

주요용어 : 보건의료서비스 이용평가, 데이터웨어하우스, 데이터마이닝

일개지역의 보건의료서비스 이용 평가 - Y지역의 대학병원과 보건소 데이터베이스를 통하여 -*

안성희**, 이병화***

I. 서론

1. 연구의 필요성

우리나라의 의료공급체계는 다른 OECD국가들에 비해 사회보장을 포함한 공공부문이 차지하는 비중이 적어 국민의료비 문제가 심각한 수준이다. 향후 소득증가세의 둔화와 노인인구의 증가(보건복지부 2003; OECD 2003), 신기술의 도입 등에 힘입은 국민의료비 증가세가 결합할 경우 의료비 상승은 더욱 증폭될 상황에 있다(신영석, 신형웅, 신종각, 1999). 최근 한 조사결과(최병호와 신형호, 2000)에 의하면, 전 국민의료보험 실시 이후 의료비 가계본인부담 비중이 절반 정도를 유지하는 가운데도 급속한 의료비 지출증가를 보고하였는데 이는 일차 의료공급자들이 환자에 대한 고가의 서비스나 불필요한 중복 서비스를 제공한 결과이며(보건복지부, 2003) 다른 한편, 질낮은 공공의료 서비스에 대한 국민 평가라고 볼 수 있다.

사회보험의 위험보장수단이 이와 같이 약화되는 문제는 공공정책의 효율성, 형평성, 그리고 국가규제수단의 실효성 확보의 실패를 의미한다. 이러한 보건의료분배 문제는 우리나라뿐만 아니라 세계적인 공통 과제로서 보건의료체계 내에서 전문병원들과 공공의료기관이 협력하여 규제된 경쟁(Regulated Competition) 아래 계획시장(Planned Market)을 구현하고(Ham & Maynard, 1994; Van de Ven & Fschut, 1995), 컴퓨터에 누적된 거대한 정보 활용가능성을 이용하여 생산성 향상의 방안을 모색하는 배경이 되었다. 미국의 한 HMO연구기관에서는 지역단위 의료기관의 데이터베이스를 통하여 지역단위 의료기

관 성과지표를 추출하고 계획된 의료서비스의 모형을 검증하여 이상적인 모형을 제시한 바 있다(Deborah, Bradford & Northridge, 2003; Mark, Nicole & Matthew, 2005).

보건정책을 분석하기 위한 정보원천들은 지역사회 보건사업 기관에 누적된 주민 건강정보 및 연계된 외부정보들로서, 정보시스템을 통하여 다양한 환경과 상황의 정보 분석 결과를 유용한 지식으로 전환 보관함으로써 지속적인 보건사업 향상의 전략적 도구로 사용할 수 있다(Sindell, 2001). 따라서 보건복지부는 건강증진 사업이 본격화되기 시작한 2000년대 초반부터 확대된 건강사정 영역과 질적 사업관리 영역을 감당하기 위해서 보건 분야 지식경영시스템(Knowledge Management System; KMS)(Duffy, 2001; Gardner, 2003)의 도입을 진행해 왔으며(보건복지부, 2003), 최근에는 의료서비스의 질과 효율성을 높이기 위한 방안으로 민간, 공공 보건의료기관 소유의 DW와 데이터베이스를 통합하고 통합 의료지식 인프라를 구성하려는 전자건강기록시스템(Electronic Health Record System; EHRS) 구축사업에 100억 이상의 예산을 투자하고 있다. KMS와 EHR사업의 성공여부는 이천조 이상규모의 세계의료정보시장의 경쟁구도를 결정할 정도로 중요한 사안으로 금년부터 정보통신부, 산업자원부까지 막대한 투자로 가세하고 있는 실정이다.

KMS는 조직 내외에 산재한 다양한 정보를 수집하여 이를 체계적으로 분류, 가공, 축적하여 필요한 정보를 적기에 습득하게 하는 방법으로 조직 구성원들의 지적 전문성을 체계화시켜 정보와 지식을 조직차원에서 보다 효율적으로 활용할 수 있도록 하는 전략적 대안과 해결책을 제공한다. KMS는 통합되고 체계화된 지식창고인 데이터웨어하우스(Data Warehouse; DW) 기반하에 핵심성과지표(Key Performance Index; KPI)를 사용한다. 보건의료분야의 KPI는 지역단위의 여러 보건의료 조직들이 함께 추구하는 핵심목표이며 의사결정에 사용되는 핵심정보들로서 보건의료 문제점과 해결지점을 파악하

*이 논문은 2005년 가톨릭중앙의료원 성의장학학술연구비에 의하여 이루어진 것임

**가톨릭대학교 간호대학 부교수

***연세대학교 보건대학원 외래강사

<표 1> Y 지역소재 대형요양기관 진료비 지급 현황(2000.1~2004.12)

요양기관종류	진료건수(%)	총진료비(천원)(%)
종합병원	1,835,511 (58.0)	341,807,070 (91.7)
입원	136,057 (4.3)	223,399,634 (59.9)
외래	1,699,454 (53.7)	118,407,435 (31.8)
의원	1,222,553 (38.6)	27,451,613 (7.4)
치과병원	25,142 (0.8)	999,628 (0.3)
한의원	81,368 (2.6)	2,478,265 (0.7)
계	3,164,574 (100.0)	372,736,576 (100.0)

고 인구의 보건향상을 위한 지속적인 사업수행도구이다. 주요 선진국에서는 지역사회 중심 건강증진 사업 평가시스템을 구축하기 위하여 공공 및 민간 건강 서비스 기관의 DW와 CDR(Clinical Data Repository)를 통합하여 데이터가공과 질 의도구를 통하여 OECD, ISO에서 제시한 지표목록의 표준항목 세부지표를 추출하고 지역별, 시간별 추이를 비교하여 보건 자원 분배의 근거로 활용하고 있다. 우리나라에서는 병원, 의료보험 공단 등 의료기관 조직내부에서의 경영분석을 위하여 DW를 활용하고 있으나 지역단위의 보건의료정책 분석에 활용한 경우는 거의 없는 실정이다.

따라서 일개지역에 속한 주요보건의료기관의 데이터베이스에 누적된 보건의료서비스 이용 자료를 대상으로 지식분석방법을 적용함으로써 정확한 보건정책수립을 위한 정보를 생산하고자 지역사회 건강증진사업 평가연구를 수행하였다.

2. 연구목적

이 연구는 서울시 일개 지역을 대상으로 대학병원과 관할 보건소의 데이터베이스에 누적된 약 5년간 건강사업 자료를 대상으로 의료이용결과를 분석하고 효과적인 보건정책 방향을 제시하고자 시도하였다.

- 1) 일개지역의 대학병원과 보건소의 건강증진사업평가를 위한 데이터 마트를 구축 한다
- 2) 다차원 분석기법과 지식분석방법(데이터마이닝 기법, 연관분석방법)을 이용하여 보건의료서비스 이용결과를 평가한다.

