

Manuscript title: Utilization of health insurance data in an environmental epidemiology

Type of manuscript: Original article

Authors' name and affiliation:

Jongsik Ha, Seongkyung Cho, Yong Seung Shin

Korea Environment Institute, 370 Sicheong-daero, Sejong-si 339-007, South Korea

Funding sources: None.

Corresponding author's name, academic degree, address, etc:

Yong Seung Shin, Ph.D.

Address: Korea Environment Institute, 370 Sicheong-daero, Sejong-si 339-007, South Korea

Tel: (82) 44 415-7740

Fax: (82) 44 415-7644

E-mail: shiny@kei.re.kr

Short running head: Construction of episode data

Conflict of Interest: The authors have no conflict of interest to declare on this study.

Abstract

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 Statistical Classification of Disease and Related Health Problems, 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.

Key words: Episode of care, Health insurance data, Uncertainty assessment

Introduction

우리나라 건강보험자료(health insurance data)는 전체 국민이 적용 받는 국가에서 제공하는 건강보험제도로부터 생산된다. 이 자료는 전 세계적으로도 유래 없는 자료로서 우리나라의 모든 병원과 약국에서 제공된 의료서비스 관련 정보를 담고 있다. 건강보험자료는 크게 요양기관의 청구(claim) 자료와 의료서비스 대상자에 대한 의료보험 자격(eligibility) 자료로 구분된다. 이중 청구 자료는 요양기관의 의료서비스에 대한 청구와 관련된 자료로서 해당 청구건에 대한 진료 시점 및 진료와 관련된 질병정보를 가지고 있다. 반면에 자격 자료는 의료서비스를 받는 대상자의 의료급여 대상여부, 성별, 나이, 거주지 등에 대한 개인 정보를 가지고 있다.

이에 건강보험자료가 질병 발생(incidence) 또는 유병(prevalence) 자료로 가공될 수 있으면, 동일한 질병을 가진 환자 집단을 대상으로 원인적 요인을 파악하거나, 동일한 요인에 폭로된 사람들에서 발생한 질병의 양상을 파악하는 연구에 활용 될 수 있다. 특히 국가나 지역사회 등 인구집단을 연구단위로 하는 환경역학(environmental epidemiology) 분야의 생태학적 연구(ecological

study)에서는 초기 인과적 관련성(causal relationship) 가설을 유도하는데 매우 유용하게 사용될 수 있다[1]. 실례로 건강보험자료는 생태학적 연구에서 특정 질병의 발생률이 높게 나타나는 인구집단에서 인과적 요인으로 추정되는 위험요인의 발생빈도 비교, 특정한 인구집단에서 시간경과에 따른 특정한 질병의 발생양상과 그 질병의 인과적 요인으로 추정되는 위험요인의 변화 양상간의 관련성 정도를 분석하는데 활용될 수 있다.

하지만 건강보험자료를 환경역학 연구로 활용함에 있어 다양한 문제점이 야기될 수 있다. 건강보험자료의 환경역학 연구로의 활용에서 가장 크게 문제시 되는 것은 청구 자료의 특성상 동일인의 동일 질병에 대하여 여러 개의 청구 건이 발생하게 되고, 이는 여러 번의 질병 발생으로 간주하게 된다는 것이다. 나아가 이는 환경역학 연구에서 환경위험요인 노출로 인한 질병 발생 양상을 파악하는데 한계점으로 작용할 수 있다[2]. 결국 환경역학 연구에 적절한 활용을 위해서는 특정 기간 이내에 발생한 진료에 대해서 동일한 질병 발생 또는 에피소드에 속하는지 아닌지를 결정하는 일정한 기준을 필요로 한다.

