# 퇴원요약 데이터베이스를 이용한 데이터마이닝 기법의 CQI 활동에의 활용 방안

연세대학교 보건대학원

보건정보관리학과

선 미 옥

# 퇴원요약 데이터베이스를 이용한 데이터마이닝 기법의 CQI 활동에의 활용 방안

지도 이 해 종 교수

이 논문을 보건학 석사학위논문으로 제출함.

2000年 6月 日

연세대학교 보건대학원

보건정보관리학과

선 미 옥

선미옥의 보건학 석사학위논문을 인준함.

심사위원 ---------- 인

심사위원 ---------- 인

심사위원 ---------- 인

연세대학교 보건대학원

2000年 6月 日

# 감사의 글

2년 6개월이라는 대학원생활의 결실로 미흡하지만 논문을 내게 되어 보람과 부끄러움을 같이 느낍니다. 부족한 저에게 논문의 전체적인 흐름을 잡아주시고 체계적으로 정리하도록 도와주신 이해종 교수님, 전공지도 교수님으로 이러한 결실이 있기까지 대학원 생활 내내 연구하는 자세와 배움의 기회를 주신 채영문 교수님, 바쁘신 중에도 세심하게 지도해 주신 이선희 교수님께 머리숙여 감사를 드립니다.

논문을 수행하면서 많은 도움을 주신 인제대학교 강성홍 교수님, 마지막까지 성심성의껏 도와주신 호승희 선생님께 깊은 감사를 드립니다. 그리고 어려운 여건 속에서도 공부할 수 있도록 배려해 주신 병원장님 이하 직장 동료 여러분께 진심으로 감사의 마음을 전합니다.

또한 힘든 업무 중에서도 저를 도와주고 이해해 준 김명화선생, 김성주 선생을 포함한 의무기록과 식구들, 제게 처음 의무기록을 공부할 수 있게 해 주시고, 지금까지 이끌어 주신 세브란스병원 의무기록과 홍준현 과장님 그리고 저를 도와주신 주위의 모든 분들께 감사를 드립니다.

대학원생활 동안 연로하신 몸으로 집안 일을 도와 주신 시어머님, 멀리서 저를 지켜봐 주신 어머님께 깊은 감사를 드립니다.

무엇보다도 가까이서 물심양면으로 도움을 아끼지 않았던 사랑하는 남편, 그리고 바쁜 엄마를 이해해주고 건강하게 잘 자라준 두 아들 수영, 주영이와 함께 이 기쁨을 나누고자 합니다.

2000. 6.

선 미 옥

# 목 차

# 표 목 차

# 그 림 목 차

# 국 문 요 약

퇴원요약 DB를 이용한 데이터마이닝 기법의 지속적 질 관리활동에의 활용 방안을 제시하고자, 퇴원요약 DB를 토대로 병원감염발생의 현황파악을 위해 통계적 공정관리기법을 적용한 결과, 병원감염발생률에 변동이 보였으며, 이의 요인 및 패턴을 찾기 위해 데이터마이닝 기법 중 하나인 의사결정나무기법을 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 도출해낼 수 있었다. 이 중 예측분류에 대한 신뢰도가 50% 이상인 9가지의 규칙을 지침으로 제시하였다. 이익도표의 노드별 통계량과 누적 통계량을 통해 병원감염관리 대상범위를 설정할 수 있었으며, 설정된 범위에 따른 병원감염관리효과도 미리 예측할 수 있었다.

감염환자의 주진단명과 합병증진단명간의 연관성규칙을 적용한 결과, 요로감염과 그 원인균간의 지지도가 가장 높았고, 수술부위의 파괴와 수술부위의 창상감염 사이에도 높은 지지도를 보였다.

데이터마이닝 기법으로 병원감염발생규칙을 추출함으로써 병원감염의 전향적인 감시활동을 가능하게 했다. 이 연구의 결과를 기초로 하여, 퇴원요약 DB만이 아닌 다양한 병원테이터를 이용하고, 데이터마이닝 기법의 적용대상도 병원감염 뿐 아니라 더욱 확대 적용하여 정보기술을 이용한 지속적 질 관리활동을 더욱 발전시킬 수 있는 계기가 될 것으로 기대한다.

# 1. 서 론

## 1. 연구 배경

의료환경의 변화는 질병구조, 의료서비스 이용형태, 의료체계, 국민의식 등의 변화를 가져왔고, 의료계 주변여건 역시 병실료 단입제, 신용카드 수납, 의약분업, 선택진료제, 포괄수가제 확대실시, 약제비 실거래가 상환제, 자동차보험수가 인하, 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 (TOM, 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 지식경영이라 할 수 있다. 앞으로의 병원경영은 종전과는 달리 환자에게 최소의 적정서비

스를 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 축적할 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.

질 향상활동도 단편적인 사건들에만 집착한 표면적인 사건분석방법으로는 얻을 수 없는 합리적인 의사결정의 수단으로서의 자료수집 및 분석이 이루어져야 할 것이다. 자료수집방법에 있어서도 동시적인 평가는 평가대상인 환자 또는 이용객 등이 의료서비스를 이용하고 있는 중에 자료를 수집하여 질적인 측면에서 위험요소를 발견하면 즉시 개선하여 환자에게 생길 수 있는 질적 문제들을 예방할 수 있지만, 시간과 비용 또는 인력이 많이 소요되는 단점이 있으며, 후향적인 평가는 조사가 간편하고 결과적인 측면에서 종합적인 자료를 얻을 수 있지만, 사정이나 감사의 역할을 주로 하고 이미 일어난 질적인 문제는 개선할 수 없는 기존의 한계가 있다(안선경, 1999). 이에 신속한 자료 분석과 신속한 결과 제시 및 예측까지도 가능한 정보기술의 활용과 질 향상활동을 연계할 필요성이 있다.

데이터마이닝이란, 최근에 나온 데이터 분석기법으로 대용량의 데이터로부터 의미있는 중요한 사실을 찾아내어 비즈니스 업무에 적용하고자 하는 데 목적을 두고 있는 것으로, 대용량의 데이터에 숨겨져 있는 데이터간의 관계, 패턴을 탐색하고 이를 모형화 하여 업무에 적용할 수 있는 의미있는 정보로 변환함으로써 기업의 의사결정에 적용하는 일련의 과정으로 정의되고 있다(용환승, 1998). 선진국에서는 의료기관에서 정보를 지식으로 전환하여 비용관리를 돕고, 환자진료의 질을 향상시키기 위해서 데이터마이닝기법을 이용하고 있고, 이 기법을 이용함으로써 미래의 환자행동을 예측하고 진료프로그램을 향상시키는 데 이용될 수 있다. 또한, 고위험군 환자를 파악함으로써 임상의사는 오늘의 환자에게 더욱 양질의 환자관리를 할 수 있어 내일의 문제가 되지 않게 된다(Greg 등, 1999).