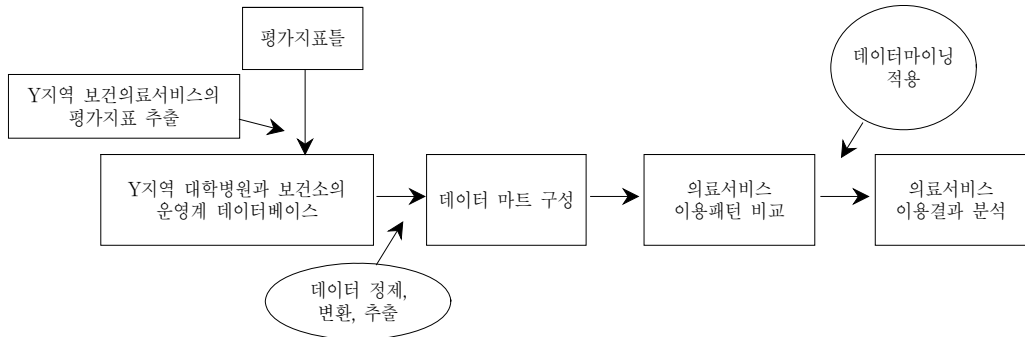
II. 연구방법

1. 연구대상

서울 Y지역은 규모 약 850만², 상주인구 약 30,090명, 세대 수 약 9,660으로 구성된 행정구역상 동단위에 해당한다(건설교통부, 2004). 이 지역의 의료기관 및 요양기관 수는 종합병원 1개소(91.7%), 의원 34개소(59.9%), 한의원 10개소(0.7%), 치과병원 2개소(0.3%), 기타 치과의원 2개소 약국 39개소이다. 이 중 Y 지역의 의료비의 58%를 차지하는 700침상규모의 대학병원과(국민보험공단, 2005)<표 1> 이지역의 관할지역인 Y보건소의 운영계 데이터베이스(Operational Database Storage: ODS)의 건강 및 행정자료 중 2000년 4월부터 2004년 4월까지 5년간 누적된 120만여 건의 자료를 대상으로 하였다.

2. 자료분석

Y지역의 건강증진사업의 평가연구에 사용된 절차는 이병화(2004)가 The Community Health Improvement Process (CHIP)모형과(Institute Of Medicine; IOM, 1991; Green & Kreuter, 2003), 지표이론으로부터 도출한 보건사업평가를 위한 지표들을 사용하여 데이터 충실도에 따라 분석에 유용한 지표를 선정하고 다차원 분석(Multi Dimensional Analysis)과 지식분석 방법을 적용하여 Y지역 건강사업 결과를 평가하였다. 다차원 분석은 병원과 보건소의 건강사업결과 추이분석을 위하여 수행결과와 관련된 영향요인에 따른 다차원 큐브를 구성하고 결과를 그래프(Pilot system)로 구현하였다. 지식분석은 데이터마이닝 기법(Cabena P, 1997)의 군집분석과 연관분석방법을 적용하였다. 연구의 전체적인 흐름도는 <그림 1>과



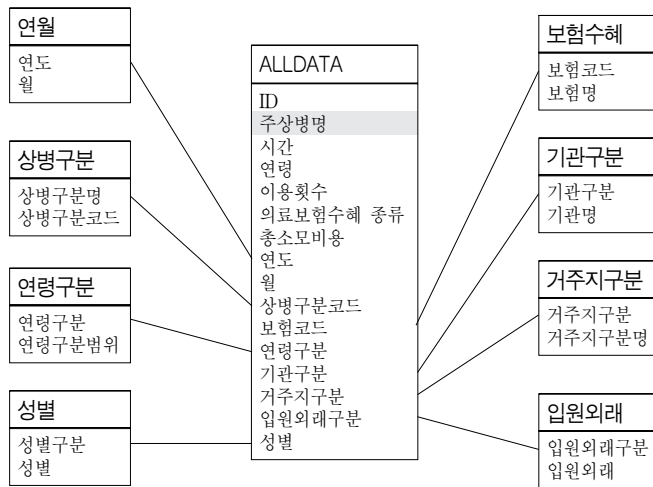
<그림 1> 연구의 흐름도

같다. 이러한 과정에 사용한 분석 도구는 Microsoft SQL(Structured Query Language) Server 2000에 포함된 OLAP(OnLine Analytical Processing) 프로그램과 추출, 변환, 전송(ETT; Extraction, Transformation, Transportation), 자료전송 서비스(DTS; Data Transformation Service) 패키지이며, 데이터 마이닝 분석과 결과 구현을 위해 SAS 9.1의 Enterprise minor, Microsoft Excel 2003을 이용하였다.

1) 자료추출 방법

Y 지역 대학병원과 관할 보건소의 ODS에서 보건지표 분석에 필요한 항목을 추출하고 데이터를 분석목적에 맞도록 정제하였다. 보건사업의 평가지표 틀에서 보건소 인구보건사정 영역의 건강결정인자와 사업성과 모니터 영역의 보건제도 성과,

서비스 특성, 보건사업 이용의 세부항목들을 사실테이블로 구성하는 데이터 모델링을 하여 건강사업 평가를 위한 데이터마트를 구성하고 스타스키마를 제시하였다<그림 2>. 또한 보건소 건강사업을 평가하기 위한 데이터 마트를 스타 스키마로 제시하였다. 이는 실측값을 지닌 1개의 사실 테이블을 중심으로 요약 정보를 지닌 7개의 차원 테이블이 별모양으로 연결된 개념도이다. 그림에서 제시한 차원 테이블의 데이터 정제과정을 설명하면 상병명(ICD 10), 거주지역은 우편번호에 따라 Y 지역을 보건행정 구역별로 나누고 거리에 따라 서열화하였으며, 합병증 진단명, 성별, 나이를 코드화 하였다. 나이는 인간의 사회 심리적 발달주기(Erickson, 1975)와 자료 충실도를 고려하여 그룹화 하였으며 연령 구분의 하위그룹은 개인 연령이 해당된다. 분석대상 중 종속변수는 총이용비용, 이용횟수, 질환명을 사용하였다



<그림 2> 일개지역건강사업평가를 위한 데이터마트

2) 군집분석과 연관성 분석

병원과 보건소의 서비스 이용특성을 파악하기 위하여 군집 분석방법을 적용하였다. 군집분석 중 K-평균 군집화(K-Mean Clustering)방법은 독립변수에 따라 종속변수를 동질성(Homogeneous)있는 그룹으로 분류하여 그 속성을 파악하는 방법이다(Wears, 2002). 군집의 수 K값은 군집간 거리가 최대화되는 수준을 파악하여 3으로 정하였다. K군집분석을 위해 컴퓨터가 수행한 데이터 마이닝 분석과정은 먼저 대상 자료 76,807개 중 무작위 선택방법으로 초기 자료(Seed Data) 2000개를 선택하여 임의값 3개의 중심 값을 선정하였다. 그 후 이 3개의 중심 값을 기준으로 각 측정값과 직선거리가 짧은 순으로, 그리고 각 군집에 대한 독립변수의 상대적 기여도(총이용비용, 이용회수, 연령, 거주지, 성별)에 따라 나머지 측정값을 재배정하여 최종 군집을 결정하고 분류하였다. 독립변수를 지역주민의 경제적 접근성(비용), 물리적 접근성(거주지)(대한간호협회 보건간호사회, 2001)과 연령(이 주영, 이순영, 박재용, 조홍준, 송태민, 2000)을 이용하였으며 이에 따라서 보건소 이용도와 지역주민 질환 특성이 달라지는가 보기 위해 군집분석을 적용하였다.

질환 연관성 분석은 신뢰도(Confidence), 지지도(Support), 향상도(Lift)로 표시하였다. 신뢰도는 조건부 확률로서 질병 A가 발생했을 때 질병 B를 포함하는 확률이며 지지도는 두 질병이 함께 발생할 확률을 의미한다. 향상도는 신뢰확률을 기대

신뢰확률로 나눈 값으로 정의하며 향상도 값이 1보다 크면 클수록 품목 간에 양의 상관관계가 있음을 의미한다.

