EHT Environmental Health and Toxicology

Open Access

Volume: 30, Article ID: e2015012, 7 pages
http://dx.doi.org/10.5620/eht.e2015012

• Original Article

eISSN: 2233-6567

Utilization of health insurance data in an environmental epidemiology

Jongsik Ha, Seongkyung Cho, Yongseung Shin

Korea Environment Institute, Sejong, Korea

Objectives In South Korea, health insurance data are used as material for the health insurance of national whole subject. In general, health insurance data could be useful for estimating prevalence or incidence rate that is representative of the actual value in a population. The purpose of this study was to apply the concept of episode of care (EoC) in the utilization of health insurance data in the field of environmental epidemiology and to propose an improved methodology through an uncertainty assessment of disease course and outcome.

Methods In this study, we introduced the concept of EoC as a methodology to utilize health insurance data in the field of environmental epidemiology. The characterization analysis of the course and outcome of applying the EoC concept to health insurance data was performed through an uncertainty assessment.

Results The EoC concept in this study was applied to heat stroke (International Classification of Disease, 10th revision, code T67). In the comparison of results between before and after applying the EoC concept, we observed a reduction in the deviation of daily claims after applying the EoC concept. After that, we categorized context, model, and input uncertainty and characterized these uncertainties in three dimensions by using uncertainty typology.

Conclusions This study is the first to show the process of constructing episode data for environmental epidemiological studies by using health insurance data. Our results will help in obtaining representative results for the processing of health insurance data in environmental epidemiological research. Furthermore, these results could be used in the processing of health insurance data in the future.

Keywords Episode of care, Health insurance data, Uncertainty assessment

Correspondence: Yongseung Shin
370 Sicheong-daero, Sejong 30147, Korea
Tel: +82-44-415-7740
Fax: +82-44-415-7644
E-mail: shiny@kei.re.kr

Received: March 4, 2015
Accepted: October 24, 2015
Published online: November 10, 2015

This article is available from: http://e-eht.org/

Introduction

Health insurance data in South Korea are generated from the national health insurance system. Such data are unprecedented in the world, containing health service-related information from all medical institutions and pharmacies in the country. Health insurance data are largely divided into claims data from medical care institutions and medical insurance eligibility data of medical service recipients. Claims data are related to billing for medical services provided by medical institutions, containing information about treatment, time of treatment, and relevant disease information. Eligibility data contain personal information of the medical service recipient, such as medical aid eligibility, sex, age, and place of residence.

Hence, when health insurance data are processed to obtain incidence or prevalence data, these may be utilized in studies to determine the causative factors of a disease in a patient group with the same disease or to ascertain the pattern of disease incidence in a population exposed to the same factors. In particular, the data may be particularly useful in generating initial causal relationship hypotheses for ecological studies in environmental epidemiology, which utilize population groups such as national

or local communities as research units [1]. For example, health insurance data can be utilized in ecological studies to compare frequencies of risk factors that are hypothesized to be causative factors in populations with a high prevalence of particular diseases and to analyze the relationship between disease prevalence and changes in the frequency of proposed risk factors over a period and within a specific population.

However, various problems may arise when utilizing health insurance data in environmental epidemiological studies. One of the biggest problems is that an individual may file multiple claims for the same disease, which appears as multiple occurrences of the disease owing to the nature of claims data. Furthermore, this may be a limitation when determining trends in disease incidence due to exposure to environmental risk factors in environmental epidemiological studies [2]. Therefore, a consistent standard is required to determine whether health-care services provided during a certain period belong to the same occurrence or disease episode, in order to effectively utilize health insurance data for environmental epidemiological studies.

In utilizing health insurance data originally intended for insurance billing purposes, Hornbrook et al. [3] proposed the concept of health-care episode. Health-care episodes regarding health problems are divided into illness, disease, care, and health maintenance, through which the beginning and end of the health problem are specified. However, the beginning and end of an episode is difficult to objectively define based on information such as "first contact with an etiological factor." As an alternative, Hornbrook et al. [3] introduced the episode of care (EoC), which defines the beginning and end of regular medical services provided by health institutions. An EoC represents a unit that defines the period from disease initiation to termination (death, termination of treatment, or recovery) as one event.

This study is about utilizing health insurance data by applying the EoC concept in an ecological study of environmental epidemiology. It aimed to provide avenues for improvement in the future application of the EoC concept after characteristic analysis of the process and results of applying the EoC concept to health insurance data.

