

# 퇴원요약 데이터베이스를 이용한 데이터마이닝 기법의 CQI 활동에의 활용 방안

선미옥\*, 채영문\*\*, 이해종\*\*\*, 이선희\*\*\*\*, 강성홍\*\*\*\*\*, 호승희\*\*

강동성심병원 의무기록과\*, 연세대학교 보건대학원\*\*, 연세대학교  
보건행정학과\*\*\*, 이화여자대학교 예방의학과\*\*\*, 인제대학교 보건관리학과\*\*\*\*

## An application of datamining approach to CQI using the discharge summary

Mi-Ohk Suhn\*, Young-Moon Chae\*\*, Hae-Jong Lee\*\*\*, Sun-Hee Lee\*\*\*\*, Sung-Hong Kang\*\*\*\*\*  
Kang\*\*\*\*, Seung-Hee Ho\*\*

Dept. of Medical Record, Kang Dong Sacred Heart Hospital\*, Graduate School of Health Science and Management, Yonsei University\*\*, Dept. of Health Administration, Yonsei University\*\*\*, Dept. of Preventive Medicine, Ewha University\*\*\*\*, Dept. of Public Health, Inje University\*\*\*\*\*

### Abstract

This study provides an application of datamining approach to CQI(Continuous Quality Improvement) using the discharge summary. First, we found a process variation in hospital infection rate by SPC (Statistical Process Control) technique. Second, importance of factors influencing hospital infection was inferred through the decision tree analysis which is a classification method in data-mining approach. The most important factor was surgery followed by comorbidity and length of operation. Comorbidity was further divided into age and principal diagnosis and the length of operation was further divided into age and chief complaint.

24 rules of hospital infection were generated by the decision tree analysis. Of these, 9 rules with predictive power greater than 50% were suggested as guidelines for hospital infection control. The optimum range of target group in hospital infection control were identified through the information gain summary.

Association rule, which is another kind of datamining method, was performed to analyze the relationship between principal diagnosis and comorbidity. The confidence score, which measures the degree of association, between urinary tract infection and causal bacillus was the highest, followed by the score between postoperative wound disruption and postoperative wound infection.

This study demonstrated how datamining approach could be used to provide information to support prospective surveillance of hospital infection. The datamining technique can also be applied to various areas for CQI using other hospital databases.

Key words :Datamining, CQI(Continuous Quality Improvement), Hospital Infection

---

주소 : (우)134-701 서울시 강동구 길동 445 강동성심병원 의무기록과  
전화번호 : (02) 2224-2152 FAX 번호 : (02) 488-0119

# I. 서 론

## 1. 연구 배경 및 필요성

의료환경의 변화는 질병구조, 의료서비스 이용형태, 의료체계, 국민의식 등의 변화를 가져왔고, 의료계 주변여건 역시 병실료 단입제, 신용카드 수납, 의약분업, 선택진료제, 포괄수가제 확대실시, 약제비 실거래가 상환제, 자동차보험수가입하, 수가차등제 등 병원에 있어서도 많은 변화를 요구하고 있다(이경미, 1999). 이로 인한 병원의 수익성 감소는 병원 생존을 위협하고 있으며, 고객들의 양질의 의료서비스에 대한 기대와 욕구는 병원으로 하여금 기존의 의료서비스와는 차이가 있는 의료서비스를 생산할 것을 요구하고 있다. 병원에서 환경변화에 적응하기 위한 여러 형태의 노력 중 하나로 Continuous Quality Improvement (CQI, 지속적 질관리) 또는 Total Quality Management (TQM, 총체적 질관리)에 관심을 두는 병원들이 늘어나게 되었다(조우현, 1999).

의료기관의 전산화도 급격히 발전하여 많은 병원에서 처방전달시스템이 구현되었으며(채영문, 1994), 더 나아가 몇몇 병원에서 의무기록전산화를 구현함에 따라 이제까지 종이에 보관되던 진료정보가 컴퓨터에 보관됨에 따라 이러한 정보를 효율적으로 활용할 필요가 있다(김옥남 등, 1999).

병원은 경쟁력을 확보하기 위해 여러 가지 방안을 마련하고 이를 실천하고 있는데, 이 중 하나가 정보기술을 활용하는 병원의 지식경영이라고 할 수 있다. 앞으로의 병원경영은 종전과는 달리 환자에게 최소의 적정서비스를 제공하는 형태로 변화해야 할 것이며, 선진국에서는 이미 이러한 시대적 요청에 부응하여 병원에서 진료정보의 효율적 활용을 위한 데이터웨어하우스, 데이터마트, 데이터마이닝 등의 정보기술을 기반으로 진료정보를 효율적으로 활용하고 있다

(Internetsite [<http://www.cio.seoul.kr/main4.html>])

진료데이터의 가장 기본이 되는 것은 의무기록과의 퇴원요약정보라 할 수 있겠다(강성홍 등, 1999). 이것은 매일 매일의 퇴원기록을 대상으로 워크시트를 이용하여 여러가지 항목으로 분석하는 것을 말하며(홍준현, 1996) 환자에 대한 기본정보를 포함하고 있다. 특히 미국에서는 국가차원에서 Uniform Hospital Discharge Data Set을 만들어 관리하면서 진료정보 데이터베이스의 가장 기본적인 항목을 설정할 때, 이 내용을 중심으로 하고 있다(이정화, 1996). 많은 병원에서 질 향상에 관련된 주요 데이터가 전산화되어 있으므로, 이러한 데이터는 병원의 업무처리를 위해 구축된 데이터이지만 질 향상활동에도 활용할 필요가 있다.

질 향상을 위해서는 객관적인 자료를 분석해서 얻은 정보가 필수적이며(조우현, 1999), 질 향상 활동에서 가장 먼저 생각해야 하는 것이 현재의 상태 또는 현상을 파악하고 이를 죽적할 수 있는 자료의 형태로 만들어서 질 향상활동의 대상영역(Target Population)을 찾아내는 것이다(안선경, 1999). 업무과정의 개선을 통하여 지속적인 질 향상을 가져오기 위해서는 밝혀진 대상에 대한 의료의 질의 변동을 밝힘으로써, 변이를 조절하고 동일 수준의 의료의 질을 유지하기 위한 모니터링이 요구되어진다. 이를 위한 방법으로써 산업체에서 공정관리하는 기법인 Statistical Process Control (SPC, 통계적 공정관리)이 보건의료분야에 90년대 후반부터 적용되고 있는 추세이다(박성현 등, 1998) 질 향상활동도 단편적인 사건들에만 집착한 표면적인 사건분석방법으로는 얻을 수 없는 합리적인 의사결정의 수단으로서의 자료수집 및 분석이 이루어져야 할 것이다(안선경, 1999). 이에 신속한 자료 분석과 신속한 결과 제시 및 예측까지도 가능한 정보기술의 활용과 질 향상활동을 연계할 필요성이 있다.