III. 연구결과

1. Y지역 병원 및 보건소의 사업 우선순위

Y지역의 3차 의료기관인 대학병원과 보건소의 최근 5년간 상위 10순위 질환을 분석한 결과 대학병원 이용자의 주요 질환은 수익성 위주의 비교적 다양하고 구체적인 질환자의 이용이 높았고, 보건소는 만성질환의 의존도가 높았다. 이들이 전체 빈도수에 차지하는 비율은 병원은 22.4%인 반면 보건소의 경우 59.0%를 차지하였다<표 2>. 병원과 보건소를 이용하는 질환자의 질병특성은 차이가 있었는데 병원은 암질환, 신체상해, 증상관리와 진단을 위한 서비스 이용비용이 높은 순위를 차지하였다. 이에 비해 보건소는 감기와 고혈압, 당뇨, 관절증, 치아질환, 여성의 골반내 염증 관리를 위한 이용 빈도가 높았다.

2. 거주지역에 따른 의료보험 수혜 형태

Y지역 만성질환 대상자들의 병원, 보건소 선택요인을 비교한 결과 모두 거주지에 따라서 의료보험 수혜형태에 유의한 차이가 있었다(p= .000)<표 3>< 표 4>. 분석대상인 Y지역의

<표 2> 병원, 보건소의 다빈도 이용질환 순위(2000~2004)

순위	병원(상위 22.4%)		보건소(상위 59.0%)	
	질환명	빈도(%)	질환명	빈도(%)
1	신생아 관리(Z38.0)	4615(6.3)	탄광부진폐증(J60)	2756(17.6)
2	기타 및 상세불명의 복통(R10.4)	3554(4.9)	본태성 고혈압(I10)	2659(16.9)
3	자궁경부의 염증성 질환(N72)	2628(3.6)	합병증이 없는 인슐린 비의존성 당뇨병(E11.9)	1356(8.6)
4	생세불명의 노인성 백내장(H25.9)	1535(2.1)	요도염 및 요도증후군(N34)	546(3.5)
5	상지부위의 다발성 신체손상(T00.2)	1085(1.5)	다발성 관절증(M15)	650(4.1)
6	상세불명의 열(R50.9)	1033(1.4)	치아맹출의 장애(K00.6)	446(2.9)
7	산전관리(Z34.9)	1025(1.4)	상아질의 우식증(K02.1)	443(2.8)
8	호흡곤란(R06.0)	976(1.3)	만성치주염(K05.3)	414(2.6)
9	본태성 고혈압(I10)	886(1.2)	(K02.1)	329(2.1)
10	두통(R51)	857(1.2)	자궁경부의 염증질환(N76)	326(2.1)

총 빈도(%) = 2,003,564(100)

<표 3> 보건소이용자의 거주지역에 따른 보험수혜형태

N(%)

거주지 구분	보험수혜 종류	의료급여			건강보험 직장보험	비급여	계	x ² -value (DF)
		장애인, 행려	의보1종	의보2종				
Y지역		539(4.8)	273(8.0)	7(19.4)	37(4.2)	16(3.5)	872(5.4)	27(401.1) (p= .000)
Y지역 10Km 이내 지역		193(1.7)	54(1.6)	0	3(0.3)	0	250(1.6)	
Y지역 10Km 이의 서울지역		8,536(75.8)	2,047(59.8)	29(80.6)	759(85.3)	402(87.8)	11,773(73.2)	
서울외 지역		1,298(11.5)	664(19.4)	0	43(4.8)	23(5.0)	2,028(12.6)	
기타		698(6.2)	387(11.2)	0	48(5.4)	17(3.7)	1,150(7.2)	
계(100.0)		11,264	3425	36	890	458	16,073*	

*거주지 미기재 10,934명(40.1%) 제외

5년간 총 의료이용건수는 2003,564이며, 병원과 보건소를 이용한 대상자는 각각 76,807명(74%), 27,007명(26%)으로 총 103,814명(100%)에 해당하였다. <표 3>에서 보건소 인근지역이나 지방소재자보다 서울시에 거주하는 무료수혜서비스 해당자의 보건소 이용률이 높았다. 또한 기타는 의료보험종류가 기재되지 않은 경우로 전체의 7.2%에 해당하였다. <표 4>에서 병원이용자의 의료보험종류는 보건소에 비해 다양하였고, 특히 재외국인, 자동차 보험, 군지역보험에 해당하는 경우 보건소보다 병원을 이용하고 있었고, 지역보험이나 직장보험 해당자의 병원이용률이 높았다.

3. 병원과 보건소의 질환별 총 이용비용 추이

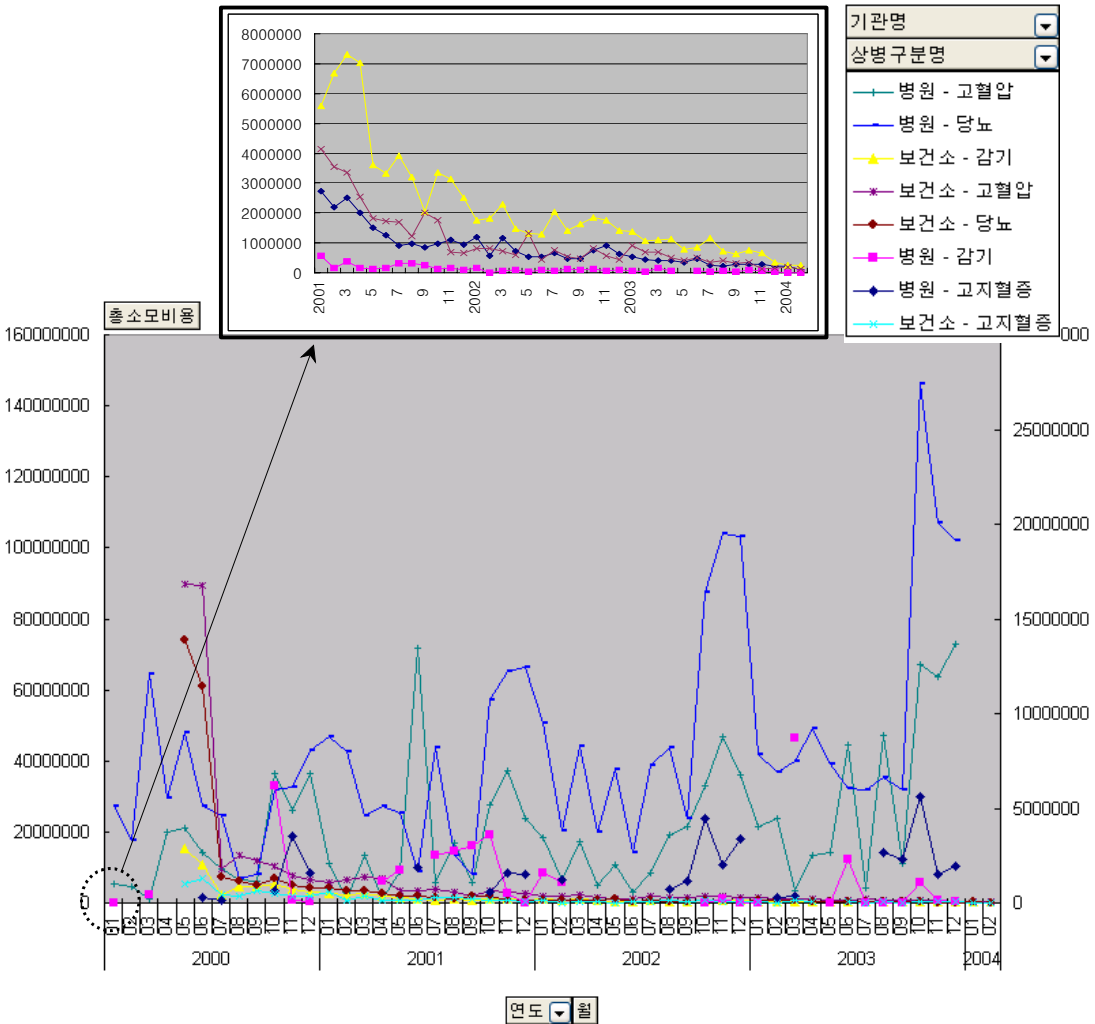
병원과 보건소 사업결과 중 환자별 총 이용비용에 대하여 기간과 질병을 차원으로 하여 OLAP 큐브를 구성하고 지표추이를 비교하였으며 그래프로 제시하였다<그림 3>.