보험료 청구 목적의 건강보험자료를 역학연구로 활용함에 있어, Hornbrook

등(1985)은 보건의료 에피소드(health care episode) 개념을 소개하였다[3]. 보건의료 에피소드는 건강상의 문제 발생을 질환(illness), 질병(disease), 진료(care), 건강유지(health maintenance)로 구분하고 이를 통한 건강문제의 시작(begin)과 끝(end)을 구체화하는 것이다. 하지만 실제로 '병인과의 첫 번째 접촉'과 같은 정보로 에피소드의 시작과 끝을 객관적으로 정의하기는 어렵다. 이에 대한 대안으로서 Hornbrook 등은 보건기관의 정규 의료서비스의 시작과 끝을 정의하는 진료에피소드(Episode of Care; 이하 EoC) 도입하였다[3]. EoC는 질환의 발생부터 종료(즉 사망, 치료 종료, 회복 등)까지를 하나의 사건(event)으로 측정하는 단위(unit)를 의미한다.

본 연구는 환경역학 분야의 생태학적 연구에 건강보험자료 활용으로서 EoC 개념을 적용해보는 것이다. 나아가 건강보험자료에 EoC 개념을 적용한 과정과 결과물에 대한 특성 분석 및 향후 EoC 개념 적용의 개선 방향을 제안하는 것이 목적이다.

Materials and Methods

1. Constructing of Episode Data

본 연구에서 환경역학 관련 생태학적 연구에 건강보험자료 활용으로서 EoC 개념 적용은 2003 년부터 2010 년까지의 열사병(ICD10 code: T67)을 대상으로 하였다. 더불어 EoC 개념 적용에 있어 서울지역의 전체 연령집단을 대상으로 하였다. 열사병의 건강보험자료는 국민건강보험공단(National Health Insurance Service)으로부터 받았다. 수집한 열사병 관련 건강보험자료는 청구건에 대한 정보와 의료서비스 대상자의 자격 정보를 포함하고 있는데, EoC 개념 적용과 관련한 항목으로 개인 ID, 거주지, 진료개시일, 진료형태(입원, 외래 등), 주/부상병명, 입/내원 일수 등을 포함하고 있다.

국민건강보험공단으로부터 받은 건강보험자료는 3 단계에 걸쳐 EoC 자료로 구축되었다(Figure 1). 첫 번째 단계로 건강보험자료의 원시자료(raw data)에 EoC 자료를 구축함에 있어 필요로 하는 개인 ID, 성별, 연령, 진료당시 주소, 진료개시일, 주/부상병명 등에 대한 변수들을 가공한다. 두 번째 단계로 대상 질병에 대한 EoC 자료를 구축함에 있어 필요로 하는 의료서비스(health care

service) 및 무진료기간(clean period)을 설정하게 된다. 마지막으로 개인별 EoC 자료 구축, 그리고 시간, 지역, 인구집단이 한정된 자료 추출로 진행된다.

Figure 1 의 EoC 자료 구축과 관련하여 EoC 적용에 있어 가장 중요한 요소는 의료서비스와 무진료기간이다[4]. 의료서비스는 특정 건강문제 또는 질병을 이겨내기 위해 요양기관으로부터 제공된 진료, 약 처방, 입원 그리고 수술에 이르기까지의 일련의 과정으로 정의된다. 반면에 무진료기간은 동일 의료서비스 수진자에 대해 의료서비스가 제공되지 않는 기간을 의미한다. Figure 2 는 개인의 의료서비스 제공에 따른 EoC 개념을 보여준다. 해당 개념도에서는 특정 기간 내에 이루어진 의료서비스를 동일한 EoC 로 판단할 것인가의 여부를 결정하는 기준이 무진료기간임을 보여주는데, 연속하여 동일한 의료서비스가 제공되어도 특정 기간의 무진료기간을 넘어서면 또 다른 EoC 의 발생으로 간주하게 된다.

2. Uncertainty Assessment

건강보험자료에 대한 EoC 개념의 적용 과정 및 결과물에 대한 특성 분석은 불확실성 평가(uncertainty assessment)를 통해 진행하였다. 불확실성 평가는 분석

과정 및 가공 결과물을 인지, 해석, 그리고 결론 내리는 과정에서 잠재적인 불확실성의 정도 및 최종 결과물에 대한 영향 정도를 파악하는데 유용하다.