선진국에서는 의료기관에서 정보를 지식으로 전환하여 비용관리를 돋고, 환자진료의 질을 향상시키기 위해서 데이터마이닝기법을 이용하고 있고(용환승, 1998), 이 기법을 이용함으로써 미래의 환자행동을 예측하고 진료프로그램을 향상시키는데 이용될 수 있다. 또한, 고위험군 환자를 파악함으로써 임상의사는 오늘의 환자에게 더욱 양질의 환자관리를 할 수 있어 내일의 문제가 되지 않게 된다(Greg 등, 1999).

이에 본 연구에서는 퇴원환자의 퇴원요약 데이터베이스를 이용하여 질적 문제 중 병원감염 발생의 변동을 SPC 기법을 활용하여 찾아낸 후, 데이터마이닝 기법을 활용하여 병원감염발생규칙을 도출해 냈으므로써, 병원감염 관리 및 효율적인 CQI 활동에의 활용방안을 제시하는 데에 그 의의가 있다고 할 수 있다.

## 2. 연구 목적

본 연구는 효율적인 질 향상 활동을 위해 퇴원환자의 퇴원요약 데이터베이스를 토대로 SPC 기법을 활용한 병원감염발생의 변동을 파악한 후, 데이터마이닝 기법을 활용하여 병원감염 발생의 규칙을 찾아내어 CQI 활동에의 활용방안을 제시하고자 하였다. 구체적인 목적은 다음과 같다.

첫째, SPC 기법을 활용하여 병원감염발생의 현황 및 문제점을 분석한다.

둘째, 데이터마이닝 기법을 활용하여 병원감염 발생의 규칙을 추출하고, 추출된 규칙에 대한 타당성을 검증한다.

셋째, 데이터마이닝 기법을 통해 도출된 결과의 CQI 활동에의 활용방안을 제시한다.

## II. 연구대상 및 방법

### 1. 연구 대상

1996년 1월1일부터 1999년 12월 31일까지 서울시에 있는 K 대학병원에서 퇴원한 88,593명의 퇴원요약 데이터에서 질적 문제(Quality Problem)에 해당되는 환자 884명 중 병원감염발생이 343 건으로서 가장 높은 비율인 36%를 차지했다. 본 연구에서는 병원감염관리방안을 찾기 위해 병원 감염환자를 연구대상으로 선정하였다. 병원감염 발생 환자 총 수술받은 242명과 단순 임의추출로서 수술받은 병원감염 미발생환자 1,281명을 선정하여 총 1,523명을 연구대상으로 하였다.

### 2. 연구 방법

#### 가. 분석도구

Access를 이용하여 구축된 분석용 파일을 다음과 같이 분석했다.

첫째, 기술통계를 작성하여 연구대상자의 일반적 특성 파악을 위해 SAS의 통계패키지를 사용하였다.

둘째, 병원감염 발생률의 변동 및 추이를 분석하기 위해 SAS의 SQC를 사용하였다.

셋째, 병원감염 발생의 규칙 및 패턴을 찾기 위해 SAS의 Enterprise Miner를 사용하였다.

#### 나. Statistical Process Control 기법을 활용한 병원감염 발생의 변동 및 추이 분석

의료서비스에서 요구되는 의료의 질이나 비용효과적인 의료제공의 목표를 달성하기 위한 SPC 기법 중 하나인 관리도 분석을 이용하였으며, 이 중 병원감염 발생률의 변동을 파악하기 위해 종속 변수가 계수형이며, 부분군의 크기가 동일하지 않으므로 불량률 관리도를 사용했다(Raymond 등, 1997).

#### 다. 의사결정나무분석을 통한 병원감염발생의 분류 및 예측

병원감염발생의 분류 및 예측을 위해 데이터마이닝 기법 중의 하나인 Decision Tree를 사용하였고, 변수의 성격이 범주형 데이터이고 예측변수와 목표변수간의 관계를 찾아야 할 때 가장 유용

한 방법인 CHAID 방법을 적용하였다(최종후 등, 1998).

#### 라. 연관성 규칙을 통한 병원감염발생환자의 합병질환 특성 분석

병원감염발생환자군의 합병질환 특성분석을 위해 데이터마이닝 기법 중의 하나인 연관성 규칙을 적용했다. 연관성규칙의 평가는 신뢰도, 지지도, 향상도의 3가지 기준에 의해 이루어 진다. 신뢰도란 A를 구매했을 때 B를 구매할 확률이며, 지지도란 A와 B가 동시에 구매될 확률이다. 향상도는 규칙 ' $A \Rightarrow B$ ' 가 의미가 있다면 전체 거래의 내역에서 품목 B를 포함하고 있는 거래의 비율보다는 품목 A가 구매된 거래내에서 품목 B를 포함하고 있는 거래의 비율이 더 클 것이다. 즉 규칙 ' $A \Rightarrow B$ '의 리프트는 실제의 신뢰도를 독립 가정 하에서의 신뢰도로 나눈 값으로 그 값이 '1'보다 크면 양의 연관관계, '1'보다 작으면 음의 연관관계가 있다고 본다(강현철 등, 1999).

## III. 연구 결과

### 1. Statistical Process Control을 이용한 병원감염발생의 변동 추이

퇴원요약 DB에 포함되어 있는 질 개선 관련 항목인 질적 문제 중 병원감염의 추세분석을 위해 SPC 기법을 이용한 관리도로서 불량률 관리도 분석을 하였다.

그림 1과 같이 질적 문제 중 병원감염 발생률을 분석한 결과, 신뢰구간이 2 sigma일 때 중앙선 값이 21.3%로서 1996년과 1999년에 관리한계선을 벗어났으며, 특히 1999년이 문제가 됨을 알 수 있었다. 즉 불량률 관리도 분석을 통해 병원감염 발생률의 변동이 있음을 알 수 있었으며, 본 연구에서는 병원감염 발생에 대한 심층분석을 하기로 하였다(그림 1).