병원과 보건소이용자의 질환별 총 이용비용의 추이를 비교한 결과 병원의 고혈압, 당뇨, 고지혈증 환자의 의료서비스 이용패턴은 2000년 8월 이후 주로 가을과 겨울에 집중되어 있는 경향을 보였으며 2003년 10월과 12월 사이의 고혈압 이용비용이 급증하였다. 이에 반해 보건소의 고혈압, 당뇨, 고지혈증 관리의 이용비용은 2000년 7월 이후 지속적으로 감소하는 경향을 보였다.

<표 4> 병원이용자의 거주지역에 따른 보험수혜형태

N(%)

거주지 종류	보험수혜 종류	비 급여	재외국 인	자동차 보험	지역 보험	군지역 보험	직장, 공무원 사립학교	의료급여		기타	계	x ² -value (DF)
								보호1종	보호2종			
Y지역		26 (7.3)	7 (11.5)	34 (6.8)	285 (4.4)	2 (4.0)	758 (7.5)	3 (0.5)	2 (1.8)	3,413 (5.8)	4,530 (5.9)	78(656.5) (p= .000)
Y지역 10km 이내 지역		15 (4.2)	4 (6.6)	39 (7.8)	548 (8.5)	1 (2.0)	864 (8.5)	26 (4.4)	4 (3.6)	4,777 (8.2)	6,278 (8.2)	
Y지역 10km 이외 서울지역		158 (44.6)	19 (31.1)	249 (49.9)	3,305 (51.1)	5 (10.0)	4,914 (48.4)	195 (33.3)	49 (43.8)	27,951 (47.7)	36,845 (48.0)	
서울외 지역		156 (43.9)	31 (50.8)	177 (35.5)	2,323 (36.0)	42 (84.0)	3,609 (35.6)	362 (61.8)	57 (50.8)	22,397 (38.3)	29,154 (38.0)	
계		355	61	499	6,461	50	10,145	586	112	58,538	76,807	



<그림 3> Y지역의 시간별 질환별 총이용비용 추이

4. 병원, 보건소 이용자의 군집특성

총 이용비용, 물리적 접근성(대한간호협회 보건간호사회, 2001)과 성, 연령(이주열 등, 2000)에 따라 병원, 보건소 이용집단, 지역주민 질환 특성이 달라지는가 보기 위해 군집분석을 적용하였다. 그 결과 보건사업 이용의 총 이용비용(1.0), 연령(0.5), 이용횟수(0.3)의 순에 따라 동질성 있는 3개 집단이 분류되었다. Y지역의 보건의료서비스 이용 특성은 물리적 접근성과 성 특성보다는 총비용, 연령, 이용횟수에 따라 질환특성이 다르게 나타났다. 이 군집특성별로 의미 있는 질환을 높은 순으로부터 정리하여 <표 5>와 <표 6>에 제시하였다. 보건소

이용자의 개인기록 중 거주지나 보험 정보 누락 등 분석에 충분하지 않은 자료 17,363명(64.3%)을 분류대상에서 제외하고 9,644건만 사용하였다.

병원 이용자의 군집 I은 가장 많은 수를 포함하고 있었으며 평균연령이 38.1세 이었으며 병원이용 횟수와 총비용이 다른 군집에 비해 중간순위에 속하였다. 이들이 병원을 이용한 이유는 주로 만성질환, 혈액계 및 기타 암질환 때문이었다. 특히 병원에 방문하는 고혈압, 당뇨병에 해당하는 환자들은 합병증을 동반한 경우에 해당하였다. 군집 II는 병원이용횟수와 총 이용비용이 세군 중 가장 작은 집단으로 평균연령이 35세였다. 이들은 주로 증상을 관리하기 위하여 병원을 찾은 경우에 속

<표 5> 병원 이용자의 군집 특성

특 성	군집 I	군집 II	군집 III
	(75,032명)	(1133명)	(642명)
총 이용비용*(원)(M±SD)	3,750,092.8 ± 7155905.5	110,218,4 ± 9744056.0	80,560,447.3 ± 18244851.9
이용횟수**(M±SD)	23.3 ± 30.5	4.5 ± 5.5	195.5 ± 3,131.2
연령(세)***	38.1 ± 24.4	66.5 ± 9.4	35.3 ± 20.3
질환(%)			
암질환	위장계암질환 5.8 혈액계암질환 2.8 유방암 1.2 자궁 및 난소 0.7 기타 암 1.2	-	-
고혈압	고혈압 1.6 고혈압성신장질환 2.2	-	-
당뇨병	-	합병증(-)1.0 합병증(+1.2)	당뇨병성 망막병증 5.9
기타	고지혈증 0.2 기타 86.4	호흡곤란 7.1 복통 3.6 두통 3.6 저체온증 3.6 기타 79.9	급성단구성백혈병 11.8 만성단구성 11.8 급성림프아구성 백혈병 23.5 빈혈 11.7 기타 35.3
계	100.0	100.0	100.0

M : Mean, SD : Standard Deviation

군집 분류기여도 : *- 1.0, **0.34, ***0.54

하였다.

군집 III은 병원 이용횟수와 이용비용이 가장 높은 집단으로 주로 백혈병 환자에게 속하였고 당뇨환자 중 망막병증에 해당하는 경우가 5.9%에 해당하였다. 이에 반해 보건소의 경우는 이 집단분류 특성으로 총 이용비용만이 유의하였으며($p < 0.01$), 총 이용비용에 의해 세 집단으로 분류하였다<표 6>. 보건소의 군집 I은 총 이용비용에 비해 이용횟수가 가장 많았고 평균 연령은 69.8세이며 관절증을 가장 많이 포함하였다. 이에 비해 평균연령이 71.0세인 군집 III은 군집 I에 비해 평균연령은 크게 차이가 없으나 이용횟수는 작고 총 이용비용은 높았다. 이들의 고혈압 관리비용이 가장 높았다. 군집 II는 보건소 이용횟수가 가장 작고 만성질환 보유율이 낮은 평균연령 48세 집단으로 이용비용이 가장 낮았다.

5. 만성질환의 질환연관성 분석

보건소의 만성질환자의 질병 연관성 결과 중 의미 있는 질병을 선택하여 신뢰도, 향상도, 지지도가 높은 순으로 11개를 선정하여 정돈하였다<표 7>. 주 상병은 첫번째 보건소 방문시 받은 주 진단이며 이를 중심으로 재방문시의 진단을 분석하였다.