불확실성 평가를 위해서는 다양한 형태의 불확실성을 종합적으로 평가할 수 있는 개념이 필요하다. 불확실성 평가의 전문가들은 3 차원의 불확실성 평가체계를 제안했는데, 첫째 불확실성의 위치(location), 둘째 불확실성의 수준(level), 그리고 불확실성의 원인(nature)으로 구분하여 평가하였다[5]. 불확실성의 위치(location)는 특정 평가단계에서 불확실성이 나타날 수 있는 지점과 관련된다. 이러한 불확실성의 위치는 문맥의 불확실성(context uncertainty), 모형구조의 불확실성(model uncertainty), 입력 자료의 불확실성(input uncertainty), 그리고 모형결과의 불확실성(model outcome uncertainty)으로 세분화된다. 불확실성의 수준(level)은 지식의 존재(knowledge exists)를 나타내는 완전한 무지에서 완벽한 앎까지의 범위와 관련된다. 불확실성의 수준은 미래 상황이나 결과에 대해 확실성이 있는 경우의 결정(decision)에서부터 통계적 불확실성(statistical uncertainty), 시나리오 불확실성(scenario uncertainty), 인지적 무지(recognised ignorance), 그리고 알고 있지 못하다는 것조차 알지 못하는 완전한 무지(total ignorance)로 세분화된다. 마지막으로 불확실성의 원인(nature)은

지식의 불확실성(epistemic uncertainty)과 가변성의 불확실성(variability uncertainty)으로 세분화된다. 지식의 불확실성은 연구자의 지식이 불완전함으로서 인해 발생하는 불확실성을 의미하며, 가변성의 불확실성(variability uncertainty)은 본질적인 변동성으로 인해 나타나는 불확실성을 의미한다.

불확실성 평가는 불확실성 매트릭스(uncertainty matrix)를 통해 시각적으로 나타낼 수 있다. 불확실성 매트릭스는 불확실성이 어디에 나타나고, 이러한 불확실성의 수준 및 원인이 어떤 특징을 가지는지에 대한 목록을 만드는 것과 같다. 본 연구의 건강보험자료에 대한 EoC 개념의 적용 과정 및 결과물에 대한 특성 분석은 불확실성 매트릭스를 활용하여 정성적 차원에서 불확실성을 확인하고 특성화하였다.

Results

1. Application of the Episode of Care Concept

1.1. Application Process

Table 1 은 본 논문에서의 열사병에 대한 EoC 적용 관련 사항을 정리한 것이다.

첫 번째로 의료서비스 정의와 관련하여 건강보험자료의 2003 년부터 2010 년을 기간으로 청구건의 주상병(main diagnosis) 또는 부상병(sub diagnosis)에 해당 질병의 ICD-10 코드인 T67 이 포함된 입원 청구건으로 한정하였다. 더불어 입원이 가능한 요양기관의 청구건으로 한정하였으며, 기본적으로 청구건에 대한 청구 및 개인 정보에 결측값이 없는 것으로 한하여 가공하였다(Table 1).

두 번째로 무진료기간은 환자의 의료서비스가 이루어지지 않은 기간으로서 본 연구에서는 동일 수진자의 첫 번째 입원 후 퇴원을 하고 다시 입원하는 기간인 '퇴원 후 재입원'까지의 기간으로 정의하였다. 이와 더불어 본 연구에서는 동일인의 입원 청구건에 대한 '퇴원 후 재입원 기간'을 무진료기간으로 활용함에

있어, 개인별 입원 의료서비스의 연속성 여부를 확인하여 무진료기간을 수정·보완하였다. 세부적으로는 개인별 입원 청구건들의 퇴원 후 재입원 간격에 대한 빈도 분석(frequency analysis)을 통해 동일한 입원으로 묶을 수 있는 최대 일수(days)를 평가하였다. 더불어 의료서비스 정의별 최대 일수에 대한 민감도 분석(sensitivity analysis)으로서 의료서비스 정의와 관련된 주/부상병 및 요양기관 종별 정의에 따라서 최대 일수를 평가하였다(Table 1).

1.2. Application Results

본 논문에서 EoC 적용은 의료서비스를 입원으로 한정하여 진행했다. 이 경우에 무진료기간의 적용은 입원기간이 아닌 기간으로 비교적 명확해지지만 연속된 입원에 대한 분리된(separated) 청구건을 구별해 내는 것이 중요해진다. 이에 본 논문에서의 적용결과는 개인별 입원 청구건들의 퇴원 후 재입원 간격에 대한 빈도분석 결과에 중점을 두어 서술하였다.

