RESEARCH ARTICLE

A Comparison of the Cancer Incidence Rates between the National Cancer Registry and Insurance Claims Data in Korea

Hee Jung Seo[1], In-Hwan Oh[2]*, Seok-Jun Yoon[3]

Abstract

   Although much health services research has been conducted using national health insurance claims data in Korea, the validity of this method has not been ascertained. The objective of this study was to validate the use of claims data for health services research by comparing incidence rate of cancers found using insurance claims data against rates of the national cancer registry of Korea. An algorithm to estimate incidence rates using claims data was developed and applied. The claims data from 2005-2008 were acquired and the patients admitted to hospitals due to cancer in 2008 without admission to hospital from 2005-2007 by the same diagnosis code were regarded as incident cases. The acquired results were compared with the values from the National Cancer Registry of Korea. The incidence rate of all cancers found using claims data was 363.1 per 100,000 people, which is very similar to the 361.9 per 100,000 rate of the national cancer registry. Also the age-, gender- and disease-specific rates between the two data sources were similar. Therefore, national health insurance claims data may be a worthwhile resource for health services research if appropriate algorithms are applied, especially considering the cost effectiveness of this method.

Key words: Cancer - incidence rate - national cancer registry- claims data - Korea

Introduction

   Korea has a social health insurance system in which the National Health Insurance Corporation (NHIC) is an exclusive provider of social health insurance (Lee et al., 2008; Oh et al., 2011). The NHIC and complementary medical aid program of Korea have achieved a status of universal health services coverage for Korea (Lee et al., 2008; Jeong, 2011). The insurance claims of both the NHIC and complementary medical aid are paid by the NHIC, so the claims data from the NHIC are representative of medical service usage throughout Korea (Oh et al., 2011).

   But although the NHIC claims data cover almost the entire Korean population, the main purpose for these claims data is for reimbursement (Riley, 2009; Cheng et al., 2011; Oh et al., 2011). Also it is known that identifying an epidemiologic parameter such as the incident rate of diseases by using claims data is difficult (Riley, 2009). Therefore various algorithms have been developed and applied for the use of health services research to cover these shortcomings (Lee et al., 2006; 2011; Park et al., 2006; Yoon et al., 2007; Kim et al., 2008; Oh et al., 2011). Previous studies have used NHIC claims for the study of medical cost, incidence rates, and the prevalence of disease in the Korean population (Lee et al., 2006; 2011; Park et al., 2006; Yoon et al., 2007; Kim et al., 2008; Oh et al., 2011). For example, Lee et al. (2006) determined the incidence of cancer cases using data reflecting the utilization of medical services in a specific year without usage of these medical services during the previous three years. NHIC claims data have been used for the purpose of scientific and clinical research, but there is a paucity of research rigorously assessing the validity of this practice (Lee et al., 2002).

   The Korean Ministry of Health and Welfare has had a hospital-based cancer registry system (Korea Central Cancer Registry, KCCR) since 1980 (Shin et al., 2005, Lee et al., 2010). In 2002, the Korea National Cancer Incidence Database (KNCIDB) was constructed merging the KCCR and other population data bases. Because the KNCIDB data reflected mostly hospital data, the

[1]Department of Public Health, Graduate School of Korea University, [2]Department of Preventive Medicine, School of Medicine, Kyung Hee University,[3]Department of Preventive Medicine, Korea University College of Medicine, Seoul, Republic of Korea
For correspondence: ihoh@khu.ac.kr

cancer cases in this database were mostly confirmed by a physician, suggesting a fairly high accuracy of diagnosis. Other population data such as death certificate also are used as a supplementary data for covering the cancer patients who did not use medical service. This national cancer registry is used as an official source of national cancer data in Korea (Shin et al., 2005; Jung et al., 2010), and was also included in the Cancer Incidence in Five Continents Volume IX (Curado et al., 2008).