앞서 언급한 바와 같이 의료분야에서 처방전달시스템, 의무기록전산화가 구현되고 대용량의 데이터가 축적됨에 따라 이제 단순한 통계패키지로 내용을 분석하는 것은 한계가 있으며, 병원 업무를 위해 구축된 대용량의 데이터로부터 지식을 추출하여(Prather 등, 1997) CQI 활동에 활용하기 위해서 최근에 데이터 분석도구로 각광을 받고 있는 데이터마이닝 기법의 활용이 요구된다.

이에 본 연구에서는 퇴원환자의 퇴원요약데이터베이스를 이용하여 질적 문제 중 병원감염발생의 변동을 SPC 기법을 활용하여 찾아낸 후, 데이터마이닝 기법을 활용하여 병원감염발생규칙을 도출해 냄으로써, 병원감염관리 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 병원감염발생의 규칙을 추출하고, 추출된 규칙에 대한 타당성을 검증한다.

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

# Ⅱ. 이론적 배경

## 1. Continuous Quality Improvement의 개념과 현황

### 가. Continuous Quality Improvement의 개념

CQI란 미국병원들이 당면한 문제를 극복하기 위하여 일반 산업에서 사용되던 질관리 프로그램인 TQM을 병원산업에 도입한 것으로, 고객중심의 사고 속에 과학적인 방법과 직원들의 참여를 바탕으로 업무과정의 개선을 통하여 지속적으로 의료의 질을 향상시키려는 노력을 말한다.

### 나. 질 향상 활동의 특성에 따른 분류

질 향상 활동을 그 특성에 따라 분류하면 질관리(Quality Control; QC), 적정진료관리(Quality Assurance; QA), 지속적인 질관리(Continuous Quality Improvement; CQI)로 구분할 수 있다. QC는 병원에서의 의료의 질과 관계가 있다고 생각되는 구조적인 요인(시설, 장비, 인력 등)을 중심으로 접근하는 것으로, 미국 병원들이 구조적인 면에서 표준화되지 못하였던 시기에 주로 사용된 방법이었으며 초창기 병원 신임제도의 주된 내용이었다.

QC는 의료의 질 향상 활동의 가장 기초적인 접근방법으로 사용하기 용이하다는 장점이 있으나, 구조적인 요인과 의료의 질과 직접적인 관계가

약하다는 점이 단점으로 지적되어 왔다.

QA는 임상평가 지침을 이용하여 병원에서 수행하고 있는 진료의 적절성을 평가하는 데 초점을 두어 주로 임상적인 활동을 평가의 대상으로 한다. 의료의 최종산물(outcome)보다는 과정을 중심으로 평가하여 기준에 미달하는 진료를 찾아내어 이를 시정하는 데 초점을 두는 방법이다.

CQI는 QA와는 달리 임상적인 활동만이 아니라, 병원 내 모든 서비스를 대상으로 하며, 잘못되거나 기준에 미달한 것을 찾아내기보다는 병원서비스의 전체적인 질을 향상시키는 데 초점을 두며, 사후감시보다는 사전예방 활동에 치중하는 특성이 있다.

### 다. 미국에서의 의료의 질 향상 활동의 발전

미국에서 의료기관들이 의료의 질에 관심을 갖기 시작한 것은 20세기 초에 각종 의료기관들이 설립되면서 의료기관의 시설과 진료수준에 차이가 많아 양질의 의료를 국민들에게 제공하지 못하게 됨에 따라 의료계의 자율적인 노력으로 시작되었다. 의료의 질과 관련된 활동이 공식적으로 처음 시작된 것은 미국외과학회(American College of Surgeons)가 기준을 설정하여 병원에 대한 신임제도를 실시한 것이라 할 수 있다.

### 라. 우리나라에서의 의료의 질 향상 활동의 발전

우리나라 병원에서 의료의 질과 관련된 논의가 실제로 시작된 것은 1981년 대한병원협회가 표준화심사를 실시하면서부터라고 할 수 있다. 물

론 그 이전에도 학계에서 의료의 질에 대한 내용이 다루어지기는 하였으나, 병원계에서 QA라는 개념이 논의되기 시작한 것은 표준화심사 내용에 적정진료관리(QA)와 의료이용도조사(Utilization Review)가 포함되어 있었기 때문이다. 표준화심사는 처음에는 미국의 JCAH(Joint Comission on Accreditation of Hospitals)가 사용하던 조사표를 그대로 번역하여 사용하였다. 그러나 조사표가 우리나라 병원 실정에 맞지 않아 심사를 하는 사람이나 심사를 받는 병원들이 혼란을 겪게 되자, 1987년 심사조사표를 우리 실정에 맞게 수정, 보완하는 작업이 있었다. 그 후에도 적정진료관리와 관련된 조사표에 많은 문제들이 지적되어 1994년 한국의료QA학회에서 적정진료관리와 관련된 조사표를 다시 보완하여 현재 사용하고 있다.

한편 정부에서는 문민정부시절 의료보장개혁위원회의 건의에 따라 병원서비스의 개선을 위하여 정부 주도로 의료기관에 대한 서비스 평가를 1995년부터 실시하였다. 의료기관 서비스평가는 표준화심사와 비슷한 내용으로 실시되었으나, 의료기관의 구조적인 특성보다는 진료과정에 초점을 둔 것이 특징이라 하겠다. 그러나, 한 의료기관을 대상으로 정부와 병원협회에서 유사한 질 평가를 하는 것에 대해서는 비판이 있는 실정이다.

병원 단위에서 의료의 질 향상 활동은 적정진료관리활동이 주였으며, 이는 표준화 심사와 서비스평가를 받기 위한 활동이었다. 병원이 의료의 질 향상 활동의 필요성을 인지하고 자발적으로 CQI 등의 의료의 질 향상 활동을 실시하기 시작한 것은 비교적 최근의 일이다. 병원간의 경쟁이 심해지고 병원의 경영수지가 악화되면서 병원들이 고객 즉 의료이용자의 중요성을 인식하게 되었고, 고객만족을 통한 경쟁력을 확보하기 위하여 의료의 질 향상 활동에 관심을 갖기 시작하였다. 따라서 우리나라 병원에서

자발적으로 의료의 질 향상 활동이 시작된 것은 그 역사가 매우 일천하다고 할 수 있다.

### 마. Continuous Quality Improvement의 발전

CQI는 의료기관에서 고객의 욕구를 충족시킬 수 있는 양질의 의료서비스를 제공하기 위하여 지속적으로 개선을 이루어 나가는 체계적인 과정이라고 할 수 있다. CQI는 원래 1960년대 일본의 기업들이 제품의 질을 향상시키고 불량률을 낮추기 위하여 품질향상조(Quality Circle)를 구성하여 통계적인 기법을 바탕으로 전사적으로 시행하던 품질향상 운동에 기초를 두고 있다. 이 당시 Juran, Deming과 같은 이 분야의 전문가들이 일본기업에서 TQM 활동을 지원하면서 TQM이 기업 내 품질향상운동으로 자리를 잡게 되었다. TQM 활동의 결과, 일본기업들이 생산하는 제품들이 점차 세계시장에서 품질을 인정받게 되었으며, 특히 1980년대에 들어와서 '오일쇼크' 이후 미국시장에서 경쟁력을 갖게 되었다. 일본제품들의 도전으로 경쟁력을 잃어가던 미국기업들이 일본기업을 벤치마킹(bench marking)하면서 TQM에 관심을 갖기 시작하였고, 1980년대에 Ford를 포함한 미국 대기업들이 TQM을 본격적으로 도입하면서 미국 내에서도 TQM이 기업에 널리 퍼지게 되었다.