### 2. 병원감염발생의 규칙 및 위험집단

#### 가. 병원감염발생의 분류 및 예측

##### 1) 병원감염발생의 계층별 패턴

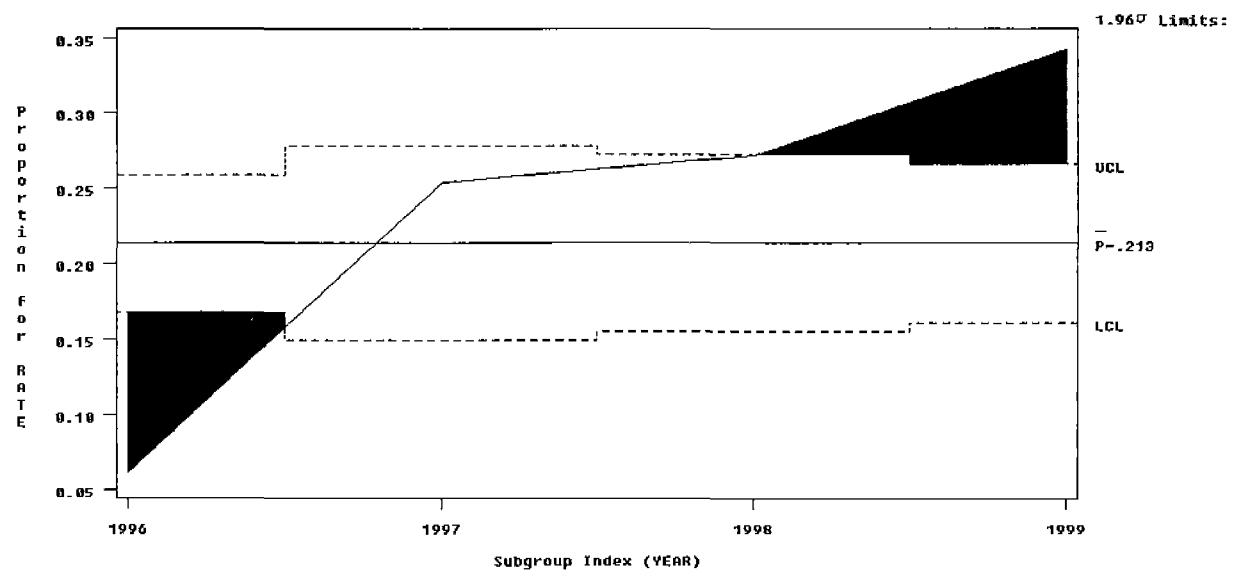


그림 1. 연도별 병원감염률의 불량률 관리도

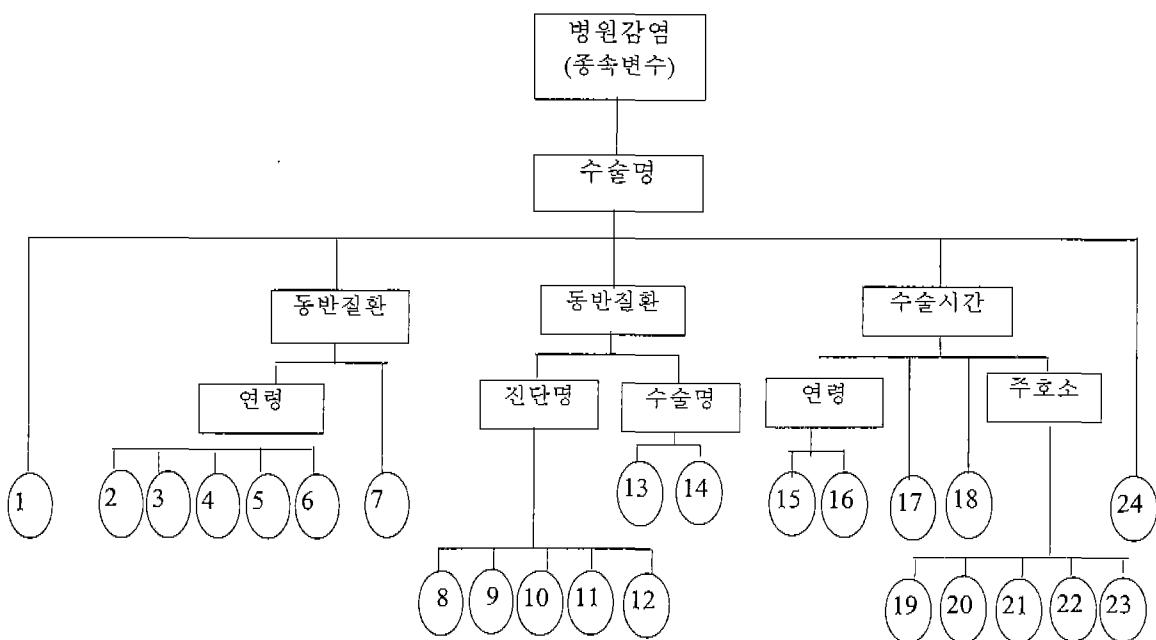


그림 2. 병원감염발생에 관한 의사결정도

그림 2와 같이 의사결정나무 분석을 한 결과, 병원감염발생 계층은 총 24단계로 분류되었다. 병원감염을 결정짓는 가장 중요한 변수는 의사결정 나무의 맨 위쪽에 위치하는 변수인 수술명이었다.

수술명 다음으로 동반질환과 수술시간이었으며, 동반질환 유무에 따른 연령, 진단명, 수술명이 나왔다. 또한 수술시간 다음으로 연령과 주호소가 주요 변수임을 알 수 있었다. 병원감염발생에 관한 각 계층별 유형은 다음과 같다(그림 2).

①번 계층을 결정하는 주요 규칙은 수술명이었다. 이 계층은 후두 및 기관수술을 시행받은 환자군으로서 병원감염 발생확률이 100%인 그룹이다.