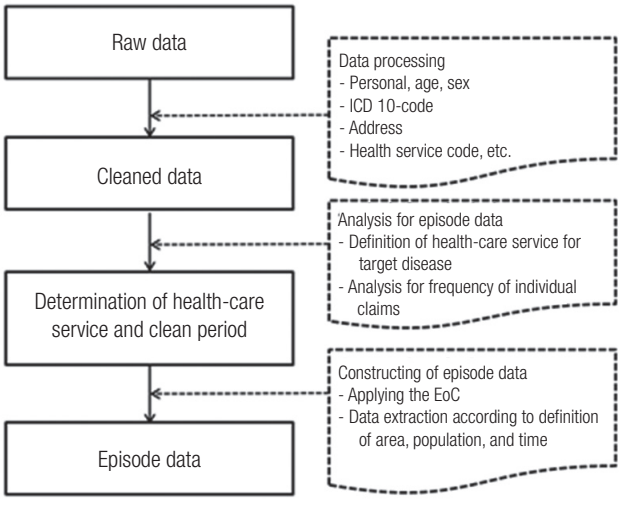

Figure 1. The procedure for the construction of the episode data. ICD-10, international Classification of Disease, 10th revision; EoC, episode of care.

Materials and Methods

Construction of Episode Data

In this study, we applied the EoC concept to heat stroke (International Classification of Disease, 10th revision [ICD-10] code: T67) claims from 2003 to 2010 as utilizing health insurance data in an ecological study of environmental epidemiology. In addition, the EoC concept was applied to all age groups in Seoul. Health insurance data on heat stroke were obtained from the National Health Insurance Service (NHIS). The heat stroke-related health insurance data contained information on claims and eligibility data of medical service recipients, including items relevant to the application of the EoC concept, such as individual identification (ID), place of residence, medical service start date, form of medical service (hospitalization, outpatient clinic, etc.), major/minor disease name, and days of hospitalization or visits.

Health insurance data from the NHIS were constructed into EoC data through 3 steps (Figure 1). First, the variables from the raw health insurance data that were required for construction of the EoC data were processed, including individual ID, sex, age, address at the time of receiving medical care, start date of medical care, and major and minor disease names. Second, periods of receiving health-care service and clean periods were defined, which is required for construction of the EoC data on the target disease. Third, individual EoC data were constructed and time-specific, location-specific, population-specific data were extracted.

With regard to the EoC data construction shown in Figure 1,

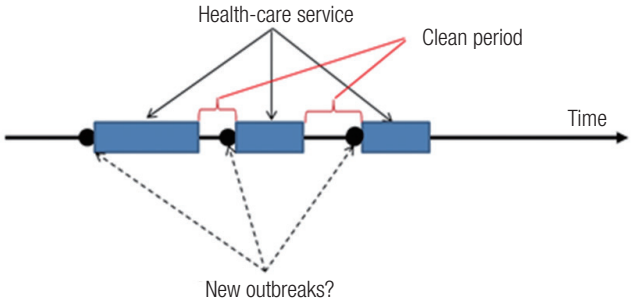

Figure 2. The episode of care concept.

the most important factors in applying EoC are the periods of receiving medical service and clean periods [4]. Medical service was defined as a series of events that are performed to overcome a particular health problem or disease, such as medical services provided by medical institutions, prescription of medication, hospitalization, and surgery. Clean period was defined as the time in which no medical service was provided to the same recipients. Figure 2 shows the EoC concept in accordance with an individual's medical care receipt. The schematic diagram indicates that consecutive provision of the same medical service was categorized into a distinct EoC when the clean period exceeded a certain period. Thus, the duration of the clean period determined whether medical services provided during a period belonged to the same EoC.

Uncertainty Assessment

Characteristic analyses of the process and results of the application of the EoC concept to health insurance data were performed by conducting an uncertainty assessment. Uncertainty assessment is useful in determining the degree of potential uncertainty and its impact on the final results in the processes of analyzing and recognizing, interpreting, and concluding from the processed results.

For uncertainty assessment, a concept for collectively evaluating various forms of uncertainties, is required. Experts suggested a three-dimensional uncertainty assessment, divided and analyzed into location, level, and nature of uncertainty, in this order [5]. Location of uncertainty is related to the point in which uncertainty may arise in a particular evaluation stage. Location of uncertainty can be subdivided into context, model, input, and model outcome uncertainties. Level of uncertainty is related to existence of knowledge, ranging from total ignorance to total knowledge and subdivided into decision, in which future situations or results are certain; statistical uncertainty; scenario uncertainty; recognized ignorance; and total ignorance, which indicates unawareness of the ignorance itself. Lastly, the nature of uncertainty is subdivided into epistemic and variability uncertainties. Epistemic uncertainty signifies uncertainty that arises from imperfection of the knowledge of the researcher, and variability uncertainty signifies uncertainty that arises from intrinsic variability.