한편 의료기관에서는 1980년 중반부터 TQM에 관심을 갖기 시작하였다. 미국에서는 Harvard Community Health Center, Hospital Corporation of America, 유타주의 Intermountain Health System 등이 의료분야에 있어 TQM의 선두적인 역할을 담당하였다. 한 조사에 의하면 미국 내 병원 중

60% 정도가 CQI/ TQM활동을 본격적으로 실시하고 있으며, 실시하고 있지 않는 병원의 85%도 1-2년 이내에 CQI프로그램을 실시할 계획이라고 한다. 우리나라에서는 1990년대 중반부터 TQM/ CQI에 대한 논의가 시작되어 현재 몇몇 병원에서 CQI프로그램을 운영하고 있으나, 아직은 초창기라 할 수 있다(조우현, 1999).

현재 우리나라의 질 향상 활동으로는 1981년 병원표준화심사제도의 도입, 전문직능단체의 보수교육 및 병원 내 활동으로서 Case Review, Mortality Conference, Drug Review 등의 전통적인 질관리 활동에서, 1994년 한국의료 QA학회의 창립과 함께 질 향상 활동이 전개되었다. 활동사례로는 질적 수준에 대한 평가작업, 의료기관내 재원사용의 합리성을 도모하려는 이용도조사, 환자만족도 조사가 주된 내용을 이루고 있다(이선희, 1999).

## 2. 데이터마이닝의 개념

### 가. 데이터마이닝의 출현 배경

정보기술의 발달로 인간의 데이터 수집능력은 무한히 증가하였다. 또한 수많은 데이터들을 저장하고, 필요한 데이터를 찾으려는 노력과 그들 데이터로부터 매우 가치있는 정보나 지식을 추출하고자 하는 욕구가 증대하고 있다. 과거에는 이러한 창의적인 작업은 전적으로 인간들의 몫인 것으로만 생각되어 왔으나, 오늘날 정보통신기술의 발전과 추론, 인지 등 학문의 진전은 이러한 작업의 많은 부분이 컴퓨터에 의해 수행될 수 있게 되었다. 그러나 이러한 정보들은 대개 방대한 데이터 더미에 파묻혀 있어 기존에 사용하던 DBMS(Database Management System) 혹은 데이터베이스 기법으로는 좀처럼 발견하기 어려웠다. 따라서 이러한 요구에 대한 솔루션으로, 데이터로부터 무엇인가를 발굴한다는 뜻에서 "데이터마이닝(Data Mining)" 개념이 발생하였다(정현, 1999).

### 나. Knowledge Discovery in Database와 데이터마이닝

'데이터마이닝'과 'KDD(Knowledge Discovery in Database)'란 용어의 정확한 의미에 대해서 많은 학자들이 비슷한 말로 간주하는 데 대해 약간의 의견차이가 있어 왔다. 1995년의 몬트리얼의 제1회 국제 KDD학술대회에서 'KDD'는 데이터로부터 지식을 추출하는 전 과정을 설명하는 뜻으로 해석되었다.

이때 지식이란 데이터 요소들 간의 연관성이나 패턴을 의미한다. 그리고 '데이터마이닝'은 KDD과정에서 탐사 단계만을 뜻하는 의미로 사용한다고 제안되었다.

KDD에 대한 공식적인 정의는 "데이터로부터 암시적이며 이전에 알려지지 않은 잠재적으로 유용한 지식을 추출하는 것"이다. 그러므로 지식은 새롭고 쉽게 알 수 없으며, 인간이 활용할 수 있어야만 한다. KDD는 새로운 기술이기보다는 기계-학습, 통계학, 데이터베이스 기술, 전문가 시스템 그리고 데이터 가시화 등의 분야가 모두 기여를 한 학제적인 연구분야에 속한다(그림 1)(용환승, 1998).

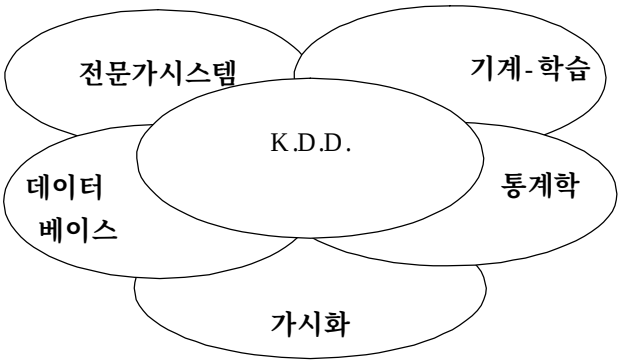


그림 1. 학제간 연구 분야인 데이터 마이닝

### 다. 데이터마이닝의 관련분야와 용어

1) KDD(Knowledge Discovery in Database)

데이터베이스 안에서의 지식(데이터간의 연관성이나 패턴)발견이라는 표현은 일반적으로 데이터마이닝과 가장 유사한 의미로 사용되고 있다. 그러나 보다 정확하게 표현한다면, KDD는 지식을 추출하는 전 과정을 의미하고, 데이터마이닝은 온라인분석처리(OLAP: On-Line Analytical Processing)이나 데이터웨어하우징(Data Warehousing)등과 마찬가지로 전체적인 KDD과정 중 한 과정인 탐사 단계를 의미한다고 할 수 있다.

2) 기계학습(Machine Learning)

인공지능(AI)의 한 분야로서 자동적인 학습기법을 설계하고 구현하는 분야이다.

3) 패턴인식(Pattern Recognition)

공학에서 출발하였으며, 이미지 분류와 깊은 관련을 가지고 있다. 패턴인식은 데이터 베이스에서 유용한 패턴을 찾아내는 다양한 기법들을 제공한다.

4) 통계학(Statistics)

데이터마이닝의 대부분은 통계학의 한 분야라고 할 수 있다. 한 가지 예를 들어본다면, 데이터마이닝의 모델링 부문에서 가장 많이 사용되는 기법 중의 하나인 판별분석(discriminant analysis)은 1936년에 R. A. Fisher에

의해 시작된 다변량 통계분석의 한 분야라고 할 수 있다.

5) 뉴로컴퓨팅(Neurocomputing)

신경망 등과 관련된 다양한 학문적 배경을 가진 분야이다(강현철 등, 1999).