②번~⑥번 계층을 결정하는 주요 규칙은 수술명, 동반질환유무, 연령으로 나타났다. 신경계 수술로서 뇌 및 척수 수술, 호흡기계 수술로서 흉막, 늑막 및 종격동 수술, 소화기계 수술로서 간 및 혀장 수술을 시행받은 환자군으로서 동반질환이 있으며, 연령별로는 41-50세일 때 병원감염 발생확률이 94.1%, 21-30세 및 51-60세일 때 68.5%, 61세 이상일 때 61.5%, 20세 이하일 때, 46.2%, 31-40세일 때 38.5 %였다. 이 중 연령이 41-50세일 때 병원감염 발생 확률이 가장 높으며, 31-40세일 때가 가장 낮았다.

⑦번 계층은 동일한 수술을 받은 환자군으로서 동반질환이 없는 경우이며, 병원감염 발생확률은 14.3%로 나왔다.

⑧번~⑫번 계층을 결정하는 주요 규칙은 수술명, 동반질환 유무, 진단명으로 나타났다. 수술명이 소화기계 수술로서 타액선 수술, 위 절개 및 절제술, 장관 절개 .절제 및 문합술, 기타 장관수술, 직장 및 직장-S상 결장 및 주위조직 수술, 담낭 및 담관 수술, 복벽절개술, 호흡기계 수술로서 후두수술, 폐 및 기관지 수술, 순환기계 수술로서 심장 판막 및 종격수술, 심장혈관 수술, 기타 혈관수술, 혈액 및 임파계의 수술로서 골수 및 비장수술, 비뇨기계 수술로서 방광수술, 요도수술, 내분비계 수술로서 갑상선 및 부갑상선 수술, 근골격계 수술로서 골의 절개. 절제 및 분리술, 관절의 절개 및 절제술, 관절재건술을 시행받은 환자군으로서 동반 질환이 있으며, 진단명이 신경계 질환, 주산기 병태, 기타 증상 ·건강 상태에 영향을 주는 증상일 때 병원감염 발생확률이 100%, 내분비 ·영 양 및 대사 질환 일 때 75%, 순환기기계 질환, 비뇨생식기계 질환, 선천성 질환, 기형 ·변형 및 염색체 이상일 때 48.9%, 감염성 및 기생충성 질환, 신생물, 호흡기계 질환, 소화기계 질환, 손상 및 중독에 의한 질환일 때 28%, 정신 및 행동장애, 근골격계 및 결합조직의 질환, 임신, 출산 및 산후기 질환일 때 10.9 %이었다. 여기에서도 역시 진단명이 신경계를 포함하여 주산기병태 및 기타 증상 ·건강상태에 영향을 주는 증상일 때 병원감염 발생확률이 가장 높게 나왔다.

⑬번~⑭번 계층은 수술명이 ⑧번~⑫번 계층과 동일하고 동반 질환이 없으며, 수술명이 장관

절개 ·절제 및 문합술일 때 병원감염 발생확률은 100%였고, 이 외의 수술일 때 3% 였다. 소화기계 수술 중 장관 절개 ·절제 및 문합술일 때의 감염발생률이 높게 나왔다.

⑯번~⑯번 계층을 결정하는 주요 규칙은 수술명, 수술시간, 연령으로 나타났다. 코 ·입 및 인두 수술로서 혀 수술, 편도 및 아데노이드 수술, 임파계 수술, 소화기계 수술로서 총수돌기 수술, 비뇨생식기계 수술로서 신장 수술, 전립선 및 정액낭 수술, 난소 수술, ·자궁 절개 및 절제술, 근골격계 수술로서 안면골 및 판절수술, 그 외의 골 수술, 골절 및 탈골 정복술, 근 ·건 ·건막 및 활액낭 수술, 기타 골 ·판절 수술, 피부 및 피하 조직 수술을 시행 받은 환자군으로서 수술시간이 60분 이하이며 연령이 61세 이상일 때 병원감염 발생확률은 13.2%, 60세 이하일 때 0.5%였다. 이 그룹은 수술시간이 짧은 비교적 경증의 환자이지만 연령이 61세 이상인 경우의 병원감염율이 연령이 낮은 그룹에 비해 높음을 알 수 있었다.

⑰번~⑲번 계층은 수술명이 ⑯번~⑯번 계층과 동일하고 수술시간이 61분~120분일 때 병원감염 발생확률은 10.7%였고, 수술시간이 121분~180분일 때에는 13.3%였다. 비교적 수술시간이 긴 수술을 받은 환자군의 병원감염 발생확률이 높음을 알 수 있었다.

⑳번~㉓번 계층을 결정하는 주요 규칙은 수술명, 수술시간, 주호소로 나타났다. 수술명은 ⑯번~⑯번 계층과 동일하고 수술시간이 181분 이상이며 주호소가 신경계통일 때 병원감염 발생확률은 100%, 소화기계통일 때 60%, 외상일 때 22.7%, 근 ·골격계통일 때 14.3%, 비뇨기계통일 때 0%이었다. 수술시간이 가장 길고 주호소가 신경계통일 때 병원감염 발생확률이 100%임을 알 수 있었다.

㉔번 계층은 내분비계 수술로서 기타 내분비선 수술, 결막을 제외한 안과수술, 귀 수술, 혀 ·타액선 ·편도 및 아데노이드를 제외한 코 ·입 및 인두 수술, 호흡기계수술로서 기타 폐 및 기관지 수술, 순환기계 수술로서 심장 및 심장 주위 수술, 혈관의 절개 ·절제 및 교합수술, 신경계 수술로서 뇌 및 말초신경 수술, 소화기계 수술로서 식도수술, 위수술, 항문수술, 헤르니아 수술, 비뇨생식기계 수술로서 요관수술, 기타 요도수술, 고환수술, 정관 ·부고환 수술, 음경수술, 나팔관 수술, 자궁경부수술, 기타 자궁 부속기 수술, 질 및 맹낭 수술, 외음 및 회음 수술, 산과계 수술로서 제왕절개술, 근 ·골격계 수술로서 손의 근 ·건 및 건막 수술, 외피계 수술로서 유방 수술, 기타 수술을 시행 받은 환자군으로 병원감염 발생확률은 0%였다.

## 2) 병원감염 발생의 규칙

표 1과 같이 병원감염 발생의 분류와 예측의 과정을 나무구조에 의한 추론규칙에 의해

표현하였으며, 이를 통해 대상군별 환자의 특성을 규명하고 이로써 각 대상군, 즉 병원감염 발생군과 병원감염 미발생군을 결정짓는 규칙을 생성할 수 있었다. 또한 이와 같은 예측분류에 대한 신뢰도를 측정함으로써 예측가능성을 첨수화하였다.

