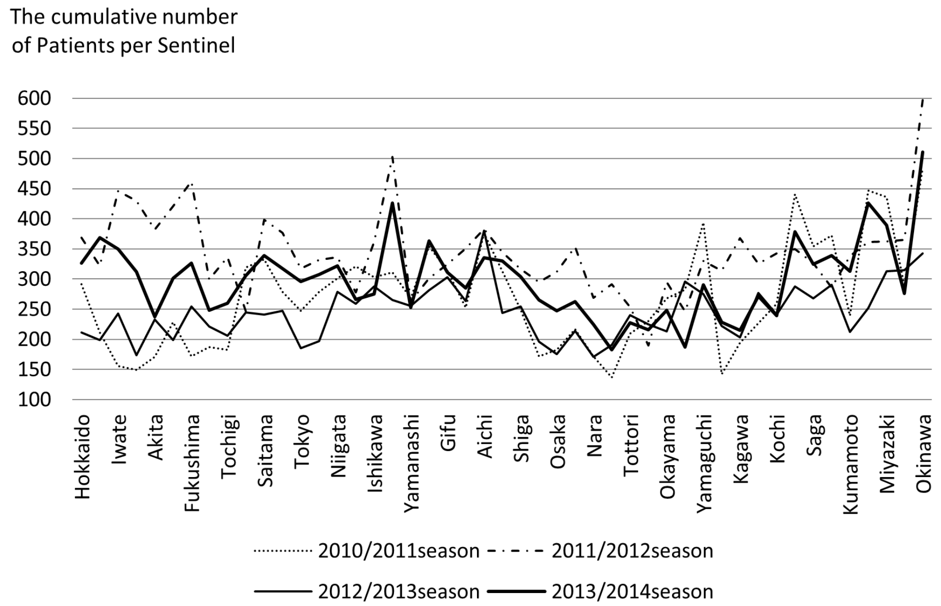

Figure 2. Cumulative number of patients per sentinel in NOSSID by prefecture for the four seasons. *Note*: NOSSID is the National Official Sentinel Surveillance for Infectious Diseases. This figure shows the cumulative number of influenza patients in one season per sentinel by prefecture. The influenza season for NOSSID extends from the 36th epidemiological week to the 35th epidemiological week of the following year. The epidemiological week in Japan is defined as Monday through Sunday. The first epidemiological week is the week including the first of January if that day was Monday-Wednesday. If that day was Thursday-Saturday, then the first epidemiological week starts from the first Monday of the year. The bold black line shows the situation in the 2013/2014 season, the thin black line shows the 2012/2013 season, the gray line shows the 2011/2012 season, and the dashed line shows the 2010/2011 season.

season. Spearman's rank correlation coefficients for the four seasons were 0.7823, 0.3907, 0.4961 and 0.4543, and their p-values were less than 0.00005, 0.0066, 0.0004 and 0.0013, respectively.

It shows some discrepancies between the rankings of the NDBEMC and NOSSID. For example, Okinawa prefecture was the top prefecture in three of the four seasons in the NOSSID ranking, but it was never ranked on top in the NDBEMC. In the exceptional season of 2012/2013, Okinawa prefecture was ranked second in

Table 1. Ranking by prefecture of the incidence rate per 1,000 population in the NDBEMC and the cumulative number of patients per sentinel in NOSSID for the four seasons