### 라. 데이터마이닝 접근방법론

1) 샘플링(Sampling)

데이터마이닝은 대체로 수십메가에서 수십기가에 이르는 대용량의 데이터를 기반으로 한다. 그러나 방대한 양의 데이터를 살피는 것은 시간의 측면에서만 보아도 많은 인내를 요하게되는 작업이 될 수 있다. 이때 고려해야 하는 과정이 바로 샘플링이다. 샘플링이란 방대한 양의 데이터(모집단)에서부터 모집단을 닮은 작은 양의 데이터(샘플: 표본)를 추출하는 것이다. 이는 데이터마이닝 작업을 하는 데 있어서 시간과 비용을 절감하여 효율적인 작업진행을 돕고 이를 통하여 데이터 마이닝 작업의 무게를 조금 가볍게 할 수 있는 절차이다. 그러므로 샘플링이 적절한가에 대한 질문을 던져보고 필요하다면 적절한 방법을 이용하여 샘플데이터를 생성한다. 통계학에서는 이미 샘플링을 하는 여러 가지 방법론들이 개발되었는데, 그 중 가장 널리 알려진 방법이 단순임의추출법(Simple Random Sampling)으로, 이는 간단하게 설명하면 모집단을 골고루 섞어 놓은 후 원하는 양만큼을 무작위로 추출해내는 것이다. 이 외에도 층화추출법(Stratified Sampling), 계통추출법(Systemic Sampling) 등이 자주 거론되는 방법으로

각 상황에 맞는 샘플링 방법을 선택하여 사용한다.

샘플링과정에서 가장 주의할 점은 모집단을 닮지 않은, 한쪽으로 치우친 샘플이 추출되는 경우로써, 이는 샘플링 후 기본적인 몇 가지의 자료탐색을 거쳐 반드시 확인하여야 한다.

2) 탐색(Exploration)

데이터의 탐색과정에서는 이미 알고 있는 사실들을 수치화하는 작업을 시작으로 하여 보유하고 있는 수많은 변수들의 관계를 살펴보는 단계이다. 예를 들어 '백화점의 A상품구매 고객들은 주로 특정 요일에 집중되어 있다' 거나, 'B상품의 구매고객은 주로 여자다' 등의 모르고 있던 정보 또는 기존에 현업 관계자의 느낌으로만 알고 있던 사실들을 확인할 수 있게 되는 경우도 있고, 또 가끔은 현업에 있는 사람들이 확신하고 있던 사실들이 틀렸다는 것도 확인할 수 있다. 이처럼 탐색단계는 데이터의 모양을 면밀히 둘러보아 정보화할 수 있는 기반을 잡아가는 과정이다. 이 단계에는 실제 한 모집단으로 알고 있던 것이 두 개 이상으로 나뉠 수도 있고, 수십개 이상의 많은 변수를 실제 중요한 정보를 주는 소수의 변수로 축소시킬 수 있으며, 기존에 있던 변수들 외에 새로운 변수를 생성해야 할 필요성을 느끼게 될 수도 있다.

3) 변형 및 조정(Modification)

한 신용카드회사의 거래정보를 상상해보자. 그 거래정보에는 각 고객의 지난 달 카드사용내역을 가지고 있을 것이다. 그러나 분석자의 요구 또는 현업 담당자의 요구에 의해 꼭 필요한 정보 - 각 고객의 지난 달 1회

평균 카드사용금액 - 는 보유하고 있지 않을 수 있다. 또한 각 고객의 연령정보는 있지만 너무 다양하여 10대, 20대, 30대 등으로 크게 나누어야 할 수도 있다. 이러한 정보는 기존의 변수를 이용하여 새로이 생성해야 할 변수가 된다. 이러한 작업들을 고려하는 단계가 바로 변형 및 조정단계이다. 이 단계에서 생성 또는 수정된 변수는 차후 모형화 단계에서 아주 중요한 정보로써 활용될 수 있을 것이다.

4) 모형화(Modeling)

데이터마이닝 과정에서 가장 중요한 단계로서, 앞서 선행되었던 단계에서 선정된 주요한 변수를 사용하여 다양한 모형 - Neural Networks, CHAID, CART, 일반화선형모형 등의 전통적인 통계적 모형 등을 적합해 보는 단계이다. '신용카드의 도용방지모형은 Neural Network가 가장 적합하다' 등의 알려져 있는 모형이 있기는 하나, 그렇지 않는 경우에는 다양한 모형에 적합한 후 예측력이 가장 뛰어난 모형을 선택하는 것이 일반적이다.

5) 평가(Assessment)

적합한 두 개이상의 모형의 효과를 비교하여 가장 좋은 모형을 선택할 수 있다. 이렇게 선택된 모형은 실제 모집단에 반영하여 그 효과를 재평가할 수 있으며, 이로 인하여 새로운 문제를 제기하여 이전 작업을 다시 반복할 수도 있을 것이다(Internetsite[http:/ / www.cio.com/ archive/ 100196_ inws_content.html]).

## 마. 데이터마이닝 기법

### 1) 의사결정나무분석

의사결정나무분석(decision tree analysis)은 의사결정규칙을 나무구조로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석방법으로, 예측 과정이 나무구조에 의한 추론규칙(induction rule)에 의해 표현된다(Lee GH 등, 1998). 데이터마이닝에서의 의사결정나무는 탐색(exploration)과 모형화(modeling)의 특성을 지니며, 사전에 이상치(outlier)를 검색하거나 분석에 필요한 변수를 찾아내고 분석모형에 포함되어야 할 교호효과를 찾아내는 데 사용될 수 있고, 그 자체가 분류 또는 예측 모형으로 사용될 수 있다. 의사결정나무분석은 하나의 나무구조를 이루고 있으며, 마디(node)라고 불리는 구성요소들로 이루어져 있고, 목표변수(target variable)인 뿌리마디(root node)로부터 시작하여 예측변수(predictor variable)에 의해 각각 가지(branch)가 끝마디(terminal node)에 이를 때까지 자식마디(child node)를 계속적으로 형성해 나감으로써 완성된다. 의사결정나무 형성을 위한 다양한 기준들이 제안되어 있는데, 이는 하나의 부모마디로부터 자식마디들이 형성될 때 예측변수의 선택과 병합이 이루어지는 기준인 분리기준(splitting criterion), 정지규칙(stopping rule), 가지치기(prunning)방법으로서 이들을 어떻게 결합하느냐에 따라 서로 다른 의사결정나무형성방법들이 만들어진다(Choi JH 등, 1998). 1980년대 이후 CHAID, CART 등 다양한 알고리즘이 제안되어 왔다.

### 2) 연관성 규칙

임의의 단위에 발생한 사건들의 묶음을 트랜잭션이라 하고, 대용량의 트랜잭션들이 데이터베이스에 누적된 환경에서 사건부류 혹은 트랙잭션간의 상호관계를 발견하는 작업을 데이터베이스상의 연관성규칙 탐사로 정의할 수 있다(Lee DH, 1998). 연관성규칙 탐사는 관계형 데이터베이스에서 함수적 종속성(functional dependency)을 추출하는 기법과는 달리 통계적 방법에 의해 연관성이 있는 항목들 사이의 규칙성을 추출하는 과정이다. 연관성규칙(association rules)에는 사용자에 의해 주어지는 임계치인 최소지지도(minimum support)와 최소신뢰도(minimum confidence)가 적용된다(Kim SM, 1998).