Cancer is the most prominent cause of burden in Korea (Yoon et al., 2007), accounting for 1,525 disability-adjusted life years (DALY) per 100,000 annually. The national cancer registry offers nationwide data on cancer incidence that is not available for other diseases. Therefore, the KNCIDB could be regarded as an accurate standard for cancer statistics, although the cancer registry is not flawless (Miller et al., 2009). By comparing the cancer incidence of the KNCIDB and claims data for the NHIC, the validity of using claims data for health services research could be examined (Nattinger et al., 2004; Miller et al., 2009; Cheng et al., 2011). If the accuracy of claims data in terms of cancer statistics was found to be substantial, then the use of claims data for the study of other diseases could also be considered valid. The aim of this study was to investigate the incidence of cancer using two different datasets, the KNCIDB and NHIC claims. Differences in these two datasets by cancer type and age were also examined.

Materials and Methods

This study compared NHIC claims data and KNCIDB registry data in Korea for cancer. The incidence rate of cancer was estimated using NHIC claims data and a variation of an algorithm we developed previously (Yoon et al., 2007; Cheng et al., 2011). Similar algorithms have been used to estimate epidemiologic parameters for various diseases in previous studies (Kim et al., 2008; 2010; Lee et al., 2011; Oh et al., 2012). All cancers were divided into different disease groups according to the international classification of diseases (ICD)-10 diagnosis code groups (WHO, 2012). The groups included lip, oral cavity and pharynx (C00-C14), esophagus (C15), gastric (C16), colon and rectum (C18-C21), liver (C22), pancreas (C25), lung (C33-C34), breast (C50), cervix uteri (C53), corpus uteri (C54), ovary (C56), prostate (C61), bladder (C67), and leukemia (C91-C95). Any other cancers were categorized as "other cancers." When there were multiple reasons for medical service use, the principle cause was used for the classification. For the incidence calculation, following algorithm was applied. First, the patients who admitted the hospital for cancer (C00-C97) in 2008 were selected by using NHIC claims data. Considering the characteristics of cancer, only the hospital admission cases were included for this study. After that, the NHIC claims data of the selected patients diagnosed during 2005-2007 were acquired to determine incident cases. Any patients admitted to the hospital

Table 1 The Incidence of Cancer in Korea for 2008 Determined Using NHIC Claims Data Compared with That Determined Using the KNCIDB (per 100,000 people)

Cancer type	ICD-10 code	Sex	NHIC claims date		KNCIDB	
			Cases	Crude rate	Cases	Crude rate
Lip, oral cavity and pharynx	C00-C14	Total	3,907	7.9	2,557	5.2
		Male	2,364	9.5	1,899	7.7
		Female	1,543	6.3	658	2.7
Esophagus	C15	Total	2,130	4.3	2,181	4.4
		Male	1,915	7.7	1,986	8.0
		Female	215	0.9	195	0.8
Stomach	C16	Total	26,614	53.9	28,078	56.8
		Male	17,947	72.5	18,898	76.3
		Female	8,667	35.2	9,180	37.2
Colon and rectum	C18-C21	Total	20,921	42.3	22,623	45.8
		Male	12,362	49.9	13,536	54.7
		Female	8,559	34.7	9,087	36.9
Liver	C22	Total	15,118	30.6	15,663	31.7
		Male	11,273	45.5	11,776	47.6
		Female	3,845	15.6	3,887	15.8
Pancreas	C25	Total	4,414	8.9	4,320	8.7
		Male	2,525	10.2	2,404	9.7
		Female	1,889	7.7	1,916	7.8
Lung	C33-C34	Total	17,115	34.6	18,774	38.0
		Male	12,288	49.6	13,384	54.1
		Female	4,827	19.6	5,390	21.9
Breast	C50	Total	13,003	26.3	12,659	25.6
		Male	82	0.3	75	0.3
		Female	12,921	52.4	12,584	51.1
Cervix uteri	C53	Total	4,197	8.5	3,888	7.9
		Male	0	0.0	0	0.0
		Female	4,197	17	3,888	15.8
Corpus uteri	C54	Total	1,766	3.6	1,562	3.2
		Male	0.0	0.0	0	0.0
		Female	1,766	7.2	1,562	6.3
Ovary	C56	Total	2,270	4.6	1,816	3.7
		Male	0	0.0	0.0	0.0
		Female	2,270	9.2	1,816	7.4
Prostate	C61	Total	6,072	12.3	6,471	13.1
		Male	6,072	24.5	6,471	26.1
		Female	0	0.0	0	0.0
Bladder	C67	Total	4,076	8.3	3,230	6.5
		Male	3,310	13.4	2,582	10.4
		Female	766	3.1	648	2.6
Leukemia	C91-C95	Total	2,832	5.7	2,561	5.2
		Male	1,622	6.6	1,432	5.8
		Female	1,210	4.9	1,129	4.6
All cancers	C00-C97	Total	179,413	363.1	178,816	361.9
		Male	92,952	375.5	93,017	375.7
		Female	86,461	350.8	85,799	348.1