Uncertainty assessment can be visually represented through an uncertainty matrix, which shows where uncertainties arise and a list of the characteristics of the levels and causes of these uncertainties. Characteristic analyses of the process and results of the EoC application to health insurance data in this study were performed by using an uncertainty matrix, in which uncertainties were confirmed and characterized qualitatively.

Results

Application of the Episode of Care Concept

Application Process

Table 1 summarizes the items related to the application of EoC to heat stroke in this study.

First, medical services were defined as health insurance claims from 2003 to 2010 for hospitalization with ICD-10 code T67 as main diagnosis or subdiagnosis. In addition, it was limited to claims from institutions that allow hospitalization, and only the claims with no missing variables in claim or individual data were included.

Second, in this study, clean period was defined as the "period between discharge and rehospitalization," which signifies the period between the time a patient was discharged from the first hospitalization and the time that the same patient was rehospitalized. In addition, in order to handle clean periods for hospitalization claims related to the same patient, we amended and supplemented clean periods by confirming the continuity of hospitalization for individuals. Specifically, the maximum number of days that could be categorized as the same hospitalization was evaluated by frequency analysis of rehospitalization after discharge by using hospitalization claims at the individual level. In addition, a sensitivity analysis was performed on the different definitions of medical services, according to the definitions of

Table 1. The health-care service and clean period for episode data

Steps	Detailed contents
Health-care service	- Period: Jan 1, 2003 - Dec 31, 2010 - Diagnosis code: T67 (ICD-10 code) in main or subdiagnosis - Treatment type: inpatient treatment (i.e., hospitalization) - Institutions: medical care institutes which is possible hospitalization - Additional information: claims with clear information such as personal ID, age, address, and hospitalization period etc.
Clean period	- Clean period: non-hospitalization period - Connectivity whether or not of rehospitalization after a discharge from a hospital: analysis of variable for rehospitalization interval after a discharge from a hospital in individual claims

ICD-10, International Classification of Disease, 10th revision; ID, identification.

medical institutions, and major and minor diseases associated with a definition of medical services.

Application Results

In this study, for the application of EoC, medical services were restricted to hospitalization. In this case, definition of clean period becomes relatively clear as a period of no hospitalization, but distinguishing separate claims for a continuous period of hospitalization becomes an important task. As such, the results of this study focused on frequency analysis results for periods between discharge and rehospitalization indicated in hospitalization claims at the individual level.

Figure 3 is a schematic representation of the reduction in the total number of claims with the period between discharge and rehospitalization for all hospitalization claims due to heat stroke in all age groups in Seoul. The period between discharge and rehospitalization shows the highest distribution at 0 days for hospitalization claims due to heat stroke and almost no difference is seen afterward. This characteristic was consistently shown regardless of main/subdiagnosis and institutional definitions. As such, this study established clean periods by regarding 0 day as the time of first discharge and rehospitalization for heat stroke as the same claim, and applied the concept of EoC accordingly.

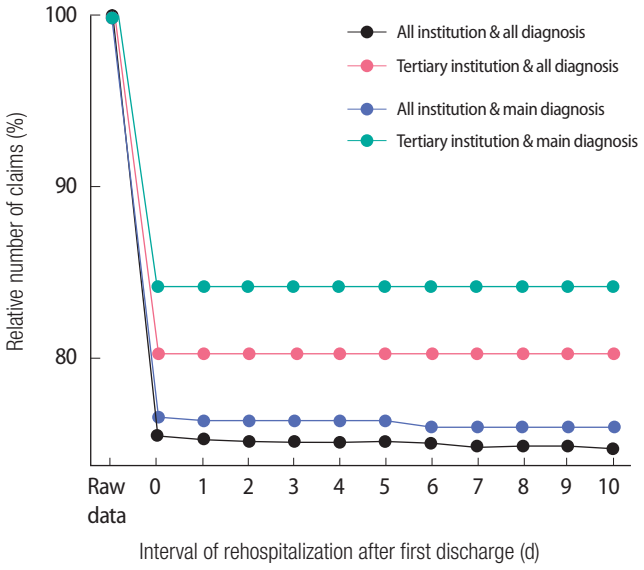

Figure 3. Relative number of claims according to interval of rehospitalization after the first discharge (d). All diagnosis means that the claim of heat stroke at discharge is the main diagnosis or subdiagnosis.