Figure 3 은 서울지역 전체연령 열사병 입원 청구건의 퇴원 후의 재입원 간격에 따른 전체 청구건의 감소 정도를 도식화한 것이다. 열사병 입원 청구건의 퇴원 후 재입원 간격은 0 일에 가장 많은 분포를 가지고 이후에는 거의 변화를 보이지 않고 있다. 이러한 특성은 주/부상병명에 따른 질병 및 요양기관 정의에

상관없이 동일한 경향을 보였다(Figure 3). 이에 본 논문에서는 열사병 입원 의료서비스에 대해 첫 퇴원 이후 재입원 간격이 0 일인 경우에는 동일한 청구건으로 간주하여 무진료기간을 설정하였으며, 이에 근거하여 EoC 개념을 적용하였다.

Figure 4 는 서울지역 전체연령에 대한 열사병 입원 일별 청구건수와 EoC 자료구축 결과를 보여주고 있다. Figure 4 의 좌측은 EoC 를 적용하기 전이며, 우측은 EoC 를 적용한 후이다. EoC 적용 전후에 대한 시계열적 변화를 명확히 확인할 수는 없지만, EoC 적용 전에 비해 적용 후에 일별 건수의 전체적 변동(deviation)이 감소한 것을 확인할 수 있다.

2. Uncertainty Assessment for Episode Data of Heat Stroke

Table 2 는 열사병에 대한 EoC 적용 과정에서 나타날 수 있는 불확실성에 대한 불확실성의 위치, 수준, 그리고 원인을 불확실성 매트릭스로 작성한 결과이다.

문맥상의 불확실성(context uncertainty)은 EoC 의 적용과 관련하여 의료서비스 정의와 관련된다. 지금까지 건강보험자료를 활용한 다수의 환경역학 연구가 수행된 바 있는데, 특정한 기준에 근거하여 가공된 것은 아니다. 김선영 등(2006)은 일별 대기오염수준의 천식 급성(acute asthma) 건강영향평가라는 연구에서 천식 급성의 질병 발생(incidence) 여부를 야간 방문, 공휴일 방문, 응급실 방문의 청구건을 대상으로 추출(입원 및 외래)하고, 이후 과거력(medical history)을 이용하여 과거력이 있는 경우에 제외하는 과정으로 질병 발생여부를 정의한 바 있다[6]. 하지만 박춘석 등(2006)은 천식 유병률 확인을 통한 질병부담(disease burden) 추정이라는 연구에서 천식의 유병률을 추정하기 위해서 2 회 이상의 천식 청구건(주상병 또는 부상병)과 천식 관련 치료약제가 처방된 경우로 한정하고, 기타 세부적인 조건들에 해당하는 청구건은 제외하는 방식으로 가공한 바 있다[7]. 일반적으로 건강보험자료의 가공에 있어 의료서비스 조건은 해당 자료를 이용하여 활용하고자 하는 목적에 따라서 달라진다.

본 논문의 Table 1 에서 적용한 열사병에 대한 EoC 개념은 에피소드 자료의 한 사례라고 할 수 있다. 특히 환경위험노출에 따른 질병 발생 여부를 확인하는 목적에서의 의료서비스 정의는 실제 열사병 발생 여부에 대한 불확실성을 야기할

수 있다. 열사병 발생으로서 의료서비스 정의는 입원 및 외래 환자, 주상병 및 부상병 진단코드, 그리고 의료서비스를 제공한 요양기관 종류에 따라서 다양하게 달라질 수 있다. 또한 약제정보, 응급실 방문 여부 등의 추가적인 정보를 통해 더욱 세분화될 수 있다. 이에 의료서비스 정의와 관련된 불확실성들은 시나리오적인 불확실성, 그리고 지식적인 불확실성의 특성을 가진다(Table 2).