절개, 절제 및 문합술, 기타 장관수술, 직장 및 직장-S상결장 및 위조직수술, 담낭 및

표 1. 병원감염발생의 규칙

예측 집단	규칙	수술명	동반 질환	수술시간	연령	진단명	주호소	예측분류 에 대한 신뢰도(%)
병원 감염 발생 그룹	1	후두 및 기관수술	-	-	-	-	-	100.0
	2	<sup>1)</sup> 타액선수술	유	-	-	<sup>6)</sup> 신경계질환 등	-	100.0
	3	<sup>2)</sup> 혀수술 등	-	3시간 이상	-	-	신경계 및 감각계	100.0
	4	장관외 절개, 절제 및 문합술	무	-	-	-	-	100.0
	5	<sup>3)</sup> 뇌수술 등	유	-	41-50	-	-	94.1
	6	타액선수술 등	유	-	-	내분비, 영양 및 대사질환	-	75.0
	7	뇌수술 등	유	-	21-30, 51-	-	-	68.5
	8	뇌수술 등	유	-	61-	-	-	61.5
	9	혀수술 등	무	3시간이상	-	-	소화기계	60.0
병원 감염 미 발생 그룹	1	<sup>4)</sup> 안검수술 등	무	-	-	-	-	100.0
	2	혀수술 등	-	3시간이상	-	-	비뇨기계	100.0
	3	혀수술 등	-	1시간이내	-60	-	-	99.5
	4	<sup>5)</sup> 장관수술을 제외한 수술	무	-	-	-	-	97.0
	5	혀수술 등	-	1-2시간	-	-	-	89.3
	6	타액선수술 등	유	-	-	<sup>7)</sup> 정신 및 행동장애 등	-	89.1
	7	혀수술 등	-	1시간이내	61-	-	-	86.8
	8	혀수술 등	-	2-3시간	-	-	-	86.7
	9	혀수술 등	-	3시간이상	-	-	근골격계	85.7
	10	뇌수술 등	무	-	-	-	-	85.7
	11	혀수술 등	-	3시간이상	-	-	외상 및 부작용	77.3
	12	타액선수술 등	유	-	-	<sup>8)</sup> 감염성 및 기생충질환 등	-	72.0
	13	뇌수술 등	유	-	31-40	-	-	61.5
	14	뇌수술 등	유	-	-20	-	-	53.8
	15	타액선수술 등	유	-	-	<sup>9)</sup> 순환기계질환	-	51.1

- 1) 타액선수술 등 : 타액선수술, 후두수술, 폐 및 기관지수술, 심장 판막 및 종격수술, 심장혈관수술, 기타 혈관수술, 골수 및 비장수술, 위 절개 및 절제술, 장관

담관수술, 복벽절개술, 방광수술, 요도수술, 갑상선 및 부갑상선수술, 골의 절개절제 및 분리술, 관절의 절개 및 절제술, 관절재건술

규칙	Node	Node:n	Node:%	Response:n	Response:%	Gain(%)	Index(%)
1	2	31	2.00	31	12.80	100.00	628.93
2	24	4	0.26	4	1.65	100.00	628.93
3	25	1	0.006	1	0.41	100.00	628.93
4	33	1	0.006	1	0.41	100.00	628.93
5	15	17	1.11	16	6.60	94.12	591.82
6	23	4	0.26	3	1.23	75.00	471.69
7	16	54	3.54	37	15.28	68.50	430.81
8	17	26	1.70	16	6.61	61.50	386.79
9	32	5	0.32	3	1.23	60.00	377.35
10	22	45	2.95	22	9.09	48.90	307.54
11	18	13	0.85	6	2.47	46.20	290.56
12	19	13	0.85	5	2.06	38.50	242.13
13	21	175	11.49	49	20.24	28.00	176.10
14	31	22	1.44	5	2.06	22.70	142.76
15	30	7	0.45	1	0.41	14.30	89.93
:							

2) 혁수술 등 : 혁수술, 편도 및 아데노이드수술, 임파계수술, 충수돌기수술, 신장수술 전립선 및 정액낭수술, 난소수술, 자궁질개 및 절제술, 안면골 및 관절수술, 그 외 골수술, 골절 및 탈골정복술, 근·건·건막 및 활액낭수술, 기타 골관절수술, 피부 및 괴화조직수술

3) 뇌수술 등 : 뇌 및 두개골 뇌막수술, 척수수술, 흉막늑막종격동수술, 간수술, 췌장수술

4) 안검수술 등 : 안검수술, 누관수술, 결막을 제외한 안과계수술, 귀수술, 혁·타액선 및 아데노이드수술을 제외한 코·입·인두수술, 기타 폐 및 기관지수술, 심장 및 심장주위수술, 혈관의 절개·절제 및 교합수술, 뇌 및 말초신경수술, 식도수술, 위수술, 항문수술, 헤르니아수술, 요관수술, 기타요도수술, 고환수술, 정관부고환수술, 음경수술, 나팔관수술, 자궁 경부수술, 기타 자궁부속기수술, 질 및 맹낭수술, 기타 내분비선수술, 계왕절개술, 손의 근·건 및 건막수술, 유방수술, 치료적 교정장치의 치환 및 제거, 결석 및 이물질의 제거

5) 장관수술을 제외한 수술 : 2)타액선수술 등에서 장관 절개·절제 및 문합술을 제외한 수술

6) 신경계질환 등 : 신경계질환, 주산기병태, 기타증상, 건강상태에 영향을 주는 증상

7) 정신 및 행동장애 등 : 정신 및 행동장애, 근골격계 및 결합조직의 질환, 임신, 출산 및 산욕기 질환

8) 감염성 및 기생충성 질환 등 : 감염성 및 기생충성 질환, 신생물, 호흡기계질환, 소화기계 질환, 손상 및 중독에 의한 질환

9) 순환기계질환 등 : 순환기계질환, 비뇨생식기계질환, 선천성질환, 기형, 변형, 염색체 이상

표 2. 병원감염발생 환자군의 이익도표  
(Node-by-Node 통계량)

### 3) 병원감염발생 위험집단의 선정

병원감염 발생의 계층별 패턴과 규칙을 추출한 후, 병원감염 발생그룹을 세분화하여 다른 그룹에 비해 병원감염 발생의 위험도가 높은 감염관리 대상을 찾아내기 위해 이익도표를 활용하였다.