Figure 4 shows daily counts of heat stroke-related hospitalization claims for all age groups in Seoul and the results of the EoC data construction. Figure 4A shows results before the application of EoC, and Figure 4B shows results after the application of EoC. Although time sequential changes before and after the application of EoC are not clear, the overall deviation of daily counts before the application was reduced after the application EoC.

Uncertainty Assessment for Episode Data on Heat Stroke

Table 2 shows the uncertainty matrix of location, level, and nature of uncertainties that may occur during the process of applying EoC to heat stroke claims.

Context uncertainty is related to the definition of medical services relevant to the application of EoC. Although a number of environmental epidemiological studies utilizing health insurance data have been performed to date, the data were not processed based on specific criteria. In their study that evaluated the health impact of daily air pollution levels on acute asthma, Kim [6] defined disease incidence by extracting data based on claims for night or holiday visits and emergency room visits, eliminating patients with a medical history of asthma. However, in another study that estimated disease burden through confir-

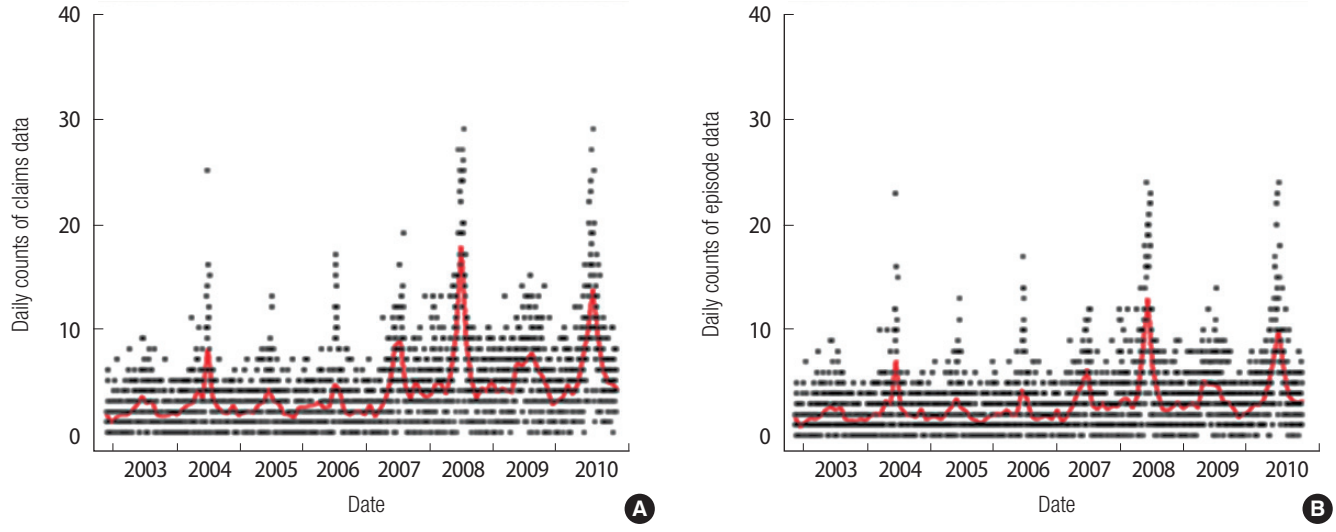

Figure 4. Time series trends of claims (A) and episode (B) data before and after application of the episode of care concept to heat stroke.

Table 2. The uncertainty assessment for episode data of heat stroke using uncertainty matrix

Location		Level			Nature	
		Statistical uncertainty	Scenario uncertainty	Recognized ignorance	Epistemic uncertainty	Variability uncertainty
Context	Operational definition of health care service					
	Claim for outpatient vs. inpatient		+[a]		+	
	Claim in main and/or subdiagnosis		+		+	
	Claim in specific medical care institutions		+		+	
Model	Methodology for the setting of clean period					
	Clinical experience vs. data analysis		+		+	
	Variation by time, place, and person of clean period			+		+
Inputs	Information of health insurance data					
	Medical claim data vs. care data	+			+	
	Utilization of additional information		+		+	
	Claim characteristics of health insurance data					
	Spatial-temporal characteristics by health care provider			+		+
	Change of treatment patterns			+		+

[a]The sign of + indicate whether the characterization is or not in the uncertainty matrix.

mation of asthma prevalence, Park et al. [7] estimated asthma prevalence by limiting their data to cases of two or more asthma claims (main diagnosis or subdiagnosis) and cases with prescription of asthma-related medications, while eliminating claims that correspond with other specific criteria. Generally, the medical service criteria used to process health insurance data differ according to the purpose that the data are being used for.