모형구조의 불확실성(model uncertainty)은 EoC 의 적용과 관련하여 무진료기간의 설정과 관련된다. 건강보험자료에 EoC 의 적용은 다양하게 제시될 수 있는데, Wingert 등(1995)은 31 개 상병에 대해서 특정 질병별, 활용 목적별로 EoC 구축 방법을 여섯 가지로 구분하여 설명한 바 있다[4]. 특히 EoC 개념 적용에서의 질병의 특성은 질병의 완치 가능성, 나아가 무진료기간의 설정과 관련된다. 예를 들어 당뇨병처럼 한번 질병에 이환되면 평생 완치가 불가능한 경우에는 최초 발생이후에 이루어진 모든 당뇨병 의료서비스 이용이 하나의 에피소드라고 볼 수도 있다. 반면 당뇨병 그 자체가 아니라 혈당조절의 실패로 저혈당증으로 쇼크에 빠져서 입원하는 것만 본다면 이런 것으로 한 사람에게 여러 차례 다른 에피소드로 반복될 수 있다[8].

본 논문의 Table 1 에서 적용한 열사병에 대한 무진료기간의 설정은 의료서비스 정의와 연계하여 설정한 것이다. 즉 열사병 발생여부는 입원 의료서비스의 청구 여부로 정의되기 때문에 이에 따른 무진료기간은 입원 의료서비스가 제공되지 않은 기간으로 정의할 수 있다. 하지만 개인별 입원 의료서비스의 연속된 입원인지 아닌지 여부를 실제 청구건 자료의 특성을 분석하여 최대 0 일 이내로 설정하였다. 실제로 무진료기간 설정은 김재용(2008)의 기상변화에 따른 건강영향평가 및 김호(2010)의 대기오염 변화에 따른 건강영향평가 연구에서 입원 에피소드(실제 입원개시~입원종료) 구축 시에 개인별 입원 의료서비스의 연속된 입원인지 아닌지 여부를 각각 최대 2 일, 최대 0 일 이내로 정의한 사례가 있다[9, 10]. 그럼에도 불구하고 열사병에 대한 무진료기간 설정은 자료 분석 이외에 임상적 경험에 근거할 수 있기 때문에 시나리오적인 불확실성 및 지식적 불확실성의 특성을 가진다. 더불어 이러한 무진료기간은 시간적(time), 공간적(place), 그리고 인구집단별(population)로 다를 수 있으나 확인할 수 없었기 때문에 인지된 무지의 불확실성 및 본질적인 불확실성의 특성을 가진다(Table 2).

마지막으로 입력(input) 자료와 관련된 불확실성은 건강보험자료 관련 정보 및 본래 가지고 있는 특성과 관련된다. 먼저 건강보험자료는 크게 의료서비스에 대해 일반인을 대상으로 하는 청구 자료(claim data)와 기초생활수급자를 대상으로 하는 보호 자료(care data)로 구분되며, 의료서비스를 정의함에 있어 약제 정보(pharmaceutical information), 응급실 방문 정보 등을 포함하고 있다. 이에 본 논문에서 청구 자료만을 활용하여 정의한 의료서비스는 각각 통계적인, 시나리오적인 불확실성, 그리고 지식적인 불확실성의 특성을 가진다(Table 2). 더불어 건강보험자료는 의료공급자의 시공간적인 청구관행 차이 및 진료 형태의 차이 등과 같은 본질적인 특성을 가지고 있다[11]. 이에 본 논문에서 건강보험자료의 본질적인 특성을 고려하지 못한 EoC 개념의 적용은 인지된 무지의 불확실성 및 본질적인 불확실성의 특성을 가진다(Table 2).

Discussion and Conclusion

본 논문의 목적은 건강보험자료에 EoC 개념을 적용하여 환경역학 연구자료로 구축하고 불확실성 정도를 평가해 보는 것이다. 이러한 목적과 연계하여 건강보험자료를 질병 발생 및 이환 여부에 대한 역학 자료로 활용하기 위한 EoC 개념의 적용 이유는 비교적 명확하다. 건강보험자료의 청구건 자료를 역학 연구로 활용함에 있어 동일한 질병으로 서로 분리된 여러 차례의 의료서비스 이용이 발생하는 경우에 이를 하나의 건수로 간주하지 않는다면 심각한 오류에 빠질 수 있기 때문이다. 특히, 건강보험자료는 월별 분리청구를 하고 있어서 하나의 요양기관에서 연속된 입원 청구건이라고 할지라도 한 달을 넘어가면 2 건 이상으로 분리 청구한다. 만일 같은 질병으로 다른 요양기관으로 전원(transfer)된 경우라면 이러한 문제는 더 심각해진다.