*NHIC, National Health Insurance Corporation (Korea); KNCIDB, Korea National Cancer Incidence Database

within the same ICD code group during 2005-2007 were excluded because these cases were not considered to reflect cancer incidence for 2008. Disease-specific cancer incidence rates were measured, and total incidence rates were estimated by summing the disease-specific cancer incidence rates. The midyear population for 2008 was

used to calculate crude incidence rates. The midyear population is also used in KNCIDB data (Curado et al., 2008; KSIS, 2008). Age-specific incidence rates were measured for 18 age groups (0-4, 5-9, ….., 85 and older) for all cancers and disease-specific incidence rates were determined for 8 age groups (0-9, 10-19,….. 70 and older). These cancer incidence rates calculated from the NHIC claims data were compared with the KNCIDB's cancer incidence rate for 2008 (Jung et al., 2010). Cancer rates categorized by age, gender and disease were compared by their ICD code group. Every disease's ICD code was identical for both datasets. All statistical analyses were performed using SAS 9.2 (SAS Institute Inc., Cary, NC, USA).

Results

The number of cases and incidence rates of cancer in Korea for 2008 determined using NHIC claims data are presented in Table 1. There were 179,413 estimated cases in 2008, and the incidence rate for all cancer was 363.1 per 100,000 people, 375.5 in 100,000 for males and 350.8 in 100,000 for females. Among the cancer types, the incidence rate of stomach cancer was found to be the highest in Korea with an incidence rate of 72.5 in 100,000 for males, 35.2 in 100,000 for females and 53.9 in 100,000 for the total population. Colon and rectum cancer (42.3 per 100,000), lung cancer (34.6 per 100,000), liver cancer (30.6 per 100,000) and breast cancer (26.3 per 100,000) followed. Among the various cancers, stomach cancer was the leading cause for males, whereas breast cancer was the leading cause for females by incidence (52.4 per 100,000).

The comparison with the results from NHIC claims data to those from KNCIDB data is also shown in Table 1. The incidence rate of cancers from KNCIDB data was found to be 361.9 per 100,000 people, 375.7 in 100,000

for males and 348.1 in 100,000 for females. The two data sets were found to provide similar values for the incidence of cancer (363.1 in 100,000 versus 361.9 in 100,000), although the incidence rates for males was found to be slightly higher when using the KNCIDB data, and the incidence rate for females was higher with the NHIC claims data. The breakdown for the frequencies for the various types of cancer was also found to be similar for both the NHIC claims data and KNCIDB data. Stomach cancer was the most common with the

Table 2. The Incidence of Cancer in Korea for 2008 Determined Using NHIC Claims Data Compared with that Determined Using the KNCIDB according to Age Group (per 100,000 people)