지혈증을 지닌 사람이 다시 고혈압으로 방문할 확률은 90.0%이며 고혈압을 갖지 않을 확률보다 2배 높고, 전체 만성질환자 중 두 질병을 함께 가지고 있는 확률은 10.2%로 나타났다. 또한 두통으로 방문한 사람이 고혈압을 가질 확률은 90.0%이며, 고혈압이 아닌 다른 질환일 경우보다 2배 높은 확률을 가지고 있고 이 두 질환을 모두 가지고 있을 확률은 만성질환자의 9.6%에 해당하였다. 합병증이 없는 인슐린비의존성 당뇨병 환자가 감기로 재방문할 확률은 86.7%로 나타났다. 고혈압과 다른 다발성 관절증, 인슐린 비의존성당뇨병으로 재방

<표 6> 보건소 이용자의 군집 특성

특성	군집 I	군집 II	군집 III
	(9,274명)	(13명)	(357명)
총 이용비용*(원)(M±SD)	932,222.0 ± 132468.9	34,537.5 ± 53452.3	402,455.2 ± 113313.0
이용횟수(M±SD)	202 ± 104.1	10.7 ± 15.8	123.7 ± 43.3
연령(세)	69.8 ± 8.0	48.0 ± 22.1	71.0 ± 6.5
질환(%)			
위장계질환	-	-	5.8
고혈압	20.0	16.7	49.2
당뇨병	20.0	0.01	26.1
관절염	40.0	0.24	5.8
감기	20.0	-	11.5
기타 질환	0	83.1	1.6
계	100.0	100.0	100.0

M : Mean, SD : Standard Deviation

군집 분류기여도 : * - 1.0

총 군집분류대상 : 9,644(결측치 17,363 제외)

<표 7> 보건소 방문자의 연관질환

주상병	연관질환	Lift	Support (%)	Confidence (%)	Transaction Count(N)
지단백질 대사 장애 및 기타 지혈증(E78.0)	고혈압(I10)	2.0	10.2	90.0	18.0
두통(R51)	고혈압(I10)	2.0	9.6	90.0	17.0
합병증이 없는 인슐린 비의존성 당뇨병(E11.9)	감기(J00)	1.6	29.6	86.7	3.0
고혈압(I10)	다발성 관절증(M15)	4.4	9.5	86.4	38.0
고혈압(I10)	비인슐린의존성당뇨병(E11.9)	2.0	26.3	78.3	2.0
합병증이 없는 인슐린 비의존성 당뇨병(E11.9)	다발성 관절증(M15)	2.0	21.2	63.3	38.0
만성치주염 (K05.3)	위장계질환(K31)	4.5	5.5	62.9	22.0
만성치주염(K05.3)	상세불명부위의 소화성궤양(K27)	1.9	7.1	61.2	30.0
만성치주염(K05.3)	합병증이 없는 인슐린 비의존성 당뇨병(E11.9)	1.8	15.8	57.1	28.0
기타 및 상세불명의 복통(R10.4)	자궁경부의 염증성 질환(N72)	4.1	3.65	53.7	8.0
만성치주염(K5.3)	위염 및 십이지장염(K29)	2.1	14.5	30.3	23.0

문할 확률은 86.4%, 78.3%이며 다른 이유보다 각각 4.4배, 2.0배 높았다. 만성치주염을 이유로 보건소를 방문하는 사람이 다시 위장계 질환, 소화성 궤양, 위염 및 십이지장 궤양으로 방문할 확률은 62.9%, 61.2%, 30.3%로 나타났으며 다른 질환 동반율보다 4.5, 1.9, 2.1배 높고, 이 질환을 함께 가지고 있을 확률은 만성질환자의 7.1%, 15.8%, 14.5%에 해당하였다. 또한 보통으로 진단받은 사람이 자궁경부의 염증성 질환으로 재방문할 확률은 4.1배 높고 다른 질환일 확률보다 53.7% 높은 것으로 나타났다<표 7>.

IV. 논 의

그 동안 건강증진사업과 지방자치체 도입 이후 지역사회 중심의 건강관리사업에 대한 국가적인 노력이 얼마나 성과가 있었는지에 대한 다각도의 평가가 있었다(강성홍과 최순호, 2001; 서미경 등, 2003). 그러나 이는 공공사업분야에 국한되거나 민간병원내부의 경영이나 마케팅 전략수립 차원의 일회적인 평가수준이어서 지역단위의 통합적이고 지속성 있는 평가가 되지 못하였다. 사실상 국가적인 차원의 보건의료정책의 기본 방향은 형평성과 효율성의 기본원칙을 동시에 보장하기 위한 의료분배방법을 설계하고 구현해야 한다. 이는 세계적인 관심으로 의료분야에 균형성과지표, DW, 데이터마이닝 등과 같은 지식경영기법을 도입하기에 이르렀다. 더욱이 차세대에는 각 보건의료 기관이 수행하는 사업결과가 수록된 EHR과 CDR을 통합하여 체계적인 보건의료 지식자원 인프라를 구축하고 보건의료정책의 정확성과 적시성을 도모하고자 하는 계획이 활발히 진행되고 있다.

이 연구는 2000년 2월에서 2004년도 12월까지 서울시 일개 지역내의 병원과 보건소의 보건의료사업 이용결과를 분석하여 데이터베이스를 이용한 보건사업평가의 유의성을 점검하고, 질환과 비용을 이용한 이용 특성 분류, 보건소 이용자의 질병 연관성에 관한 정보를 산출하였다.

이 연구의 결과분석에서 이용한 항목은 OECD, ISO의 보건사업평가를 위한 표준 성과지표 틀에서 선정하였다. 그 결과 인구보건사정과 사업성과 모니터의 영역의 대분류 항목, 중분류항목 중 건강상태의 건강조건, 건강결정인자의 개인자원, 지역사회 보건사업 특성의 지역사회와 보건기관, 보건사업 서비스 특성 중 건강증진 사업이 선정되었다. 보건사업 성과모니터는 접근성과 효과성이 선정되었다. 각 분류항목에 해당하는 상세지표로 건강조건에는 질환별 질병명, 개인자원에서는 거주지, 의료보험종류, 연령, 성별을 이용하였으며 의료이용에는 비

용, 양, 간격, 빈도, 보건서비스 특성에는 Y지역의 병원, 보건소가 선정되었다. 보건사업 서비스 특성에는 건강증진 사업, 접근성에는 지역/시간별 서비스 이용 추이, 효과성에는 다빈도 질환, 만성질환 이용 지표가 선정되었다.

이 연구의 대상인 병원 자료는 Y동 지역의 의료보험청구 비용의 91.7%를 차지하기 때문에 해당지역의 관할지역인 보건소의 자료와 함께 이지역의 민간, 공공보건의료서비스 결과를 분석하기에 충분한 자료로 판단하였다. 병원인 경우 의료보험료 부과를 위하여 구축된 데이터베이스를 사용한 결과 누락된 정보는 비교적 충실하였다. 반면 보건소의 경우 64%가 누락되는 결과를 보여주었다. 또한 병원인 경우 ICD 10코드로 된 질병명의 소수점이 누락되어 DB통합과 상세분석이 어려웠다. 이러한 이유로 병원진단자료를 이용한 연관성 분석은 연구에서 제외하였다.