표 2의 Node-by-Node 통계량을 보면, 'Node'는 노드의 번호를 나타내고, 'Node:n'은 마디에 속하는 개체의 수, 'Respn:n'은 마디에 속하는 목표범주의 개체의 수, 'Gain(%)'은 마디에 속하는 목표범주의 개체의 수/마디에 속하는 개체의 수, 'Index(%)'는 특별한 마디에서의 목표 범주의 비율/전체에서의 목표범주의 비율을 나타낸다 (최종후 등, 1998).

규칙 1은 노드 2에 해당하고 이 때의 관찰된 개체의 수는 31명이고, 이는 전체 대상자 1,523명의 2%에 해당되며, 관찰된 개체수 31명 중 31명이 병원감염 발생 환자로서 이는 전체 병원감염 발생 환자군 242명의 12.8%에 해당된다. 즉 규칙 1의 특성을 지닌 대상이 전체 병원감염 발생환자군 중 12.8%에 해당되는 것을 의미하며, 'Gain(%)'이 '100'으로 이는 규칙 1의 특성을 지닌 대상이 병원감염 발생환자군일 확률이 100%임을 나타낸다. 또한 'Index(%)'가 '628.93'으로 이는 전체 대상자 중 병원감염 발생군의 비율보다 규칙 1의 조건을 만족하는 군에서의 병원감염 발생군의 비율이 6배 높음을 의미한다.

Node	Gain(%)	Index(%)	Cumulative					
			Node:n	Node:%	Response : n	Response : %	Gain(%)	Index(%)
2	100.00	628.93	31	2.00	31	12.80	100.00	628.93
24	100.00	628.93	35	2.29	35	14.46	100.00	628.93
25	100.00	628.93	36	2.36	36	14.87	100.00	628.93
33	100.00	628.93	37	2.42	37	15.28	100.00	628.93
15	94.12	591.82	54	3.54	53	21.90	98.14	617.23
23	75.00	471.69	58	3.80	56	23.14	96.55	607.23
16	68.50	430.81	112	7.35	93	38.42	83.03	522.20
17	61.50	386.79	138	9.06	109	45.04	78.98	496.72
32	60.00	377.35	143	9.38	112	46.28	78.32	492.57
22	48.90	307.54	188	12.34	134	55.37	71.27	448.23
18	46.20	290.56	201	13.19	140	57.85	69.65	438.05
19	38.50	242.13	214	14.05	145	59.91	67.75	426.10
21	28.00	176.10	389	25.54	194	80.16	49.87	313.64
31	22.70	142.76	411	26.98	199	82.23	48.41	304.46
30	14.30	89.93	418	27.44	200	82.64	47.84	300.88
:								

수립되어진다면 다른 대상집단보다 상대적으로 높은 이득을 얻을 수 있음을 알 수 있었다

표 3. 병원감염발생 환자군의 이익도표  
(누적통계량)

다음으로 규칙 14는 노드 31에 해당되고 이 때의 관찰된 개체의 수는 22명이고, 전체 대상자 1,523명의 1.4%에 해당되며, 관찰 된 개체수 22 명 중 5명이 병원감염 발생환자로서 이는 전체 병원감염 발생환자군 242명의 2%에 해당된다. 즉 규칙 14의 특성을 지닌 대상이 전체 병원감염 환자군 중 2%에 해당하는 것을 의미하며, 'Gain(%)'가 '22.7'로 이는 규칙 14의 특성을 지닌 대상이 병원감염 발생환자군일 확률이 22.7%임을 나타낸다. 또한 'Index(%)'가 '142.76'으로 이는 전체 대상자 중 병원감염 발생환자군의 비율보다 규칙 14의 조건을 만족하는 군에서의 병원감염 발생군의 비율이 1.4배 높음을 의미한다.

이와 같은 결과는 규칙 1보다는 크지는 않지만 전체 자료보다는 상당히 높은 병원감염 발생률을 보이고 있다.

이와 같은 방식으로 규칙 15로서 마디 30을 보면, 이 마디에 대한 인덱스 점수는 약 90%정도로서 작은 인덱스 점수인데, 이에 해당하는 마디는 전체 자료보다 낮은 병원감염발생률을 갖는다는 것을 알 수 있다. 따라서 마디 30과 마디 31 사이에 교차점(전체자료보다 높은 병원감염 발생률을 보이는 마디와 낮은 병원감염 발생률을 보이는 마디의 교차점)이 있음을 알 수 있다(표 2).

따라서 병원감염발생 위험집단으로서 인덱스 점수가 높은 대상집단에 우선적인 질관리 전략이

또한 병원감염발생 위험집단의 관리 대상 범위를 설정하기 위해 병원감염발생 위험집단에 대한 최적의 세분화가 필요했다. 이익도표의 누적 통계량은 병원감염관리의 대상집단 선정시 의사결정에 필요한 정보를 제공해주고 있다. 예를 들어 가장 병원감염 발생률이 높은 마디 2 단을 취한다면 전체 자료의 2%(31/1,523)의 환자를 대상으로 하여 전체 감염환자군의 12.8%(31/242)의 병원감염 환자군을 관리할 수 있다. 여기서 추가적으로 마디 24를 포함한다면 2.3%의 환자를 대상으로 하여 전체 감염환자군의 14.5%의 병원감염 환자군을 관리할 수 있다. CQI활동의 효율성을 높이기 위해서는 감염관리 노력에 비해 병원감염 관리효과가 큰 집단에 우선적인 감염관리 활동전략이 수립되어져야 한다. 그러기 위해서는 마디 23까지를 병원감염 관리대상으로 포함해야 할 것이다. 이는 전체 대상자 중 병원감염 발생군의 비율보다 규칙 6까지의 조건을 만족하는 군에서의 병원감염 발생군의 비율이 6배 높을 뿐 아니라, 병원감염 전체 환자 중 3.8%를 감염관리 대상집단으로 관리하면 전체 병원감염환자군의 23.1%의 병원감염 관리효과를 가져올 수 있기 때문이다(표 3).