지지도는 특정항목집합의 통계적 중요성을 나타내는 수치 값으로, 예를 들면, "두 개의 항목이 동시에 일어날 확률'을 의미한다. 신뢰도는 연관성규칙의 강도(strength)를 나타내는 척도로서, 예를 들면, '품목A가 구매되었을 때 품목B가 추가로 구매될 확률인 조건부 확률'이다(호승희, 1999).

### 바. 데이터마이닝의 주요 활용분야

데이터마이닝은 통계적인 기법을 사용하지 않으면서도 데이터의 패턴을 발견해 주는 기법으로서 예측을 해준다는 데 가장 큰 장점이 있다(Lawrence, 1997). 데이터마이닝을 이용하여 분석할 수 있는 병원의 주요 정보는 다음과 같다(Internetsite[http:/ / www.cio.com/ archive/ 100196_inws _content.html]).

#### 1) 환자마케팅

기존 환자군의 특성을 이용하여 어떤 환자들의 재진율이 높고, 그 원인이 무엇인지를 알려주는 등 환자의 의료기관 이용특성을 알려주므로 이를 근거로 환자에 대한 마케팅을 실시하여 병원의 환자수를 증가시킬 수 있으며, 더 나아가 병원의 수익성을 높일 수 있는 방안을 마련하는 데 도움을 줄 수 있다.

2) 의료시설배치

데이터마이닝을 이용하여 환자의 동선을 파악할 수 있으므로, 이에 근거하여 의료시설을 재배치하면 환자들의 동선을 줄이면서 병원의 혼잡도를 완화할 수 있다.

3) 미수금관리

미수금 발생확률이 높은 환자를 기존자료에 근거하여 입원시부터 찾아내 주므로 입원시부터 미수금이 발생하지 않도록 중점관리할 수 있도록 하여 병원의 미수금 발생을 줄일 수 있다.

4) 진료행위 프로토콜 개발

의사들의 진료행위와 치료결과에 근거하여 가장 효율적인 치료패턴을 찾아주므로, 이에 근거하여 진료행위 프로토콜을 만들 수 있고, 또한 의료의 질 향상을 도모할 수 있게 한다.

5) 진료비 삭감 예방

기본의 진료비 삭감내역을 분석하여 삭감의 원인을 찾아주므로 이에

근거하여 진료비 삭감을 막을 수 있는 방안을 찾아 줄 수 있어서 병원의 진료비 삭감에 의한 손실을 방지할 수 있다.

6) 진단

검사결과와 각종 진단자료에 근거하여 의사들의 정확한 진단을 도와 줄 수 있다.

7) 재원일수 및 진료비 예측

기존 환자들의 재원일수 및 진료비 자료에 근거하여 새로운 환자가 입원할 경우에 이들의 재원일수와 진료비를 예측하여 줄 수 있다.

8) 자원사용량 예측

병원에서 필요로 하는 각종 의료자원을 예측하여 주므로 병원의 효율적인 물자관리에 기여할 수 있다.

9) 추정예산

각종 자료에 근거하여 병원에서 필요로 하는 예산을 추정하여 줄 수 있다.

10) 의학연구

## 3. Statistical Process Control의 개념

### 가. Statistical Process Control의 정의

Statistical Process Control(SPC, 통계적 공정관리)이란 공정에서 요구되는 품질이나 생산성 목표를 달성하기 위하여 통계적 방법으로 공정을 효율적으로 운영해 나가는 방법을 말한다. SPC는 세 가지 단어의 합성어이며, 'S(Statistical)'는 통계적 자료와 분석기법의 도움을 받아서, ‘P(Process)'는 공정의 품질변동을 주는 원인과 공정의 능력상태를 파악하여, ’C(Control)‘는 주어진 품질목표가 달성될 수 있도록 PDCA사이클을 돌려가며 끊임없는 품질개선이 이루어지도록 관리해 나가는 활동이다. PDCA란 관리의 네 가지 구성요소로서 계획(Plan), 실시(Do), 검토(Check), 조치(Act)를 의미한다.

### 나. Statistical Process Control과 Total Quality Management

우리나라에서의 품질관리활동은 1961년에 공업표준화법이 제정되고, 1967년에 품질관리법이 공포되면서 시작되었으며, 이 QC활동은 그동안 품질향상과 수출증대에 많은 기여를 해왔으나, 아직도 QC 체제의 보급과 확산이 미흡하고, 이를 도입한 기업들도 상당수가 기대한 만큼의 성과를 거두지 못하고 있는 것도 현실이다. 이것은 그 동안 우리 기업의 QC활동이 생산현장 및 제품 중심으로 추진되어 온 데다, 90년대에 들어오면서 까다로운 고객요구, 국내시장개방, 고임금 등 경영환경의 변화에 적응하기 위한

경영전략수단으로서의 종래의 QC활동이 한계를 가지고 있기 때문이다.

따라서 최근 지금까지의 현장위주, 제품위주의 QC체제에서부터 한 차원 더 나아가 제품의 기획, 설계로부터 사후 서비스(after-service)에 이르기까지 전 단계에 걸친 경영전략적 차원에서의 전사적 품질경영(TQM: Total Quality Management)이 필요하게 되었다.

TQM은 최고경영자의 리더십 아래 품질방침(QP: Quality Policy), 품질관리(QC: Quality Control), 품질향상(QI: Quality Improvement), 품질보증(QA: Quality Assurance) 등과 같은 수단에 의해 기업의 모든 부서에서 전사적으로 실행하는 종합적인 경영관리체계이다.

TQM을 수행하는 데는 각종의 과학적 관리기법이 사용되어야 그 운영의 효율성을 증대시킬 수 있다. TQM과 관련된 대표적인 과학적 관리기법으로 SPC를 들 수 있다(박성현 등, 1998).

### 다. Statistical Process Control 기법

SPC는 의료서비스에서 요구되는 의료의 질이나 비용 효과적인 의료제공의 목표를 달성하기 위한 통계적 방법으로, 주요 기법으로서 특성요인도, 히스토그램, 파레토도표, 산점도, 관리도, 층별 연관도, 매트릭스 도법, 검정추정, 회귀분석, 주성분 분석, 군집분석, 체크시트, 친화도법, 계통도, 요인배치실험법, 판별분석, 인자분석 등이 있다.

관리도는 1924년 벨 전화연구소의 W. A. S. Hewhart에 의해 처음으로 소개되었으며, 관리상태에서의 품질 특성치의 평균을 나타내는 중심선과 중심선 상하에 한 쌍의 관리한계선(control limits), 그리고 관리도는 공정의

상태를 나타내는 품질 특성치로 구성된 그래프를 말한다. 관리도는 중앙선의 값을 평균을 사용하며 신뢰구간의 개념에 따라 sigma의 범위를 정한다(박성현, 1998). 또한 종속변수의 특성과 부분군의 크기에 따라 관리도의 유형이 달라진다.