건강보험자료는 진료에 대한 보험료 청구 목적으로 수집된 자료이기 때문에 건강영향 지표의 역학 자료로 활용하는데 있어 본질적인 문제가 있다. 하지만 충분한 제한점과 엄밀한 사용, 그리고 타당성 검정 등을 거친다면 건강보험자료의 특징인 장기성, 일관성, 광역성을 갖춘 역학 자료로 활용될 수 있을 것이다. 본

논문에서 보고자 하는 열사병의 경우에는 적절한 대처 및 관리를 받는다면 단기간 내에 질병 이전의 상태로 돌아갈 수 있는 질병이다. 물론 상태가 심각하여 작은 의료기간으로 전원을 하거나 며칠씩 입원하는 경우도 있을 수 있다. 이러한 점을 고려한다면 비교적 짧은 기간 내의 발병건들을 동일한 에피소드로 묶는 것이 바람직하겠다. 단기간 내에 질병이 완치되므로 환자는 유병률 통계에서 빠지게 되며 이런 경우에 진료에피소드로 묶어낸 자료는 발생률 통계서 기후변화와 같은 환경역학 연구로 민감하게 반응하는 자료로 활용성이 높아질 것이다.

본 논문에서 적용해본 열사병의 EoC 자료구축은 불확실성 평가와 연계하여 의료서비스의 정의와 무진료기간의 설정인 두 가지 측면에서 개선해 볼 수 있다. 첫째, 본 논문에서 열사병에 대한 의료서비스는 입원 청구건 발생으로 정의하였다. 하지만 의료서비스는 세분화하여 더 명확하게 정의될 수 있다. 만일 열사병으로 인한 응급실 방문과 이로 인한 입원으로 정의한다면 현재의 연구결과보다 환경위험요인 노출로 인한 열사병 발생에 대한 신뢰성 있는 자료 구축이 가능할 것으로 사료된다. 이와 더불어 의료서비스의 정확한 청구 정도에 대해서는

의료서비스에 진료 관련한 약제정보를 추가함으로써 정확도를 높일 수 있을 것으로 사료된다.

둘째, 본 논문에서는 무진료기간을 서울지역의 전체연령을 대상으로 자료 특성을 바탕으로 설정하였다. 하지만 무진료기간은 단일의 자료에서 추정될 수 있을 뿐만 아니라 임상 의료인의 경험에 의해서도 정의될 수 있다. 특히 Wingert 등(1995)은 EoC 구축은 환자의 의료서비스 수진과 관련되는 것으로서 질병의 경중과 더불어 환자의 사회경제적인 수준에 의해서도 영향을 받을 수 있다고 보고한 바 있다[7]. 예를 들어 환자의 성별, 연령대, 거주지, 그리고 사회경제적 수준 등에 의해서 무진료기간은 다르게 나타날 수 있다. 향후 이에 대한 추가적인 분석 및 분석결과에 대한 임상 의료인의 경험을 바탕으로 열사병에 대한 EoC 를 적용해 볼 수 있겠다.

본 논문의 건강보험자료를 활용한 EoC 자료 구축은 질병 발생률을 정확히 파악하는 목적으로는 아직 많은 무리가 있다. 하지만 EoC 자료를 활용하여 질병 발생에 대한 시계열적 변화와 환경위험 요인에 대한 단기적인 영향 변화를 파악하는데 활용될 수 있다. 더불어 본 논문의 건강보험자료를 활용한 환경역학

연구자료 구축과 이에 대한 불확실성 평가는 지금까지 상당한 연구 자료로의 활용에도 불구하고 활용 과정에 대해 처음으로 정리해보고 이를 평가해봤다는 것에 의미가 있다. 향후 환경역학 연구에서 건강보험자료의 가공에 대한 한계 및 가공 결과물에 대한 제한된 사용에 도움이 될 것이며, 나아가 개선된 건강보험자료의 가공에 활용될 수 있겠다.