Miller(1994)는 민간차원에서 지역단위의 통합적인 보건사업 성과를 평가하기 위하여 지역사회 관할 지역 내의 민간의료기관을 포함한 다수 공공보건기관의 사업결과 DB 중 효과성에 관한 항목을 수합하여 단면적인 비교를 한 바 있다. Glouberman과 Miller(2003)는 건강증진 사업을 포함한 보건정책 분석을 위해 음주, 흡연, 영양, 건강자거보고 지표를 사용하였다. 이 연구에서는 질환명과 총 의료비용만을 사용할 수 있었다. 이는 아직 보건의료기관 자체 내에서 개발한 정보시스템 생산 자료가 지역보건의료양상, 사업증진 및 연구, 대민 서비스에 대한 정보를 산출하지 못하고 있다는 것을 증명하고 있다.

무엇보다 DB/DW의 활용에 있어서 가장 큰 과제는 자료의 빈곤이다. 각 보건기관 DB의 결과자료가 보건의료정책분석 목적에 부합하려면 체계적인 지표를 기반으로 자료가 균형적으로 수집되어야 한다. CDC와 IOM기구에서 사업성과향상을 위해 개발한 KPI 항목의 지표 역시 각 지역기관의 자료부족으로 적용하지 못하고 있는 실정이다(Parker, Margolis, Eng & Carlos, 2003). 최근 CDC에서는 국가 보건서비스 수행기준에 따라 '필수 공중보건서비스를 10가지로 제시하여 공공보건사업에서의 기획과 조정, 지역사회와의 연대, 평가 역할을 강조하고 있다. 또한 각 주에서 이를 실행하도록 공중보건 필수사업에 대한 인증제도를 도입하고, 공공보건 필수서비스에 대한 수행기준과 도구 개발, 도구에 대한 타당성 평가 등을 진행하고 있다. 이 10가지 필수 공공보건서비스에는 모니터링, 진단과 조사, 정책과 사업계획의 개발, 효과, 접근성, 질 평가, 연구가 중요한 순위 안에 포함되어 있다.

미국의 IOM(1991)에서는 보건의료서비스 평가지표를 개발하였는데 초기에 총 10개 영역에 대하여 세부지표 84개를 사

용하였으나 장기간 지표 활용과정을 통해 핵심지표로 함축되어 총 6개 영역에 대하여 23개를 사용하고 있다(Halverson & May, 2001). 우리나라도 향후 전문가와 관련기관의 합의에 의한 핵심평가지표의 선정, 조직의 지표 활용과 유용한 지표의 여과과정이 필요하다고 여겨진다.

Y지역에 소개한 병원과 보건소 이용자들은 지역외부에 거주하는 경우가 많았다. 병원은 혈액암에 관한 고난이도 서비스에 관한 인지도가 높은 이유이며, 보건소의 경우 교통편의가 높아 먼 거리의 의료급여대상자도 많이 방문하는 것으로 볼 수 있다. OLAP을 이용한 시간별 질병추이를 살펴본 결과 고혈압, 당뇨, 고지혈증 환자의 보건소와 병원의 이용패턴에 차이가 있었다. 병원을 이용하는 환자는 계절적 성향이 높은 것으로 나타났다. 보건소에서는 건강증진 사업이 본격화된 2000년 이후 만성질환 이용수가 줄었다. 이는 보건소를 근간으로 한 공공보건사업이 실적위주의 양적인 사업에서 질적으로 변화되는 결과로 해석된다. 또한 보건사회 연구원(2003)의 보고에 의하면 보건소의 기능과 역할범위가 확대되었으나 2000년 보건소의 전문인력은 1996년보다도 감축되거나 정체 상태에 있다고 하였다. 보건소 인력구조를 1996년과 비교하면 의사, 간호사 전문인력이 IMF 이후 구조 조정으로 줄어들었으며, 새로운 사업에 필요한 전문인력의 확보는 거의 이루어지지 않았으며 보건소 이용률은 도시지역에서는 특히 낮으며, 농촌 지역, 저소득층, 의료급여대상자에서 상대적으로 다소 높다는 내용과 일치한다.

이병화와 채영문(2004)은 K지역의 보건소 자료를 분석한 결과에서도 같은 양상을 보였으나 국민건강변화에 대한 전국적 조사(국민건강영양조사)에서는 이러한 패턴을 도출하지 못하였다. Frerichs and Shaheen(2001)은 자료수집의 정확성을 위한 방법론을 개발하기 위해 군집 분할의 타당성 확보를 위한 소구역 근거조사를 제시한 바 있다. 또한 군집분석 방법론을 개발하기 위한 연구들을(Campbell, Mollison, Steen, Grimshaw & Eccless, 2001; Ukoumunne & Thompson, 2001; Loeys, Vansteelandt & Goetghebeur, 2001; Chiara, Patrizia, Piero, Carlo & Perucci, 2002) 바탕으로 이 연구에서 모집단의 등분산성 가정을 필요로 하지 않는 비모수적인 방법인 데이터 마이닝 방법의 이점을 활용하여 DW에 의존한 자료 분석을 할 수 있었다(Flaherty, 2001). 이 방법은 사전 정보 없이 정보 분석이 가능한 지식분석방법이다(Breiman, 1984).

이 연구와 같이 의료분야의 정보 분석에 군집분석을 활용한 예를 들면 Harris 등(2002)은 단백질표현지도의 데이터베이스를 구성하여 광범위 유방암의 조직행렬에 군집분석을 적용하

였으며, 이를 통해 유전자 교체와 DNA, RNA의 유전자 표현 방식을 정의함으로써 계층기술을 개발하였다. 그밖에 MRI 결과분석을 통한 암 세포변화 추적연구로 간질 모니터 효과를 높이기 위한 발작증상그룹과 관련된 위험요인 규명(Haut, 2003), 약물남용과 청소년 능력과 관련된 애착-자율성 형성의 개인적 차이(Lee & Bell, 2003), 약물의존 남자의 기질과 중증도 변수에 따른 분류(Henderson & Galen, 2003), 환자와 학생의 개인적 신상의 군집분석을 이용한 우울증 지속성(Cox, Enns & Larsen, 2001), 군집분석을 통한 식사패턴 분석의 내적 타당성(Quatromoni et al., 2002), 직업적 조건의 성 차이와 여성건강(Bildt & Michelsen, 2003) 등의 연구가 있었다. Hillestad and Berkowitz(1984)와 James(1996)는 병원 마케팅 계획과 자원의 효율적 분배에 활용할 수 있는 방안을 도출하였다.

의료비용과 빈도, 양을 의료서비스의 결과로 분석하는 과정에서 의료서비스 구매 간격의 패턴이 일정하지 않아 제외한 후 의료비용의 크기만을 가지고 세 개의 군집을 나누고 군집별 질병경향을 분석하였다. 이 결과 의료서비스 기관을 결정하는데 물리적 접근성보다 비용으로 인한 접근성이 높았다. 비용 크기를 결정하는 요인은 병원과 보건소의 차이가 있었는데 병원에서는 총 이용비용과 이용횟수와 연령이, 보건소는 총 이용비용만이 이용군을 분류하는 요인이었다. Y지역에 소개한 병원은 전국을 대상으로 한 혈액암 질환 전문병원으로의 특성을 보여주었다. 즉 고난이도의 시술이 필요한 평균 35세에 해당하는 연령대의 백혈병환자와 당뇨병성 망막증의 치료비용이 가장 높았다. 두 번째로 평균 연령이 38세인 위장, 유방, 혈액계 등의 암질환, 고혈압, 당뇨, 고지혈증환자군의 비용이 높았으며 가장 많은 수를 포함하였다. 가장 진료비용이 적은 군집은 평균 66세에 해당하는 고연령군으로 각종 증상으로 인한 진단검사의 지출비용에 해당하였다. 이는 만성병의 조기발견과 합병증 관리를 위해서는 병원을 이용하고 공공보건의료기관에서는 저소득층의 일반적인 질환관리서비스를 이용하는 결과로 해석된다. 이 결과의 시사점은 저소득층이 이용하는 보건소의 만성질환자의 수는 줄었지만 아직 질환관리 중심의 서비스로써 진단, 검사 및 합병증 관리서비스가 부족하다는 점이다. 또한 병원과 보건소의 서비스별 목표에 대한 결과치가 포함된 상세정보가 부족하다는 점이다. 예를 들면 보건소의 만성질환관리 이용건수가 줄었다면 질적 서비스 결과가 평가되어야 하고, 보건소의 건강검진 서비스의 직접적인 결과로 암과 성인병 조기발견 수와 같은 직접적인 자료를 필요로 한다.