### 라. 보건의료분야에서의 Statistical Process Control

보건의료분야에서의 SPC는 의료서비스에서 요구되는 의료의 질이나 비용 효과적인 의료제공의 목표를 효율적으로 운영해 나가는 방법을 말한다. 그 도입배경은 다음과 같다.

1) 보건의료분야의 전산화로 데이터의 축적

처방전달시스템, 전자의무기록 등으로 병원데이터가 전산에 축적됨에 따라 챠트를 이용하는 질관리 방법보다 컴퓨터를 이용하는 질관리 방법이 보다 효율적이며, 이는 컴퓨터를 이용하여 질관리를 해줄 수 있는 각종 Tool이 개발되고 있기 때문이다.

2) Managed Care 시대의 도래

세계각국이 의료비를 억제하기 위해 정부에서 의료행위에 대한 각종 제한을 가하고, 특히 진료비 지불체계가 행위별 수가제에서 DRG(Diagnosis Related Groups)지불제도 등으로 변함에 따라 의료기관은 비용을 최소화하면서 양질의 의료를 제공해야 하고, 그에 따라 진료행위 프로토콜 등을 개발하여 의료서비스도 표준화된 공정에 따라 제공해야 할

필요성이 증대됨에 따라 산업체에서의 공정관리기법을 도입할 필요성이 있다(Jeffrey, 1998).

3) 의료의 질 향상의 필요성

병원간의 경쟁이 치열해지고 또한 소비자의 양질의 의료에 대한 요구가 강해지고, 환자중심의 의료가 필요해 짐에 따라 의료의 질 향상이 필요하게 되었고, 환자만족도에 관한 데이터를 수집하여 분석할 필요성이 있으며 통계적 공정관리 기법의 의료분야에 대한 도입의 필요성이 있다.

4) 팀 중심의 보건의료

이제 의료행위는 의사 1인이 하는 것이 아니라, 팀 중심의 의료가 이루어져야 하며, 이를 위해서는 의료행위의 모든 단계에서 양질의 의료가 제공되도록 해야 함에 따라 통계적 공정관리의 필요성이 증대된다.

### 마. 보건의료분야에서의 Statistical Process Control 적용분야

1) 치료결과(Clinical Outcomes)

SPC기법을 치료결과에 적용하는 것은 치료결과의 기대수준과 진료행위 프로토콜의 설정, 의사들간의 치료형태의 차이나 진료행위프로토콜과 실제 진료행위와의 차이의 확인, 환자진료에 필요한 적정자원을 사용할 수 있도록 하는 등에 그 목적이 있다.

측정지표의 예로서는 사망률, 감염율, 사고건수, 약물에 대한 부작용, 제왕절개율 등이 있다.

2) 진료과정(Clinical Process)

SPC기법을 진료과정에 적용하는 것은 장비, 정책, 절차 등에 기인한 의료의 질 변동을 밝히며, 의료의 질 변동에 제대로 대응하면서 동일 수준의 의료서비스의 질을 유지할 수 있도록 모니터할 수 있도록 하는 데에 그 목적이 있다.

측정지표의 예로서는 응급실에서의 대기시간, 검사의 에러, 검사결과의 정확성, 불필요한 수술건수 등이 있다.

3) 병원행정서비스 과정(Service Processes)

SPC기법을 병원행정과정에 적용하는 것은 병원행정에 기인한 변동을 확인하거나 개선을 시도한 후의 의료의 질향상을 모니터할 수 있도록 하는 데에 그 목적이 있다.

측정지표의 예로서는 입원시 대기시간, 진료비계산의 정확도, 환자의 불만건수, 예약부도율, 인공수정의 실패율, 검사사고 등이 있다.

4) 환자진료결과(Patient Outcomes)

SPC기법을 환자진료결과에 적용하는 것은 약물투여에 대한 반응, 기대치와 실제치의 변동 등을 파악하는 데 그 목적이 있다.

측정지표의 예로서는 혈압, 심박동, 혈당치, I/O, 체중, 체온, 혈소판수 등이 있다(대한의무기록협회, 1999).

# III. 연구 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 343건으로서 가장 높은 비율인 36%를 차지했다. 본 연구에서는 병원감염관리방안을 찾기 위해 병원감염 환자를 연구대상으로 선정하였다. 병원감염발생 환자 중 수술받은 242명과 단순임의추출로서 수술받은 병원감염 미발생환자 1,281명을 선정하여 총 1,523명을 연구대상으로 하였다.

## 2. 연구 모형

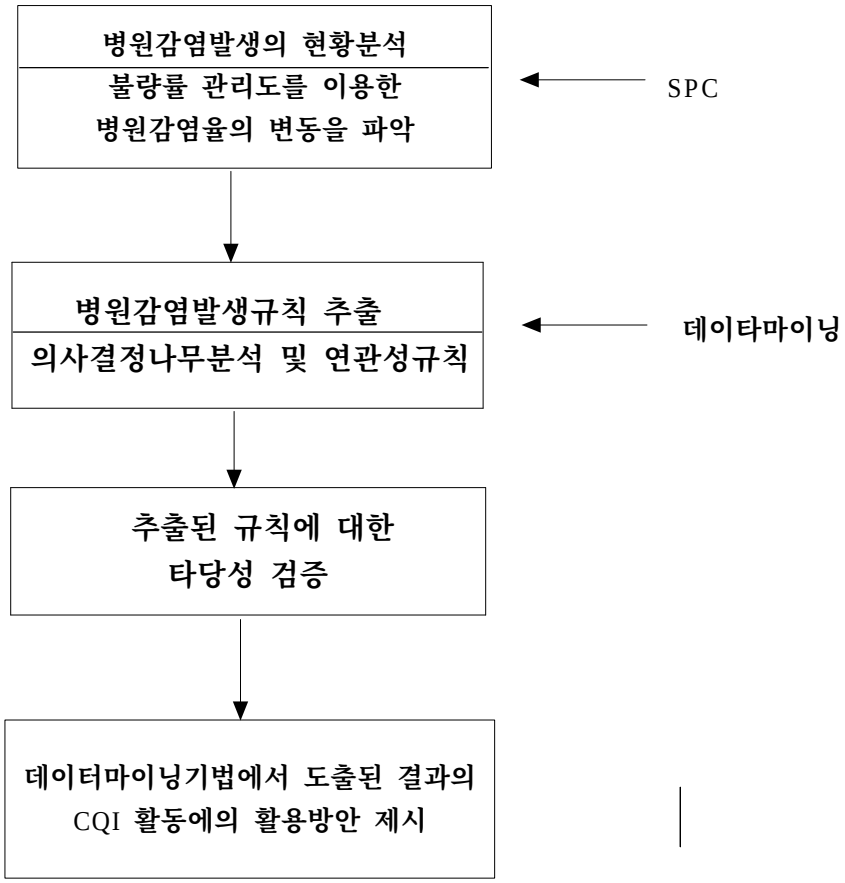


그림 2. 연구분석의 틀

이 연구는 다음의 5단계로 이루어졌다(그림 2).