Ranking	Season 2010/2011		2011/2012		2012/2013		2013/2014	
	NDBEMC	NOSSID	NDBEMC	NOSSID	NDBEMC	NOSSID	NDBEMC	NOSSID
1	Gunma	Okinawa	Kagawa	Okinawa	Hiroshima	Aichi	Tokyo	Okinawa
2	Okinawa	Oita	Okinawa	Fukui	Ehime	Okinawa	Okinawa	Oita
3	Okayama	Fukuoka	Fukui	Fukushima	Aichi	Kagoshima	Kanagawa	Fukui
4	Oita	Miyazaki	Tokushima	Iwate	Gifu	Miyazaki	Saitama	Miyazaki
5	Gifu	Yamaguchi	Yamagata	Miyagi	Kagoshima	Gifu	Gunma	Fukuoka
6	Hiroshima	Aichi	Fukushima	Yamagata	Kagawa	Hiroshima	Hokkaido	Aomori
7	Yamaguchi	Nagasaki	Tochigi	Saitama	Toyama	Nagasaki	Nara	Nagano
8	Fukuoka	Nagano	Iwate	Aichi	Tokushima	Fukuoka	Mie	Iwate
9	Aichi	Saga	Kumamoto	Akita	Nagasaki	Ishikawa	Iwate	Saitama
10	Nagasaki	Saitama	Shizuoka	Chiba	Ishikawa	Nagano	Shiga	Nagasaki
11	Toyama	Toyama	Kochi	Hokkaido	Okayama	Niigata	Fukui	Aichi
12	Mie	Gunma	Wakayama	Kagawa	Niigata	Ehime	Aomori	Mie
13	Nagano	Gifu	Tokyo	Kagoshima	Shimane	Yamaguchi	Yamagata	Hokkaido
14	Yamanashi	Mie	Akita	Miyazaki	Tottori	Saga	Osaka	Fukushima
15	Miyazaki	Fukui	Hyogo	Oita	Shizuoka	Fukui	Nagano	Saga
16	Tottori	Ishikawa	Gifu	Ishikawa	Yamanashi	Shizuoka	Kumamoto	Niigata
17	Shimane	Niigata	Chiba	Hyogo	Shiga	Toyama	Chiba	Chiba
18	Ehime	Hokkaido	Okayama	Shizuoka	Wakayama	Yamanashi	Nagasaki	Kumamoto
19	Shiga	Kagoshima	Ibaraki	Fukuoka	Miyazaki	Fukushima	Hyogo	Miyagi
20	Kumamoto	Hiroshima	Ehime	Mie	Yamaguchi	Shiga	Kagawa	Gifu
21	Saga	Chiba	Ishikawa	Kochi	Nagano	Oita	Miyagi	Kanagawa
22	Kagawa	Kanagawa	Osaka	Kumamoto	Gunma	Chiba	Oita	Gunma
23	Ishikawa	Okayama	Yamanashi	Niigata	Hyogo	Gunma	Ehime	Shiga
24	Fukui	Yamanashi	Nara	Tochigi	Kumamoto	Mie	Gifu	Yamagata
25	Kanagawa	Kochi	Gunma	Kanagawa	Ibaraki	Iwate	Fukuoka	Tokyo
26	Niigata	Shizuoka	Saitama	Yamaguchi	Mie	Kochi	Kyoto	Yamaguchi
27	Saitama	Shiga	Miyagi	Saga	Nara	Saitama	Fukushima	Shizuoka
28	Tokyo	Tokyo	Kagoshima	Ehime	Kochi	Tottori	Tochigi	Kagoshima
29	Hyogo	Kumamoto	Kanagawa	Gifu	Fukuoka	Akita	Ibaraki	Ishikawa
30	Chiba	Shimane	Hokkaido	Aomori	Fukui	Shimane	Aichi	Ehime
31	Shizuoka	Yamagata	Shiga	Tokyo	Saga	Tokushima	Tokushima	Toyama
32	Kagoshima	Ehime	Kyoto	Shiga	Tokyo	Ibaraki	Yamanashi	Kyoto
33	Kochi	Hyogo	Saga	Tokushima	Fukushima	Hyogo	Shizuoka	Hyogo
34	Yamagata	Aomori	Mie	Osaka	Yamagata	Okayama	Miyazaki	Tochigi
35	Miyagi	Tottori	Toyama	Nagano	Chiba	Kumamoto	Okayama	Yamanashi
36	Hokkaido	Kagawa	Aichi	Ibaraki	Oita	Hokkaido	Toyama	Okayama
37	Nara	Ibaraki	Aomori	Kyoto	Tochigi	Tochigi	Wakayama	Ibaraki
38	Tokushima	Tochigi	Tottori	Okayama	Iwate	Kagawa	Tottori	Osaka
39	Aomori	Osaka	Hiroshima	Wakayama	Kyoto	Yamagata	Shimane	Kochi
40	Osaka	Nara	Fukuoka	Nagasaki	Kanagawa	Aomori	Niigata	Akita
41	Ibaraki	Fukushima	Niigata	Toyama	Osaka	Kanagawa	Saga	Tokushima
42	Tochigi	Kyoto	Nagano	Nara	Akita	Kyoto	Ishikawa	Tottori
43	Kyoto	Akita	Miyazaki	Yamanashi	Okinawa	Wakayama	Hiroshima	Nara
44	Wakayama	Iwate	Oita	Tottori	Hokkaido	Tokyo	Akita	Shimane
45	Fukushima	Miyagi	Nagasaki	Gunma	Saitama	Osaka	Yamaguchi	Kagawa
46	Akita	Tokushima	Yamaguchi	Hiroshima	Miyagi	Miyagi	Kagoshima	Hiroshima
47	Iwate	Wakayama	Shimane	Shimane	Aomori	Nara	Kochi	Wakayama

Note: NDBEMC is the National Database of Electronic Medical Claims. The influenza season in the NDBEMC is defined as September to August of the following year. NOSSID is the National Official Sentinel Surveillance for Infectious Diseases. The influenza season in NOSSID extends from the 36th epidemiological week to the 35th epidemiological week of the following year.

NOSSID but fifth from the bottom in the NDBEMC. Similarly, Wakayama prefecture was ranked as the least flu-prevalent prefecture in the two seasons of 2010/2011 and 2013/2014 in NOSSID, but it was never ranked as the lowest in the NDBEMC. The exceptional case was Shimane prefecture in the 2011/2012 season. It ranked as the least prevalent prefecture in both datasets. These discrepancies between the two datasets may have caused the low rank correlation of around 0.5, even though the four p values were quite small, and the null hypothesis that the two datasets were independent was rejected in all seasons.