Age group	NHIC claims data		KNCIDB	
	Cases	Crude rate	Cases	Crude rate
0-4	427	18.5	452	19.6
5-9	384	13.3	282	9.8
10-14	485	14.0	409	11.8
15-19	683	20.3	601	17.9
20-24	974	30.2	955	29.6
25-29	2,357	58.2	2,428	60.0
30-34	4,083	102.3	4,426	110.9
35-39	7,423	162.0	7,906	172.6
40-44	11,078	262.4	11,599	274.7
45-49	17,090	390.1	17,737	404.9
50-54	19,720	562.6	19,456	555
55-59	18,420	740.9	18,456	742.3
60-64	20,547	1028	19,664	983.9
65-69	24,262	1269.5	24,285	1270.7
70-74	22,507	1605.6	21,858	1559.3
75-79	15,439	1779.6	15,098	1740.3
80-84	8,536	1792	8,320	1746.7
>85	4,998	1619.5	4,883	1582.3
Total	179,413	363.1	178,816	361.9

*NHIC, National Health Insurance Corporation (Korea); KNCIDB, Korea National Cancer Incidence Database

Table 3 The Incidence of the Five Most Common Cancers in Korea for 2008 Determined Using NHIC Claims Data Compared with that Determined Using the KNCIDB by Age Group (per 100,000 people)

Cancer type	ICD-10 code	Sex	NHIC claims data								KNCIDB							
			0-9	10-19	20-29	30-39	40-49	50-59	60-69	>70	0-9	10-19	20-29	30-39	40-49	50-59	60-69	>70
Stomach	C16	Total	0	0.1	1.9	11.8	38.7	94.3	199.2	284.4	0	0.1	2.2	13.1	42.6	102.5	207.6	290.3
		Male	0	0.1	1.4	11.3	46.1	138	310.5	493.7	0	0.1	1.7	12.4	51.7	150.6	324.7	493.4
		Female	0	0.1	2.4	12.4	31.1	50.4	100.5	163.1	0	0	2.7	13.9	33.1	54.3	103.8	172
Colon and rectum	C18-C21	Total	0	0.2	1.7	6.8	24.3	74.4	159.8	242.4	0	0.2	2	8.4	29.1	83.1	175.9	287.2
		Male	0	0.2	1.5	7.6	28.2	92.2	218.8	351.6	0	0.2	1.9	9.6	34.6	104.4	241.2	408.1
		Female	0	0.1	2	6	20.2	56.4	107.5	179.1	0	0.3	2.1	7.2	23.5	61.8	117.9	217.2
Liver	C22	Total	0.7	0.2	0.8	4.4	23.6	66.5	113.2	137.4	0.5	0.2	0.5	4.8	27.5	71.3	111.9	136.3
		Male	0.7	0.3	1	6.8	39.5	109.3	179.3	46.4	0.6	0.2	0.7	7.4	46.7	118.5	177.1	227.6
		Female	0.6	0.1	0.5	1.9	7.1	23.6	54.6	82.2	0.4	0.2	0.4	2.1	7.5	24	54.1	83.4
Lung	C33-C34	Total	0.1	0.2	0.5	2.3	10.9	43.4	136.1	262.3	0	0.1	0.4	2.4	11.8	47.1	148.7	290.8
		Male	0	0.2	0.5	2.3	12.3	60.1	223	510.5	0	0	0.3	2.4	13.8	67.5	244.2	548.7
		Female	0.1	0.2	0.5	2.2	9.3	26.7	59.2	118.4	0.1	0.1	0.4	2.4	9.8	26.6	64	141.3
Breast	C52	Total	0	0.1	2.2	19.2	57.4	59.5	47.2	27.7	0	0.1	2.4	20.4	56.8	54.7	43.4	28.7
		Male	0	0	0	0.2	0.2	0.5	1.4	2.4	0	0	0	0.2	0.1	0.4	1.4	2.1
		Female	0	0.2	4.5	38.9	116.7	118.8	87.9	42.4	0	0.2	4.9	41.4	115.5	109.1	80.7	44.2

*NHIC, National Health Insurance Corporation (Korea); KNCIDB, Korea National Cancer Incidence Database

KNCIDB data (56.8 per 100,000), colon and rectum cancer followed with a rate of 45.8 per 100,000, lung cancer with a rate of 38.0 per 100,000, liver cancer with 31.7 per 100,000 and breast cancer with 25.6 per 100,000. Again, with the KNCIDB data, stomach cancer was most frequent for males (76.3 per 100,000) and breast cancer for females (51.1 per 100,000). When compared to the NHIC claims data, the incidence rates for these most common diseases, except breast cancer, were higher with the KNCIDB data.