첫째, CQI를 위한 퇴원환자의 퇴원요약 DB 구축단계로서 퇴원환자의 퇴원요약 DB를 텍스트로 다운받아서 Access를 이용하여 PC에서 DB를 구축한 후 CQI 분석에 적합하도록 필요한 항목을 추출하고 변형했다. 둘째, SPC 기법을 활용하여 병원감염의 현황을 분석하는 단계로서 병원감염발생률을 중심으로 변동 및 추이를 분석했다. 셋째, 데이터마이닝 기법을 활용하여 병원감염발생의 규칙을 찾아냈다. 넷째, 데이터마이닝 기법을 통해 추출된 규칙에 대한 타당성을 검증했다. 다섯째, 데이터마닝 기법에서 추출된 규칙의 CQI 활동에의 활용방안을 제시했다.

## 3. 분석항목의 선정

### 가. Continuous Quality Improvement를 위한 퇴원요약 데이터베이스의 구성

퇴원요약 DB는 일일 퇴원분석지를 통하여 퇴원환자 개개인에 대한 진료내용 또는 해당 코드 및 통계사항이 포함되어 있다. 미국의 경우 UHDDS(Uniform Hospital Discharge Data Set)를 이용하고 있으며, 이것은 acute care 환자에 대한 significant minimum data sets중의 하나로, 1974년 미연방정부에 의해 지정된 Medicare, Medicaid 환자의 퇴원정보에 반드시 포함되어 있어야 하는 최소한의 항목들의 묶음이다. 1984년, 1986년에 걸쳐 두 차례 개정되어 현재까지 사용되고 있으며, 사용범위도 확장되어 미국내의 개인병원에서도 쓰이고 있다. 우리나라에서도 대부분의 병

원에서 일일 퇴원분석지의 내용을 전산에 입력하여 많은 정보를 얻고 있으며 진료통계자료 및 의료보험 청구용 자료로 사용될 뿐 아니라, 의료의 질 개선을 위한 기초자료가 되고 있다(이정화, 1996). 본 연구에서 사용할 변수인 질적 문제(Quality Problem) 중 병원감염 등 의료의 질 개선 관련 사항도 퇴원요약 DB에 포함되어 있다.

### 나. 병원감염발생 관련 변수의 선정

본 연구의 데이터는 1996년부터 1999년까지 퇴원한 환자의 퇴원요약 DB로서 환자의 입원에서부터 퇴원할 때까지 발생된 의료서비스의 내용들을 항목별로 입력함으로써 구축되었고, 총 110개의 필드로 이루어져 있으며, 그 중 중복 필드(field)를 제외하면 모두 78개의 필드로 이루어져 있다.

데이터마이닝 기법을 이용한 병원감염발생 규칙 발굴에 사용된 변수들은 표 1과 같다.

표 1. 데이터마이닝기법에서 사용된 변수

| 변수명 | 내용 |
|---|---|
| 종속변수 | 병원감염발생 유무 (1:유, 0:무) |
| 독립변수 | 성별 (1:남, 2:여) |
| | 연령 (1:-20, 2:21-40, 3:41-60, 4:61- ) |
| | 주진단명 |
| | 동반질환 유무 (1:유, 2:무) |
| | 주수술명 |
| | 수술시간 (1:-60분, 2:61-120분, 3:121-180분, 4:181- ) |
| | 주호소 |

퇴원요약 DB 중 질적 문제의 해당항목은 임상 각과에 공통으로 적용되는 평가기준으로서 그 종류는 다음과 같으며 이 중 병원감염은 ⑧번 항목에 속한다(유승흠, 1993).

① 전번에 입원의 부족한 치료 혹은 합병증으로 인한 재입원
(Readmission for complication/ incomplete management on previous hospitalization)

② 외래 치료의 부작용 또는 합병증으로 인한 입원
(Admission for adverse results of outpatient management)

③ Precedure 중 organ이 perforation,laceration, tear 혹은 그외의 injury를 받은 경우
(Perforation,laceration, tearor injury of an organ incurred)

④ **계획없이 수술방 혹은 분만실로 되돌아 온 경우**

(Unplanned return to operating or delivery room on this admission)

⑤ **약물 혹은 수혈에 대한 부작용, 부적당한 사용**

(Drug/ transfusion reaction,complication or improper utilization)

⑥ **수술/ 분만 중 혹은 후에 과도한 출혈이 있는 경우**

(Excessive blood loss at / after delivery or surgery)

⑦ **수술중 혹은 수술후 48시간이내 심정지 혹은 심근경색이 있는 경우**

(Cardiac arrest or myocardiac infarction during or within 48 hours of a surgery)

⑧ **병원 감염** (Nosocomial infection)

⑨ **심정지 혹은 호흡정지**

(Cardiac or respiratory arrest except death)

⑩ **환자사고** (Hospital- incurred patient incident)

⑪ **사망** (Unexpected death)

⑫ **기타**

### 다. 분석 방법

Access를 이용하여 구축된 분석용 파일을 다음과 같이 분석했다.

첫째, 기술통계를 작성하여 연구대상자의 일반적 특성파악을 위해 SAS의 통계팩키지를 사용하였다.

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

셋째, 병원감염발생의 규칙을 찾기 위해 SAS의 Enterprise Miner를 사용하였다.

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

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

## 5. 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 유용한 방법인 CHAID 방법을 적용하였다(최종후 등, 1998).

## 6. 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 =⇒ 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 =⇒ 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).

# Ⅳ. 연구 결과

## 1. 연구대상자의 일반적 특성

연구대상 88,593명 중 남자가 50.7%, 여자가 49.2%로 남자가 많았으며, 퇴원과로는 내과계가 48.3%, 외과계가 51.7%로 외과계가 많았고, 연령층은 20세 이하가 30.3%, 21-40세가 28.4%, 41-60세가 23.6%, 61세 이상이 17.6%로 20세 이하가 가장 많았다. 주요 질환군은 손상, 중독 및 외인에 의한 특정 기타 질환군이 14.%로 가장 많았고, 호흡기계 10.9%, 신생물 10.4% 순이었다. 퇴원형태로는 퇴원지시후 95.1%로 가장 많았고, 자의퇴원 3.7%, 전원 1.2%였다. 치료결과로는 호전 91.9%로 가장 많았고, 호전안됨 1.5%, 치료않음 4.2%, 가망없는 퇴원 0.3%, 48시간이내 사망 0.5%, 48시간이후 사망 1.2%였다. 입원경로로는 외래 56.9%로 가장 많았고, 응급실 37.8%, 신생아실 5.3%였다. 28일이내 재입원여부에서 해당 37.1%, 미해당 62.9%였다. 질적 문제의 해당여부에서 해당 1%, 미해당 99%였다(표 2 ).