Acknowledgement

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).

References

- [1] Park BJ. Health insurance claims data characteristics and research utilization considerations. In: Korean Academy of Medical Sciences. Workshop for the activation of medical science research; 2007 Feb 12; 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] Park BJ. Health insurance claims data characteristics and research utilization considerations. In: Korean Academy of Medical Sciences. Workshop for the activation of medical science research; 2007 Feb 12; Seoul: Korean Academy of Medical Sciences; 2007, p. 33-48 (Korean).

Tables

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

Steps	Detailed contents
	<ul style="list-style-type: none">-Period: 2003/01/01 - 2010/12/31-Diagnosis code: T67(ICD-10 code) in main or subdiagnosis
Health-care service	<ul style="list-style-type: none">-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	<ul style="list-style-type: none">-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

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 v.s. inpatient		+ ¹		+	
	Claim in main and/or subdiagnosis		+		+	
	Claim in specific medical care institutions		+		+	
Model	Methodology for the setting of clean period					
	Clinical experience v.s. data analysis		+		+	
	Variation by time, place, and person of clean period			+		+
Inputs	Information of health insurance data					
	Medical claim data v.s. care data	+			+	
	Utilization of additional information		+		+	

Claim characteristics of health insurance data		
Spatial-temporal characteristics by health care provider	+	+
Change of treatment patterns	+	+

¹ The sign of + indicate whether the characterization is or not in the uncertainty matrix.

Figure legends

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

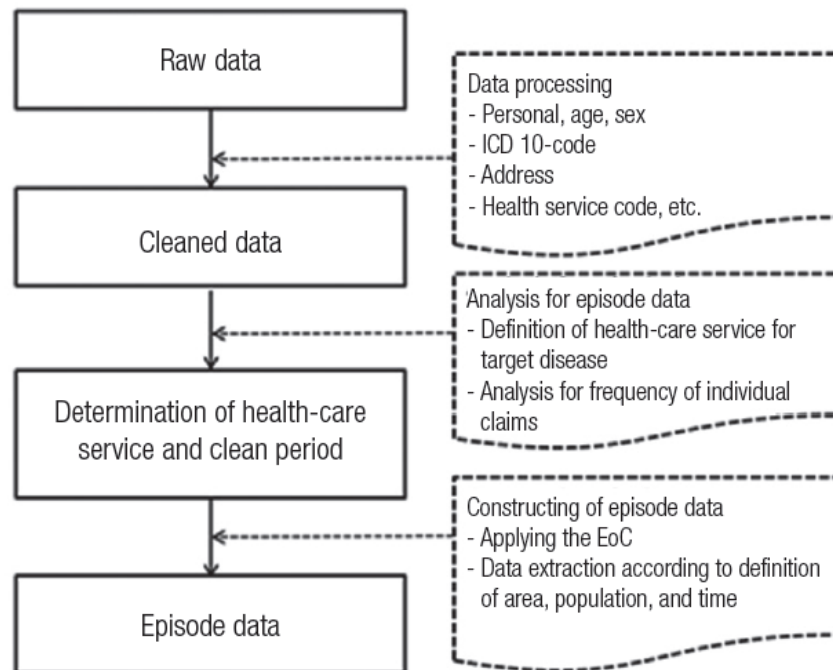


Figure 2. The episode of care concept

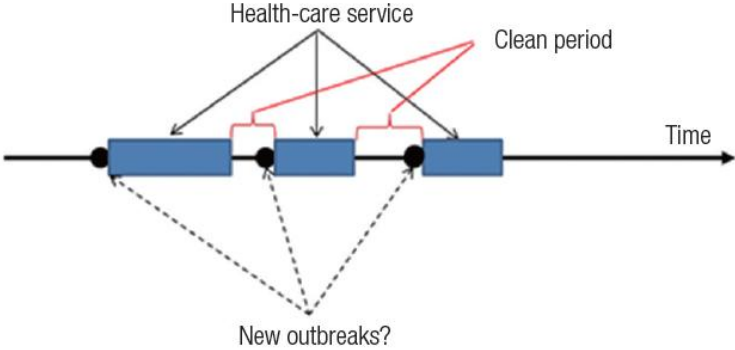


Figure 3. Relative number of claims according to interval of rehospitalization after the first discharge (days). All diagnosis means that the claim of heat stroke at discharge is the main diagnosis or subdiagnosis