Age-specific incident cases and rates are shown in Table 2. NHIC claims data, the overall incidence rate is generally increasing with ages except 0-4 and 85 and older age group. The number of cancer cases increased sharply beginning with age groups in the forties and was highest in the 65-69 age group for the NHIC data. The incidence rate was highest in the 80-84 age group due to the relatively small population of this age group. KNCIDB data showed a similar pattern to the NHIC claims data. The incidence rate was highest in the 80-84 age group and lowest in the 5-9 age group. When compared to the KNCIDB claims data, the incidence rate determined using NHIC claims data was higher in the 70 years and older age groups.

The age- and disease-specific cancer incidence rates for the five most common cancers are displayed in Table 3. In cases of stomach cancer, from 30-39 age group to older age groups, the incidence rate from KNCIDB is relatively higher than that of NHIC claims data, though the difference is quite small. In the case of colon and rectum cancer, the incidence rate from KNCIDB was greater than that found with NHIC claims from 20-29 age group. A similar trend was also noted in lung cancer. The difference in lung cancer incidence between the results found by the KNCIDB data and NHIC claims data in patients 70 years and older age group was 28.5 per 100,000 people. The two different data sources yielded similar incidence rates for liver and pancreas cancers.

Discussion

In this study, the incidence rates of cancer in 2008 determined using two different data sources, NHIC claims data and the KNCIDB registry, were compared to examine the validity of using claims data in health services research. As a result, the overall rates using the two data are quite similar. The incidence rate of all cancers using NHIC claims data was estimated to be 363.1 per 100,000 people when the algorithms we developed was applied and 361.9 per 100,000 using the KNCIDB. In the most common cancers, which are stomach, colon and rectum, lung, liver and breast, the rates of incidence by age and gender were found to be similar for both data sources.

Though overall feature of incidence rate is similar between two data sources, some difference is still observed. Various reasons could affect theses result apart from the inaccuracy of algorithm used. For instance,

when a patient uses medical care for cancer test and the result was negative finding, if a diagnosis code of claims data is not changed and still remained as cancer code, this could result in differences. Also one of characteristics of incidence pattern is the difference between genders. Namely, the incidence rate of women's disease such as breast, cervix, and ovary cancer was higher in NHIC claims data, but the incidence rate of prostate cancer was higher in the KNCIDB data. Because NHIC claims data is based on the use of medical services, whereas the KNCIDB is based on hospital data in addition to death certificates, women's greater usage of medical services could be the reason for these differences (Ladwig et al., 2000).

In Korea, many health services studies are conducted using NHIC claims data (Lee et al., 2002; 2006; 2011; Park et al., 2006; Yoon et al., 2007; Kim et al., 2008; 2010; Oh et al., 2011, Shin et al., 2012). At the same time, the validity of NHIC claims data, especially the accuracy of diagnosis code had has been debated. Lee et al. compared NHIC claims data and the Kwangju Cancer Registry to estimate the sensitivity of claims data in 1998-1999. In their study, the overall sensitivity for cancer cases was 92.8% (Lee et al., 2002). Another study compared claims data to the medical records of one hospital and found the claims estimated disease correctly in 85% of cases (Ahn, 2002). These studies showed that the ICD code in claims records has a substantial reliability to use for health services research. In the present study, NHIC claims data were used to find incidence rates using our algorithm, and the rates we discovered using the NHIC data were consistent with values found using the Korean National Cancer Registry.