We can surmise three reasons for the low rank correlation. First, in principle, the sentinels in NOSSID were chosen randomly. However, they were determined as the outcome of bargaining between the local government and the regional medical associations, and thus, in fact, the sentinels were not selected randomly. For example, if a local government would like to know about the early stage of an outbreak, they might ask larger hospitals and clinics to be sentinels. This process may bias the NOSSID data. The second reason relates to bias in the proportion of sentinels. Among the 5,000 sentinels in NOSSID, about 3,000 are pediatric clinics/hospitals and 2,000 are clinics/hospitals for adult patients. Thus, the coverage in children is much higher than that in adults, even though its ratio is not well known because the age distribution of the pediatric and adult patients was not published by prefecture. Therefore, NOSSID may reflect epidemics in children much more accurately than in adults. In contrast, the NDBEMC has no such bias in age because almost all electronic claims are included. This difference in the two datasets may have caused the low rank correlation. The third reason might be incompleteness of the NDBEMC. As explained above, the coverage of the NDBEMC was quite high but not 100%. The information defect may occur at small clinics operated by only one older doctor. In such cases, the doctor can report records by handwriting, rather than electronically, and this interim measure can cause this defect. If there were systematic variations in the proportion or size of these small clinics to which the interim measure was applied, among regions, the lack of these clinics' data might cause bias in the NDBEMC and thus a subsequent lower rank correlation with NOSSID. However, since such a defect might affect a very small proportion of the data at least in the national average, and there is no specific reason for systematic variations in the proportion or size of these small clinics among regions, the possibility of this effect might be small.

There were some limitations in this study. First, we examined regional variation in influenza, but we do not know whether similar phenomena might be observed in other infectious diseases including respiratory syncytial virus infection; gastrointestinal infections; varicella; hand, foot and mouth disease; erythema infectiosum; or pertussis. Therefore, we cannot extend the obtained result to NOSSID itself.

Second, though we do not know the age distribution of the patients with influenza in NOSSID, we will be able to obtain it from the NDBEMC with a future application. Such information on age distribution may be insightful in interpreting regional variation. As mentioned before, the sentinels in NOSSID were designed to bias to pediatric patients rather than adult. Conversely, NEDEMC does not have those bias in age distribution. This difference might cause some discrepancies in two data set.

Third, as we have examined no research in countries other than Japan, we cannot apply the obtained results to sentinel surveillance itself, especially as related to other countries. As medical claims or records in other countries are available, we need to verify the preciseness of the sentinel surveillance.

In conclusion, we found that regional variation in the influenza data in NOSSID was not imprecise statistically, but its correlation with the NDBEMC dataset was not so high. It is important to note this fact when interpreting regional variation in the NOSSID data. The examination of infectious diseases other than influenza remains as a future challenge.

Acknowledgements

We acknowledge the MHLW for their permission to use the NDBEMC.

References

1. Hashimoto S, Murakami Y, Taniguchi K, Shindo N, Osaka K, Fuchigami H, Nagai M. Annual incidence rate of infectious diseases estimated from sentinel surveillance data in Japan. J Epidemiol 2003; 13:136-141.
2. Kawado M, Hashimoto S, Murakami Y, Izumida M, Ohta A, Tada Y, Shigematsu M, Yasui Y, Taniguchi K, Nagai M. Annual and weekly incidence rates of influenza and pediatric diseases estimated from infectious disease surveillance data in Japan, 2002-2005. J Epidemiol 2007; 17:S32-S41.
3. Ministry of Health, Labour and Welfare, Japan. National Database of Health Insurance Claims and Specific Health Checkups of Japan. *https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000193322.pdf* (accessed October 5, 2018). (in Japanese)
4. Health Insurance Bureau, Ministry of Health, Labour and Welfare, Japan. Expert Meeting of Providing the Information of Medical Claims and Others. The report of providing the information of medical claims and health check-up for metabolic syndrome to the third party. 2013. (in Japanese).
5. Health Insurance Bureau, Ministry of Health, Labour and Welfare. Expert Meeting of Providing the Information of Medical Claims and Others. The guideline for providing the information of medical claims and health check-up for metabolic syndrome. (updated 2013 Aug) 2011. (in Japanese).
6. Nakamura Y, Sugawara T, Ohkusa Y, Taniguchi K, Miyazaki C, Momoi M, Okabe N. Severe abnormal behavior incidence after administration of neuraminidase inhibitors using the national database of medical claims. J Infect Chemother 2018; 24:177-181.
7. Miura K, Miyagawa N, Murakami Y, Okayama A. Outlines of the national database on medical expenditure and health check-up. Hoken Iryo Kagaku 2013; 62:31-35. (in Japanese)
8. Health Insurance Claims Review & Reimbursement services (2015): Status of medical claims by claim form (Diagnosed in April, 2015). *http://www.ssk.or.jp/tokeijoho/tokeijoho_rezept/tokeijoho_04_h27.files/seikyu_2704.pdf.* (accessed July 3, 2018). (in Japanese)
9. Nakamura Y, Sugawara T, Kawanohara H, Ohkusa Y, Kamei M, Oishi K. Evaluation of the estimated number of influenza patients from National Sentinel Surveillance using National Database of Electronic Medical Claims. Jpn J Infect Dis 2015; 68:27-29.

(Received November 15, 2018; Revised December 24, 2018; Accepted December 29, 2018)