### 3. 연관성 규칙을 활용한 병원감염 발생환자의 합병질환 특성

병원감염 환자들에게 동반되는 감염의 종류 및 특성을 파악하기 위해 병원감염 환자군의 주진단명과 합병증진단명과의 연관성 규칙을 적용했다. 요로감염이 주진단인 경우 원인균이 되는 대장균이 기타진단으로 나올 확률이 일반환자보다 4배나 높다는 것을 알 수 있다. 수술부위의 파괴라는 진단명을 가진 환자가 수술부위의 창상감염을 동반할 확률은 75%로서 수술부위의 파괴는 수술부위의 창상감염을 가져오기 쉽다는 것을 의미한다. 또한 뇌내출혈이라는 진단명을 가진 환자가 도뇨관으로 인한 요로감염을 동반할 확률이 일반환자보다 1.4배 높다는 것을 알 수 있다. 여기에서 병원감염 중 요로감염 및 수술후 창상감염이 많다는 것을 알 수 있었다(표 4).

표 4. 병원감염 환자의 주진단과 합병증진단과의 연관성

연관상병	향상도	지지도 (%)	신뢰도 (%)
요로감염 => 대장균	4.04	4.08	20.00
대장균 => 요로감염	4.04	4.08	82.35
요로감염 => 기타 원인균	3.68	3.50	17.14
수술부위의 파괴 => 수술부위의 창상감염	2.96	3.50	75.00
수술부위의 창상감염 => 수술부위의 파괴	2.96	3.50	13.79
기타 원인균 => 요로감염	3.68	3.50	75.00
뇌내출혈 => 도뇨관으로인한 요로감염	2.72	1.46	55.56
천공성 충수염 => 수술부위의 창상감염	3.29	1.46	83.33
녹농균 => 요로감염	2.80	1.17	57.14
외상성 경막하 출혈 => 요로감염	3.27	1.17	66.67
치치에 합병된 출혈 => 수술부위의 창상감염	1.75	1.17	44.44
뇌내출혈 => 도뇨관으로인한 요로감염	2.80	1.17	57.14
내과적·외과적 치료도중 무균의 실패 => 수술부위의 창상감염	2.25	1.17	57.14

### 4. 데이터마이닝 기법을 통해 도출된 병원감염발생 규칙의 Continuous Quality Improvement 활동에의 적용

최근 CQI활동이 전개되면서 병원감염관리도 예방효과를 높이기 위한 감시활동 중심으로 이루어지고 있다. 그러나 병원감염에 높은 위험도를 가진 환자를 찾아내는 데 만족할 만큼 민감한 전향적 감시체계가 없다(김정순, 1988). 본 연구에서는 병원감염 발생규칙을 추출하여 병원감염에 높은 위험도를 가진 환자군을 미리 예측해 줌으로써 병원감염 발생확률이 높은 집단에 대한 집중관리를 하게 하며, 병원감염 관리의 전향적인 감시체계의 구축을 가능하게 하였다. 또한 병원감염 관리대상의 범위를 선정하는 데 이의도표를 통해 감염관리로 인한 이익효과를

사전에 알려 줌으로써 병원감염관리 활동을 위한 전략수립시 의사결정에 유익한 정보를 제공해 줄 수 있었다.

본 연구결과 병원감염발생에 관한 예측분류에 대한 신뢰도가 50%이상인 경우를 기준으로 하여 아래와 같이 병원감염발생패턴을 선정할 수 있었다.

첫째, 수술명이 후두 및 기관수술인 경우였다.  
둘째, 수술명이 타액선수술, 후두수술, 폐 및

기관지수술, 심장 판막 및 종격수술, 심장혈관수술, 기타 혈관수술, 골수 및 비장수술, 위 절개 및 절제술, 장관 절개, 절제 및 문합술, 기타 장관수술, 직장 및 직장-S상결장 및 주위조직수술, 담낭 및 담관수술, 복벽절개술, 방광수술, 요도수술, 갑상선 및 부갑상선수술, 골의 절개·절제 및 분리술, 관절의 절개 및 절제술, 관절재건술 중 어느 하나를 시행받았고 동반질환이 있으며 질환명이 신경계 질환, 주산기 병태, 기타 중상, 건강상태에 영향을 주는 증상인 경우였다.

셋째, 주호소가 신경계 및 감각계이면서 혀수술, 편도 및 아데노이드수술, 임파계수술, 충수돌기수술, 신장수술, 전립선 및 정액낭수술, 난소수술, 자궁절개 및 절제술, 안면골 및 관절수술, 그외 골수술, 골절 및 탈골정복술, 근·건·건막 및 활액낭수술, 기타 골관절수술, 피부 및 피하조직수술 중 어느 하나를 시행받았고, 수술시간이 3시간 이상인 경우였다.

넷째, 수술명이 장관의 절개, 절제 및 문합술 중 어느 하나를 시행받았고 동반질환이 없는 경우였다.

다섯째, 수술명이 뇌 및 두개골 뇌막수술, 척수수술, 흉막·늑막·종격동수술, 간수술, 췌장수술 중 어느 하나를 시행받았고 동반질환이 있으면서, 연령이 41-50세인 경우였다.

여섯째, 수술명이 타액선수술, 후두수술, 폐 및 기관지수술, 심장 판막 및 종격수술, 심장혈관수술, 기타 혈관수술, 골수 및 비장수술, 위 절개 및 절제술, 장관 절개, 절제 및 문합술, 기타 장관수술, 직장 및 직장-S상결장 및 주위조직수술, 담낭 및 담관수술, 복벽절개술, 방광수술, 요도수술, 갑상선 및 부갑상선수술, 골의 절개·절제 및 분리술, 관절의 절개 및 절제술, 관절재건술 중 어느 하나를 시행받았으며, 동반질환이 있고, 진단명이 내분비, 영양 및 대사질환 중 어느 하나인 경우였다.

일곱째, 수술명이 뇌 및 두개골 뇌막수술, 척수수술, 흉막·늑막·종격동수술, 간수술, 췌장수술 중 어느 하나를 시행받았고 동반질환이 있으며, 연령이 21-30세, 51-60세인 경우였다.

여덟째, 수술명이 뇌 및 두개골 뇌막수술, 척수수술, 흉막·늑막·종격동수술, 간수술, 췌장수술 중 어느 하나를 시행받았고 동반질환이 있으며, 연령이 61세 이상인 경우였다.