이 연구에서 사용한 연관분석 결과는 고연령으로 인한 다빈

도 질환의 발생확률, 약물부작용 정보, 면역과 관련된 동반질환 정보를 산출하였다. 즉 고혈압을 지닌 환자가 관절증이나 당뇨로 재방문할 가능성 등 고연령과 소득에 관련된 질환발생 정도를 보여주었다. 또한 이미 유의하다고 알려진 질환 간 관련성만 산출할 수 있어서 매우 제한적인 분석만 가능하였다. 향후 증상과 질병, 생활습관자료의 양이 많아지면 각 변수간의 유의한 연관성과 발생확률을 밝혀 예방상담 활동에 활용할 수 있을 것이다.

이 연구의 중요한 의의는 기존의 기관별 정보시스템에 누적되었던 자료를 통합하여 세계적인 표준 보건의료서비스 평가 지표체계를 틀로 하여 지역사회 중심 보건의료사업의 결과를 분석하였다는데 의의가 있다. OLAP분석 방법은 질병의 계절적인 경향, 건강박람회 등의 건강홍보행사와 관련한 건강변화를 민감하게 보여줄 수 있으며, 데이터 마이닝 분석은 보건소 서비스 이용의 영향력과 만성질환과 같은 국가 중점사업 분석에 유용하게 활용할 수 있음을 보여주었다.

마지막으로 이 연구의 제한점은 다음과 같다.

- 첫째, 일개 지역의 병원과 보건소 사업의 결과가 기록된 데이터베이스를 분석대상으로 한 후향적 조사 방법이다.
- 둘째, Y지역의 의료이용결과 중 데이터 충실도가 떨어지는 항목은 제외하였기 때문에 민간, 공공의료사업을 충분히 비교하지 못하였다.
- 셋째, 보건소를 이용하는 Y지역 거주자 중 저소득층이 상대적으로 적다는 점을 감안하기 위해 지역별 총생산 지표를 비교하여야 한다.

이 연구를 통해 제안할 점은 보건의료사업의 평가는 기존의 자료를 충분히 활용한 지속적인 노력이 중요하기 때문에 국제적 표준 평가지표를 이용한 지역단위 보건의료기관의 통합 DW의 구축과 충실한 자료 확보 노력이 시급히 필요하다고 할 것이다.

V. 결론

본 연구는 지식분석기법을 적용하여 효과적인 의료자원 배분을 위한 기초자료를 생성하기 위하여 Y지역의 5년간의 보건 의료서비스 이용결과를 평가하였다. 2000년 2월에서 2004년도 12월까지 Y지역 내의 병원과 보건소의 데이터베이스를 이용한 보건사업평가의 유의성을 점검하고, 주 질환과 총 이용비용을 이용한 이용 특성 분류, 보건소 이용자의 질병연관성에 관한 정보를 산출하였다. 그 결과 다음과 같은 연구결과를 얻었다.

- 1) Y지역의 대학병원과 보건소의 최근 5년간 상위 10순위 질환의 전체 빈도수에 차지하는 비율은 병원은 22.4%인 반면 보건소의 경우 59.0%를 차지하였다.
- 2) 대학병원과 보건소 이용자의 거주지역에 따른 보려수혜 형태에 유의한 차이가 있었다($p = .000$). 분석대상인 Y지역의 5년간 총 의료 이용건수 2003,564건 중 병원과 보건소를 이용한 대상자는 각각 76,807명(74%), 27,007명(26%), 총 103,814명(100%)에 해당하였다. 보건소는 인근지역이나 지방소재자보다 서울시내에 거주하는 무료수혜서비스 해당자의 보건소 이용률이 높았다. 병원이용자의 의료보험종류는 보건소보다 더 다양하였다.
- 3) 병원과 보건소이용자의 질병별 총 이용비용의 추이를 비교한 결과 병원의 고혈압, 당뇨, 고지혈증 환자의 의료서비스 이용패턴은 2000년 8월 이후 주로 가을과 겨울에 집중되어 있는 경향을 보였으며 2003년 10월과 12월 사이의 고혈압 이용비용이 급증하였다. 이에 반해 보건소의 고혈압, 당뇨, 고지혈증관리의 이용비용은 2000년 7월 이후 지속적으로 감소하는 경향을 보였다.
- 4) 총 이용비용, 물리적 접근성과 성, 연령에 따라 병원, 보건소 이용집단, 지역주민 질환 특성이 달라지는가 보기 위해 군집분석을 적용한 결과 보건사업 이용의 총 이용비용(1.0), 연령(0.54), 이용횟수(0.34)의 순에 따라 동질성 있는 3개 집단으로 분류되었다. Y지역의 대학병원 이용 특성은 물리적 접근성과 성별 특성보다는 총 이용비용, 연령, 이용횟수에 따라 질환특성이 다르게 나타났다.
- 5) 보건소의 만성질환자의 질병 연관성분석 결과로 고연령으로 인한 다빈도 질환의 발생확률, 약물부작용 정보, 면역과 관련된 동반질환에 관한 상세정보를 산출하였다.

이 연구를 통해 제안할 점은 보건의료사업의 평가는 기존의 자료를 충분히 활용한 지속적인 노력이 중요하기 때문에 국제적 표준 평가지표를 이용한 지역사회 보건의료 정책의 수립과 지속적인 평가가 이루어져야 한다는 점이다. 이를 위해서는 보건지표체계에 의한 지역단위 보건의료기관의 통합DW의 구축과 자료충실성을 높이는 정책이 필요하다고 본다.

참고문헌

- 강성홍, 최순호(2001). 데이터 마이닝을 이용한 보건소의 건강증진 사업의 효율화 방안. *의료정보학회지*, 7(2), 37-48.