The concept of EoC applied to heat stroke in this study, as shown in Table 1, can be considered as a case of episode data. In particular, the definition of medical services used to determine the effects of exposure to environmental risk on disease incidence may cause uncertainties in determining the actual incidence of heat stroke. The definition of medical services specific to heat stroke incidences can vary for inpatients or outpatients, major or minor diagnosis, and the types of institution that provided the medical service. In addition, it can be further subdivided through additional information such as medication information and whether or not it was an emergency room visit. As such, uncertainties pertaining to the definition of medical services have characteristics of scenario and epistemic uncertainties (Table 2).

Model uncertainty is related to the establishment of clean periods with regard to the application of EoC. Various applications of EoC to health insurance data are possible. Winqert et al. [4] divided 31 main diagnoses into 6 types and explained the methodology of constructing EoCs based on the particular disease or purpose of utilization. In particular, in applying the EoC concept, the characteristics of a disease are related to the possibility of recovery and, subsequently, the duration of clean periods. For example, in the case of diabetes, which is an incurable disease, all medical service usage pertaining to diabetes after the first incidence can be viewed as one episode. On the other hand, when cases of hospitalization due to hypoglycemic shock as a result of failed glycemic control are considered instead of diabetes itself, one patient may have multiple episodes [8].

The durations of the clean periods in heat stroke for this study, as shown in Table 1, were established according to the definition of medical services. In other words, as incidence of heat stroke was defined based on insurance claims for hospitalization, clean periods can be defined as the time in which no hospitalization service was provided. However, after analyzing the characteristics of the actual claims data, continuity of individual hospitalization service was defined. For example, two studies by Kim et al. [9] and Kim [10] defined 2 days or 0 day as the maximum duration for the clean period when determining the continuity of individual hospitalization services. A study by Kim et al. [9] evaluated the health impact of weather changes, and a study by Kim [10] evaluated the health impact of changes in air pollution on hospitalization episodes (actual start to end of hospitalization) [9,10]. Nevertheless, as the establishment of clean periods for heat stroke could be based on clinical experience instead of data analysis, it presents scenario and epistemic uncertainty. Furthermore, such clean periods may differ according to the time, place, and population, which cannot be confirmed, and therefore have known ignorance and intrinsic uncertainty (Table 2).

Lastly, uncertainties with regard to input data involve the nature of health insurance data and the distinct intrinsic characteristics of the data. First, health insurance data are largely divided into claims data for the general public and care data for recipients of national basic livelihood benefits. Moreover, the definition of medical services encompasses pharmaceutical information and information on emergency department visits. As such,

the definition of medical services based only on claims data, as used in this study, is characterized by statistical and scenario uncertainties, as well as epistemic uncertainty (Table 2). In addition, health insurance data have intrinsic characteristics such as temporal and spatial differences in the billing practices of health-care providers [11]. As such, the application of the EoC concept in this study, which overlooked the intrinsic characteristics of health insurance data, is characterized by recognized ignorance and intrinsic uncertainty (Table 2).

Discussion

This study aimed to apply the concept of EoC to health insurance data in order to establish these as research data for environmental epidemiological studies and to evaluate the level of uncertainty. The reasons for applying the EoC concept in order to use health insurance data for epidemiological studies on disease incidence and morbidity are relatively clear. When utilizing health insurance claims data for epidemiological studies, failure to group separate, multiple uses of medical services for the same disease as one episode may cause serious errors. In particular, as health insurance data are organized based on a monthly billing system, billing for continuous hospitalization in an institution is processed separately as 2 claims if it exceeds a period of 1 month. Such problems become more prominent when a patient is transferred to another institution for the same disease.

As health insurance data are collected with the purpose of claiming insurance fees for medical services, there are intrinsic problems associated with utilizing it as epidemiological data for health impact indicators. However, with appropriate verification of its limitations, strict usage, and validity, health insurance data may be utilized for its features as long-term, consistent, and wide-ranging epidemiological data. Heat stroke is a disease that can be reversed in a short time with proper care and management. Evidently, severe cases may lead to transfer to a small medical institution or prolonged hospitalization. Thus, grouping cases that occurred within a relatively short period as the same episode would be appropriate. As the disease is cured quickly, patients are excluded from the prevalence statistical analysis. In such cases, data that are grouped into medical service episodes will have higher usability as sensitive data for environmental epidemiological studies, such as the effect of climate change on incidence statistics.