아홉째, 수술명이 혀수술, 편도 및 아데노이드수술, 임파계수술, 충수돌기수술, 신장수술, 전립선 및 정액낭수술, 난소수술, 자궁절개 및 절제술, 안면골 및 관절수술, 그외 골수술, 골절 및 탈골정복술, 근·건·건막 및 활액낭수술, 기타 골관절 수술, 피부 및 피하조직수술 중 어느 하나를 시행받았고, 수술시간이 3시간 이상이며, 주호소가 소화기계인 경우였다

## VII. 결론 및 고찰

퇴원요약 DB를 이용하여 데이터마이닝 기법의 CQI 활동에 활용 방안을 제시하고자 한 대학병원에서 4년간 퇴원한 환자 88,593명의 퇴원요약 DB를 분석한 결과, 다음과 같은 결과를 얻었다.

퇴원요약 DB로부터 질관리의 일환으로 SPC 기법을 활용하여 병원감염 발생률의 변동을 파악한 결과, 병원감염 발생률이 관리한계선을 벗어나 있었으며 이에 대한 심층분석이 필요함을 알 수 있었다.

병원감염발생의 요인 및 패턴을 찾기 위해 데이터마이닝 기법 중 하나인 의사결정나무 기법을 적용한 결과, 병원감염을 설명하는 가장 중요한 변수는 수술명이었고, 그 다음으로 동반질환유무와 수술시간, 동반질환유무 다음으로 연령 및 진단명이었고, 수술시간 다음으로 연령 및 주호소였다.

의사결정나무기법을 통하여 24개의 병원감염 발생규칙을 도출해 낼 수 있었다. 이 중 예측분류에 대한 신뢰도가 50%이상인 9가지의 규칙을 지침으로 제시하였다. 이익도표의 노드별 통계량과 이익도표의 누적 통계량을 통해 병원감염관리 대상범위를 설정할 수 있었으며, 설정된 범위에 따른 병원감염관리 효과도 미리 예측할 수 있었다. 즉, 마디 23까지의 대상군은 전체 대상자 중 병원감염발생군의 비율보다 마디 23까지의 조건을 만족하는 군에서의 병원감염발생군의 비율이 6배 높은 대상군이므로 우선적인 병원감염 관리대상으로 보아야 할 것이며, 또한, 마디 23까지를 포함해서 전체 환자 중 3.8%를 감염관리하면 전체 병원감염환자의 23.1%의 병원감염을 관리효과를 예측할 수 있었다.

주진단과 합병증 진단명 간의 연관성규칙을 적용한 결과, 오로감염과 그 원인균을 나타내는 상병과의 지지도가 '4.08'로서 일반환자보다 4배나 높은 연관성을 보였고, 뇌내출혈의 진단명을 가진 환자가 도뇨관으로 인한 오로감염을 동반할 경우 지지도가 '1.46'으로서 일반환자보다 1.4배 높은 연관성을 보였다. 수술부위의 과괴라는 진단명을 가진 사람이 수술부위의 창상감염이라는 진단명을 동반할 경우 신뢰도 '75'로서 수술부위의 과괴는 수술부위의 창상감염을 가져오기 쉽다는 것을 알 수 있었다.

보건의료분야에서 SPC 기법을 CQI 활동에 활용함으로서 의료의 질 변동을 밝히며 동일수준의 의료서비스를 유지할 수 있도록 모니터링할 수 있게 해주며, 이에 대해 요인 및 패턴 분석을 위해 데이터마이닝 기법을 적용하여 방대한 데이터로부터 지식추출을 시도함으로서, 데이터만을 이용하여 정보기술을 활용한 효율적인 CQI 활동에의 적용 가능성을 부여했다는 데에 그 의미가 있다 하겠다.

제한점으로 첫째, 후향적인 방법으로서 퇴원분석 DB를 이용하였고, 변수의 제한성, 형생제 남용, 불완전한 의무기록 등으로 병원감염 발생률을 충분히 반영했다고 볼 수 없다. 둘째, 연구대상이

한 의과대학 부속병원에 국한되어있을 뿐 아니라 병원마다 환자의 특성 및 환경 등이 다르며, 이에 따른 병원감염의 현황도 다르기 때문에, 이 연구결과를 일반화시키기에는 무리가 있다.셋째, 본 연구는 병원감염발생에 대한 기존의 경험적 사실을 실증적으로 확인한 것에 대한 의미는 있지만, 이를 효과적으로 관리하기 위해 문제점이나 대체방안을 구체적으로 제시하지는 못하였다.

이 연구의 결과를 기초로 하여 퇴원요약 DB는 물론 다양한 병원데이터를 이용하고, 통계적 공정관리기법 및 데이터마이닝 기법의 적용대상도 병원감염 뿐 아니라 더욱 확대 적용하여 정보기술을 이용한 CQI 활동을 더욱 발전시킬 수 있는 계기가 되어야 할 것이다.

## 참 고 문 헌

- 강성홍, 강수원, 김미숙 등. 보건정보관리학.  
청구문화사, 1999
- 강현철, 한상태, 최종후 등. SAS Enterprise Miner를  
이용한 데이터마이닝; 방법론 및 활용.  
자유아카데미, 1999
- 김옥남, 강성홍. 데이터마이닝을 활용한  
진료데이터 분석. 대한의료정보학회지  
1999; 15(1): 194
- 김정순. 병원감염. 수문사, 1988
- 박성현, 박영현, 이명주. 통계적 공정관리. 민영사,  
1998
- 안선경. CQI의 개념과 필요성. 한국의료QA학회  
1999: 1-20
- 옹환승. 데이터마이닝. 그린, 1998
- 이경미, 21세기 의무기록분야의 대응전략.  
병원관리종합학술대회연제집, 1999; 15(1):  
237-251
- 이정화, 의무기록의 표준화를 위한  
퇴원진료정보분석. 연세대학교 보건  
대학원 석사논문, 1996
- 조우현, CQI의 개념과 필요성. 한국의료QA학회,  
1999: 1-20
- 채영문. 보건정보관리 .수문사, 1994
- 최종후, 한상태, 강현철 등. Answer Tree를 이용한  
데이터마이닝 의사결정나무분석, 1998
- 홍준현. 의무기록관리학. 고문사, 1996
- Greg Rogers, Ellen Joyner. Mining your data for health  
care quality improvement. SAS Institute,  
Inc., Cary, NC, 1998
- Internetsite[<http://www.cio.seoul.kr/main4.html>].
- Raymond GC, Robert CL. Quality with confidence in  
healthcare, 1997