- 건설교통부(2005). *건설교통통계연보 2004*. In : <http://www.moct.go.kr/DataCenter/StatisticData/>
- 국민보험공단(2005). *통계자료실*. In : <http://www.nhic.or.kr/wbm/wbmb/>
- 대한간호협회 보건간호사회(2001). *지역보건간호사업평가과정*.
- 미국 질병관리국 홈페이지. <http://www.cdc.gov>.
- 보건복지부(2003). *진료정보 공동활용을 위한 기반조성 연구*.
- 서미경, 김혜련, 한영자, 선우덕, 송현중, 김남순, 황나미, 김동진, 박인철, 오희철(2003). *보건의료서비스 공급체계 개선 방안 연구보고서*. 한국보건사회연구원.
- 신영석, 신형웅, 신종각(1999). *의료보험 진료비 증가요인과 정책과제*. 한국보건 사회연구원 연구보고서.
- 이병화(2004). *보건소 건강사업을 위한 의사결정시스템 모형구축과 활용방안*. 박사학위논문.
- 이병화, 채영문(2004). *보건평가지표체계구축. 보건정보교육학회지*, 3(2), 8-14.
- 이주열, 이순영(2001). *평생국민건강관리체계 정책방안에 관한 연구*. 보건사회연구원.
- 이주열, 이순영, 박재용, 조홍준, 송태민(2000). *평생국민건강관리 체계 구축방안에 관한 연구*. 한국 보건사회연구원.
- 최병호, 신형호(2000). *국민건강보험의 재정위기 평가와 재정안정화 방안*. 보건사회연구원 용역연구.
- Bildt C., Michelsen H.(2003) Occupational conditions exceed the importance of non-occupational conditions and ill health in explaining future unemployment among women and men. *Arch Womens Ment Health* 8, 21-23.
- Breiman L. J. H., Friedman, R.A.(1984). Classification and Regression tree. Belmont : Wadsworth, 1984. In Cabena P(1997). *Discovering datamining : Firm concept to implementation*, Perntice-Hall.
- Campbell M.K., Mollison, J., Steen, N., Grimshaw, J. M., & Eccless, M.(2000). Analysis of cluster randomized trials in primary Care. *A Practical approach Family Practice*, 8(4), 2-7.
- Chiara, M., Patrizia, S., Piero, B., Carlo, A., & Perucci, B(2001). Application of random effect ordinal regression model for outcome evaluation of two randomized controlled trials. *Statistics in Medicine*, 2, 3-5.
- Cox, B. J., Enns, M. W., & Larsen, D. K.(2001). The continuity of depression symptoms : Use of cluster analysis for profile identification in patient and student samples. *Journal of Affective Disorders*, 65, 67-73.
- Deborah, C., Bradford, G. E., & Northridge, M. E.(2003), Scar E. A joint urban planning and public health framework: Contributions to health impact assessment. *American Journal of Public Health*, 93(1), 118-121.
- Duffy, J.(2001). The tools and technologies needed for knowledge management. *Journal of Information Management*, 1(35), 34-41.
- Erickson, E. H.(1975). *Life history and the historical moment*. New york, Norton.
- Flaherty, W.(2001). Data Warehousing: How to make your statistic meaningful. *School Planning & Management United States*, 2(40), 73-53.
- Frerichs, R. R., & Shaheen, M. A.(2001). Small-community based surveys[Annual Rev]. *Public Health*, 22, 231-47.
- Gardner Inc.(2003). *Business Intelligence and Data Warehousing*. In : http://www.gardner.com/Research/focus_area/asset-48250.jsp.
- Green, L. W., & Kreuter, M. W.(2003). *Health promotion planing-an educational and environmental approach*. May field Publishing company.
- Glouberman, S., Millar, J.(2003). Evolution of the determinants of health, health policy, and health information system in canada. *American Journal of Public Health March*, 93(3), 388-392.
- Halverson, P. K., & May, G. P.(2001). Public health assessment. In : Novick LF, Mays GP, Editors. *Public Health Administration*. Jones and Bartlett Pub, 267-293.
- Ham, C., & Maynard, A.(1994). Managing the NHS markets. *British Medical Journal*, 308, 45-847.
- Harries, R. A., Yang, A., Stein, R. C., & Lucy, K.(2002). Brustein Luc, Herath A, Parekh R. Cluster analysis of extensive human breast cancer cell line protein expression map database. *Proteomics*, 2, 212-223.
- Haut, S. R.(2003). Seizure clustering during epilepsy monitoring. *Epilepsia*, 43(7), 711-715.
- Henderson, M. J., & Galen, L. W.(2003). A Classification of substance-dependeten on temperament and severity variables. *Addictive Behaviors*, 28, 741-760.
- Hillestad, S. G., & Berkowitz, E. N.(1984). *Health care mar-*

- keting plans From Strategy to Action*. Illinois.
- IOM(1991). *A role for performance monitoring*. Washington DC: Coutesy of National Academy Press.
- James, F. M.(1996). *The marketing process and health promotion programs*. Databases Put the Direct in Direct Marketing.
- Lee, J. M., & Bell, N. J.(2003). Individual differences in attachment-autonomy configurations : Linkages with substance use and youth competencies. *Journal of Adolescence*, 26, 347-361.
- Loeys, T., Vansteelandt, S., & Goetghebeur E.(2001). Accounting for correlation and compliance in cluster randomized trials. *Statist, Med.*, 20, 3753-3767.
- Mark, S., Nicole, Q., & Matthew, W.(2005). Profit-Seeking, Corporate Control, and the Trustworthiness of Health Care Organizations: Assessments of Health Plan Performance by Their Affiliated Physicians. *HSR: Health Services Research*, 40(3), 605-646.
- Miller, C.A.(1994). A proposed method for assessing public health function and practice. *American Journal of Public Health*, 84(1), 743-749.
- OECD(2003). *한국경제 보고서*. 재정경제부.
- Quatromoni, P. A., Copenhafer, D. L., Demissie, S., D'Agostino, R. B., O'Haro, C. E., Nam, B. H., & Millen, B. E.(2002). The internal validity of a dietary pattern analysis: The framingham nutrition studies. *Journal Epidemic Community Health*, 56, 381-388.
- Sindell, M. T.(2001). KM Conversation. Training D, 20-22.
- Ukoumunne, O. C., & Thompson, S. G.(2001). Analysis of cluster randomized trial with repeated cross-sectional binary measurements. *Statistics in Medicine*, 20, 417-433.
- Van de Ven, Fschut, W.(1995). The dutch experience with international markets: Health care reform through international market. *Washington Brookings Institution*, 95-118.
- Wears, R. L.(2002). Advanced Statistics : Statistical methods for analyzing cluster and cluster-randomized data. *Academy Emergency Medicine*, 9(4), 330-341.

-Abstract-

Key words : Health care service utilization, Data mining, Data warehousing

Evaluation on Utilization of the Health Care Service in One Urban Area in Korea

Ahn, Sung Hee · Lee, Byung Wha***

Purpose: This study was to evaluate the utilization of health care service and to provide supportive data for health care policy making in one urban area in Korea. **Method:** This study tested the significance of public health service using the database of an university hospital and public health center from Feb. 2000 to Dec. 2004. Data were analyzed by multidimensional analysis and data mining technique and produced the information on the classification of utilization characteristics by main disease and the total cost of use and disease association with the users of the public health center. **Results:** The Results were as follows: 1) Top 10 diseases in the area accounted for 22.4% of total frequency for the most recent 5 years in university hospital, while 59.0% in public health center. 2) There were significant correlations between university hospital and public health center user's insurance type and place of residence: It showed higher use of public health center for free service beneficiaries residing in Seoul than residents in nearby or local area. The medical insurance types for hospital users were more various than those for public health center users. 3) The use of hospital for patients of hypertension, diabetes mellitus and hyperlipidemia was tended to concentrate in mostly autumn and winter since August 2000, while the cost of using public health center for those patients has been steadily reduced since July 2000. 4) As a result of cluster analysis, there were classified into three homogeneous groups according to the total cost of using public health service, age, and the frequency of use. 5) The association

*Associate Professor, College of Nursing, The Catholic University of Korea

**Part time Instructor, Graduate School of Yonsei University

analysis on patients with chronic disease in public health center produced a detailed information on accompanying diseases related to the incidence rate of disease of high frequency due to aging, information on drug abuse and immune disease. Conclusion: The health care policy for local community should be evaluated continuously. And the policy to build an integrated data warehousing by public health indicator system and to enhance the faithfulness of data is required.