Claims data has been used for health services research in countries than Korea, such as Taiwan. In Taiwan, which is a country with national health insurance coverage of 99% of the population (Lee et al., 2008; Cheng et al., 2011), a study comparing the national health insurance research database and medical records was conducted. They conclude that the national health insurance research database is a valid resource for research due to high accuracy. Even in America, which does not have a universal coverage, various algorithms have been used to validate the usage of claims data (Nattinger. et al., 2004; Penberthy et al., 2005; Miller et al., 2009; Riley, 2009). For example, Medicare claims records, the Surveillance, Epidemiology, and End Results (SEER) Program, and clinical operative reports were compared to classify kidney cancer surgeries (Miller et al., 2009). The algorithm they studied showed that claims records agreed with clinical operative reports in 97% of cases. Another study compared Medicare data with a state cancer registry to identify the incidence of cancer cases (Nattinger et al., 2004). The overall sensitivity ranged from 51-94% based on the algorithms they used to identify breast cancer cases. This study concluded that Medicare data could be a supplement for the disease registry. Another study examined the

algorithm developed to detect the incidence of breast cancer using Medicare claims data (Penberthy et al., 2005). Their developed algorithm showed a sensitivity of 80%, suggesting that Medicare claims data is useful in health services research. Considering the high similarity between the numbers of incident cases identified using claims data and the KNCIDB in the present study, we can conclude that the claims data is a valuable resource for health services research and the algorithm that we developed and used could be validated. Furthermore, the use of claims data to estimate the incidence of other diseases could be worthwhile, especially considering the low cost of this method.

Some limitations should be considered. First, this study did not estimate the accuracy of claims data at an individual level. Because this study compared the incidence of group between two data sources because of unavailability of KNCIDB individual data, individual level comparisons could not be made. Also, because some cancers such as testis, kidney, bladder cancer and lymphoma were classified as "other cancers" in this algorithm, therefore the specific incidences of these diseases could not be presented. Furthermore, a special coinsurance rate was applied for cancer patients in Korea beginning in 2005 (MOHW, 2012) such that cancer patients' pay a 5% coinsurance rate whereas the general coinsurance rate for hospital admission is 20%. Therefore, the accuracy of the diagnosis codes in claims data for cancer patients could be higher than for other diseases. Therefore, the generalization of these results on other diseases could be limited though the diseases that the special rule of coinsurance rate applied have increased. The special coinsurance rate is now applied for patients with cardiovascular disease, cerebral disease, severe burns, and tuberculosis also (MOHW, 2012).

In conclusion, the cancer incidence rates determined using the KNCIDB registry and NHIC claims data were compared, and the ability to identify disease incidence was nearly identical between the two methods. Consequently, it is plausible that using NHIC claims data with these algorithms could serve as a data sources for estimating the incidence rates of other diseases when considering the cost effectiveness of data collection, though the algorithm may need to be improved. To our knowledge, this is the first study that validating the algorithm to identify the incidence case using NHIC claims data. Developing the reasonable algorithm to identify the epidemiologic parameter of diseases using administrative data is promising and should be encouraged especially in national health insurance countries such as Korea.

Acknowledgement

References

Ahn JH (2002). Analysis of agreement status between the diagnostic code of health insurance claim and medical record coding. Master thesis, Inje University.

Cheng CL, Kao YH, Lin SJ, et al (2011). Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. *Pharmacoepidemiol Drug Saf*, **20**, 236-42.

Curado MP, Edwars B, Shin HR (2008). Cancer incidence in five continents. International Agency for Research on Cancer, Lyon.

Jeong HS (2011). Korea's National Health Insurance-Lessons From The Past Three Decades. *Health Aff (Millwood)*, **30**, 136-44.

Jung KW, Park S, Kong HJ, et al (2010). Cancer statistics in Korea: incidence, mortality and survival in 2006-2007. *J Korean Med Sci*, **25**, 1113-21.