In this study, according to the uncertainty assessment, construction of EoC data on heat stroke could be improved in two aspects, namely the definition of medical services and the establishment of clean periods. First, medical services for heat stroke were defined as hospitalization claims in this study. However, medical services can be subdivided and defined more clearly. By defining medical services as visits to the emergency room due to heat stroke and subsequent hospitalization, we think it would be possible to construct more credible data on the effects of environmental risk factor exposure on heat stroke incidence than the current study results. In addition, accuracy at the level of medical service claims can be improved by adding pharmaceutical information that is relevant to medical care.

Second, in this study, clean periods were established based on data characteristics from all age groups in Seoul. However, clean periods can not only be estimated from short-term data but also defined according to the experience of health-care providers. In particular, Winqert et al. [4] reported that construction of EoC is related to the medical care actually received by the patients and so can be influenced by disease severity and the socioeconomic level of the patients. For example, clean periods may differ according to the sex, age, place of residence, and socioeconomic level of the patients. In the future, additional analysis and insights of health-care providers on the results could be used to apply the EoC concept to heat stroke cases.

In this study, many complications were associated with the use of EoC data constructed from health insurance data to correctly determine disease incidence. However, EoC data can be utilized to determine temporal and sequential changes in disease incidence and changes in the short-term effects of environmental risk factors. In addition, the construction of environmental epidemiological research data utilizing health insurance data and the uncertainty assessment in this study have significance, as the process was summarized and evaluated for the first time, despite its frequent utilization in various studies. In the future, results from this study will contribute to determining limitations in the processing of health insurance data and help with the limited usage of processed results in environmental epidemiological studies, while further improving the processing of health insurance data.

Acknowledgements

This study was supported by the Development of Climate-Change Health Impact Assessment and Adaptation Technologies project of the Korea Environment Institute, funded by the Eco-Innovation, Ministry of the Environment, South Korea (no. 412-111-001).

Conflict of Interest

The authors have no conflicts of interest associated with material presented in this paper.

ORCID

Yongseung Shin *http://orcid.org/0000-0002-3985-0366*

References

1. Park BJ. Health insurance claims data characteristics and research utilization considerations. In: Korean Academy of Medical Sciences. Proceedings of workshop for the activation of medical science research; 2007 Feb 12; Seoul. Seoul: Korean Academy of Medical Sciences; 2007, p. 33-48 (Korean).
2. Kim L, Sakong J, Kim Y, Kim S, Kim S, Tchoe B, et al. Developing the inpatient sample for the National Health Insurance claims data. Health Policy Manag 2013;23(2):152-161 (Korean).
3. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. Med Care Rev 1985;42(2):163-218.
4. Wingert TD, Kralewski JE, Lindquist TJ, Knutson DJ. Constructing episodes of care from encounter and claims data: some methodological issues. Inquiry 1995-1996;32(4):430-443.
5. Walker WE, Harremoes P, Rotmans J, van der Sluijs JP, van Asselt MB, Janssen P, et al. Defining uncertainty. A conceptual basis for uncertainty management in model-based decision support. Integr Assess 2003;4(1):5-17.
6. Kim SY. Air pollution effects on asthma according to socioeconomic position [dissertation]. Seoul: Seoul National University; 2006 (Korean).
7. Park CS, Kang HY, Kwon I, Kang DR, Jung HY. Cost-of-illness study of asthma in Korea: estimated from the Korea National Health Insurance claims database. J Prev Med Public Health 2006;39(5):397-403 (Korean).
8. Mehta SS, Suzuki S, Glick HA, Schulman KA. Determining an episode of care using claims data. Diabetic foot ulcer. Diabetes Care 1999;22(7):1110-1115.
9. Kim J, Ha JS, Jun S, Park TS, Kim H. The weather watch/warning system for stroke and asthma in South Korea. Int J Environ Health Res 2008;18(2):117-127.
10. Kim H. Quantitative risk assessment of air pollution effects on health by socioeconomic position [dissertation]. Seoul: Seoul National University; 2010 (Korean).
11. Kim JY. Using national health insurance data to vitalize the evidence-based health care: the current status and tasks. In: Korean Academy of Medical Sciences. Proceedings of workshop for the activation of medical science research; 2007 Feb 12; Seoul. Seoul: Korean Academy of Medical Science; 2007, p. 1-32 (Korean).