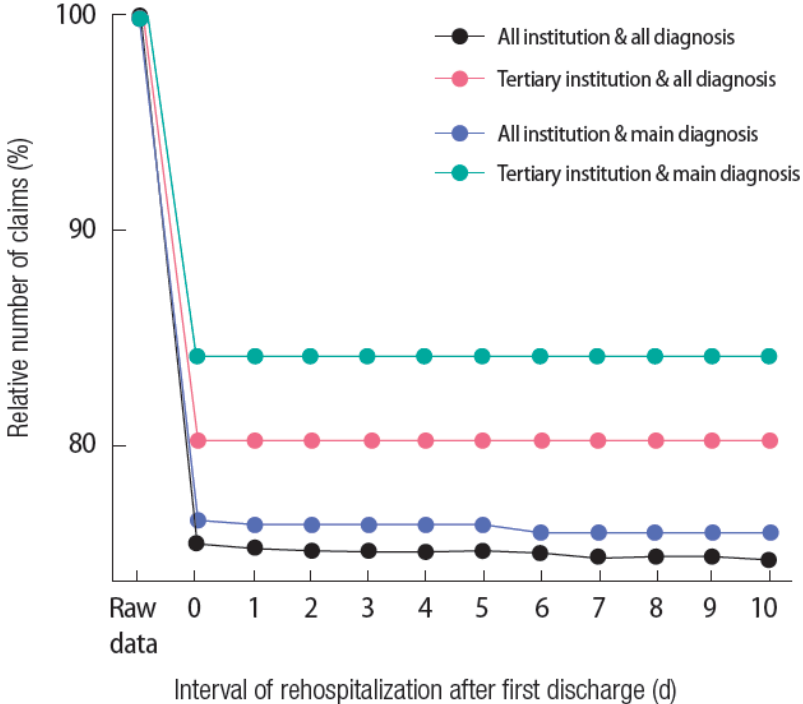
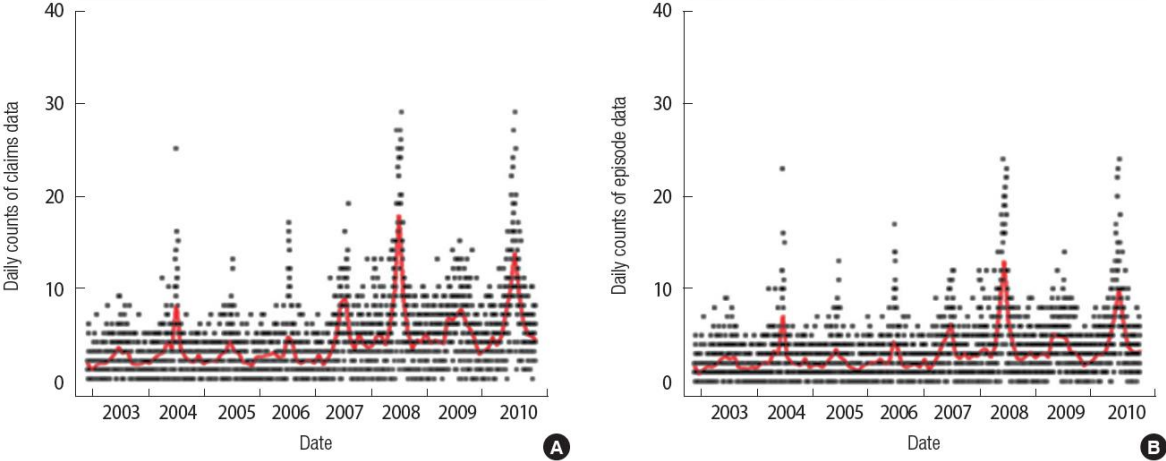


Figure 4. Time series trends of claims (A) and episode (B) data before and after application of the episode of care concept to heat stroke.



Abbreviations: None.