Kim SG, Hahm MI, Choi KS, et al (2008). The economic burden of cancer in Korea in 2002. *Eur J Cancer Care (Engl)*, **17**, 136-44.

Kim SY, Yoon SJ, Jo MW, et al (2010). Economic burden of allergic rhinitis in Korea. *Am J Rhinol Allergy*, **24**, E110-3

Korean Statistical Information Service (KSIS) (2008). Midyear population. http://kosis.kr/. Accessed 25 Mar 2012.

Ladwig KH, Marten-Mittag B, Formanek B, Dammann G (2000). Gender differences of symptom reporting and medical health care utilization in the German population. *Eur J Epidemiol*, **16**, 511-8.

Lee JY, Jung KW, Park S, et al (2010). Long-term survival of cancer patients in Korea, 1993-2007: National Cancer Registry Study. *Asian Pac J Cancer Prev*, **11**, 1459-64.

Lee SJ, Im JS, Choi JS, et al (2002). Sensitivity of medical insurance claims data using population-based cancer registry data. *J Korean Soc Med Inform*, **8**, 35-40.

Lee H, Yoon SJ, Ahn HS (2006). Measuring the burden of major cancers due to smoking in Korea. *Cancer Sci*, **97**, 530-4.

Lee SY, Chun CB, Lee YG, et al (2008). The National Health Insurance system as one type of new typology: The case of South Korea and Taiwan. *Health Policy*, **85**, 105-13.

Lee YH, Yoon SJ, Kim EJ, et al (2011). Economic burden of asthma in Korea. *Allergy Asthma Proc*, **32**, 35-40.

Miller D, Saigal C, Warren J, et al (2009). External validation of a claims-based algorithm for classifying kidney-cancer surgeries. *BMC health services research*, **9**, 92.

Ministy Of Health and Welfare(MOHW) (2012). The exemption law of a speical coinsurance rate for cancer patients.http://www.mohw.go.kr/front/jb/sjb0402vw.jsp?PAR_MENU_ID=03&MENU_ID=030405&BOARD_ID=2010&BOARD_FLAG=03&CONT_SEQ=258279&page=1. Accessed 3 Aug 2012.

Nattinger AB, Laud PW, Bajorunaite R, et al(2004). An algorithm for the use of Medicare claims data to identify women with incident breast cancer. *Health Serv Res*, **39**, 1733-49.

Oh IH, Yoon SJ, Seo HY, et al (2011). The economic burden of musculoskeletal disease in Korea: A cross sectional study. *BMC Musculoskelet Disord*, **12**, 157.

Oh IH, Yoon SJ, Yoon TY, et al (2012). Health and economic burden of major cancers due to smoking in Korea. *Asian Pac J Cancer Prev*, **13**, 1525-31.

Park JH, Yoon SJ, Lee H, et al (2006). Burden of disease attributable to obesity and overweight in Korea. *Int J Obes (Lond)*, **30**, 1661-9.

Penberthy L, McClish D, Manning C, et al (2005). The added value of claims for cancer surveillance: results of varying case definitions. *Med Care*, **43**, 705-12.

Riley GF (2009). Administrative and claims records as sources of health care cost data. *Med Care*, **47**, 51-5.

Shin HR, Won YJ, Jung KW, et al (2005). Nationwide cancer incidence in Korea, 1999~2001; first result using the national cancer incidence database. *Cancer Res Treat*, **37**, 325-31.

Shin JY, Kim SY, Lee KS, et al (2012). Costs during the first five years following cancer diagnosis in Korea. *Asian Pac J Cancer Prev*, **13**, 3767-72

World Health Orgranization (WHO) (2012). International Classification of Diseases. http://www.who.int/classifications/icd/en/. Accessed 10 Apr 2012.

Yoon SJ, Bae SC, Lee SI, et al (2007). Measuring the burden of disease in Korea. *J Korean Med Sci*, **22**, 518-23.