표 2. 연구대상자의 일반적인 특성

| 측정변수 | 측정값 | 빈도 | 백분율(%) |
|---|---|---|---|
| 성별 | 1. 남 | 44,951 | 50.7 |
| | 2. 여 | 43,632 | 49.2 |
| | 계 | 88,583 | 100.0 |
| 퇴원과 | 1. 내과계 | 42,760 | 48.3 |
| | 2. 외과계 | 45,833 | 51.7 |
| | 계 | 88,593 | 100.0 |
| 연령 | 1. 20세 이하 | 26,801 | 30.3 |
| | 2. 21-40세 | 25,163 | 28.4 |
| | 3. 41-60세 | 20,893 | 23.6 |
| | 4. 61세 이상 | 15,731 | 17.8 |
| | 계 | 88,588 | 100.0 |
| 질환군 | 1. 손상, 중독 및 외인에 의한 특정 기타결과 | 12,643 | 14.3 |
| | 2. 호흡기계질환 | 9,654 | 10.9 |
| | 3. 신생물 | 9,229 | 10.4 |
| | 4. 소화기계질환 | 8,525 | 9.6 |
| | 5. 임신, 출산 및 산욕 | 6,829 | 7.7 |
| | 6. 순환기계 질환 | 6,682 | 7.2 |
| | 7. 특정 감염성 및 기생충성 질환 | 5,829 | 6.6 |
| | 8. 비뇨생식기계의 질환 | 4,849 | 5.5 |
| | 9. 근골격계 및 결합조직의 질환 | 3,699 | 4.2 |
| | 10. 주산기에 기원한 특정병태 | 3,097 | 3.5 |
| | 11. 기타 | 17,557 | 19.8 |
| | 계 | 88,593 | 100.0 |

(계속)

| | | | |
|---|---|---|---|
| 퇴원형태 | 1. 퇴원지시후 | 88,425 | 95.1 |
| | 2. 자의퇴원 | 3,303 | 3.7 |
| | 3. 전원 | 1,023 | 1.2 |
| | 4. 사망 | 022 | 0.0 |
| | 계 | 88,593 | 100.0 |
| 치료결과 | 1. 완쾌 | 28 | 0.0 |
| | 2. 호전 | 81,458 | 91.9 |
| | 3. 호전안됨 | 1,302 | 1.5 |
| | 4. 치료않음 | 3,727 | 4.2 |
| | 5. 진단뿐 | 322 | 0.4 |
| | 6. 가망없는 퇴원 | 251 | 0.3 |
| | 7. 48hr 이내 사망 | 420 | 0.5 |
| | 8. 48hr 이후 사망 | 1,085 | 1.2 |
| | 계 | 88,593 | 100.0 |
| 입원경로 | 1. 외래 | 50,430 | 56.9 |
| | 2. 응급실 | 33,467 | 37.8 |
| | 3. 분만실 | 9 | 0.0 |
| | 4. 신생아실 | 4,682 | 5.3 |
| | 5. 기타 | 5 | 0.0 |
| | 계 | 88,593 | 100.0 |
| 28일이내 재입원 | 해당 | 32,906 | 37.1 |
| | 미해당 | 55,687 | 62.9 |
| | 계 | 88,593 | 100.0 |
| 질적 문제 | 해당 | 884 | 1.0 |
| (Quality Problem) | 미해당 | 88,709 | 99.0 |
| | 계 | 88,593 | 100.0 |

## 2. 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 3과 같이 질적 문제 중 병원감염발생률을 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 병원감염발생률의 변동이 있음을 알 수 있었으며, 본 연구에서는 병원감염발생에 대한 심층분석을 하기로 하였다(그림 3).

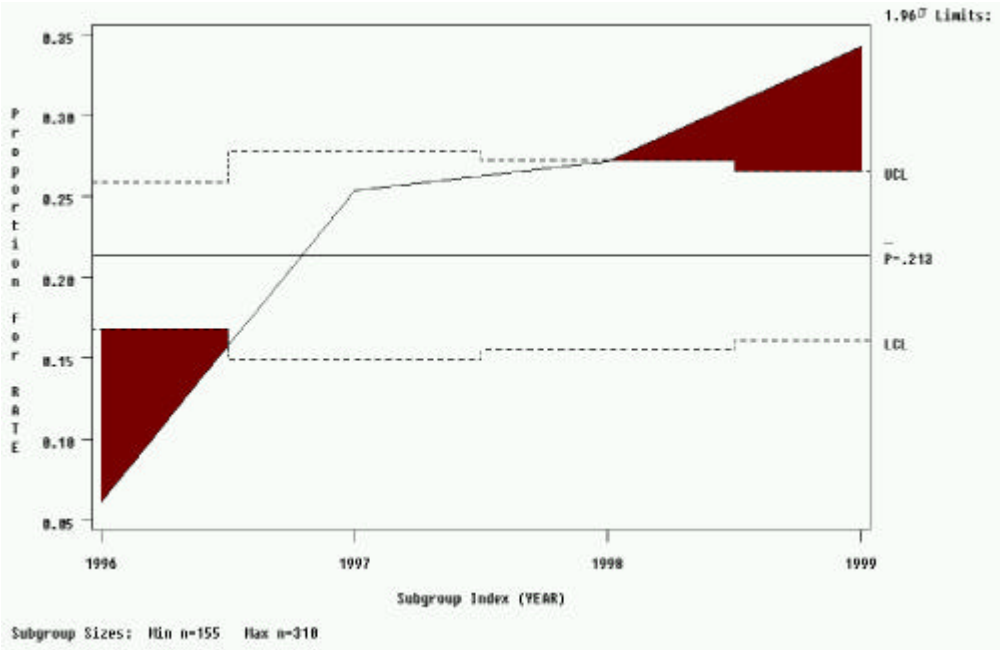


그림 3. 연도별 병원감염률의 불량률 관리도

## 3. 병원감염발생의 규칙 및 위험집단

### 가. 병원감염발생의 분류 및 예측

#### 1) 모형의 평가

병원감염발생의 분류 및 예측을 위해 데이터마이닝 기법인 의사결정나무분석을 적용하였고, 이에 대한 모형 평가를 한 결과, Root ASE가 0.27로 나타났으며, 전체 병원감염발생률 15.9% 중 의사결정모형은 병원감염이 발생할 확률이 높은 상위 10% 환자군에서 76.5 %가 병원감염이 발생할 것임을 보여주었다(표 3).

표 3. 의사결정나무 분석의 모형 평가

| Root ASE | Response rate (%) | | |
|---|---|---|---|
| | 상위 10% | 상위 20% | 상위 30% |
| 0.27 | 76.5 | 55.9 | 44.9 |

#### 2) 병원감염발생의 계층별 패턴

그림 4와 같이 의사결정나무 분석을 한 결과, 병원 감염 발생 계층은 총 24단계로 분류되었다. 병원감염을 결정짓는 가장 중요한 변수는 의사결정나무의 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 4).

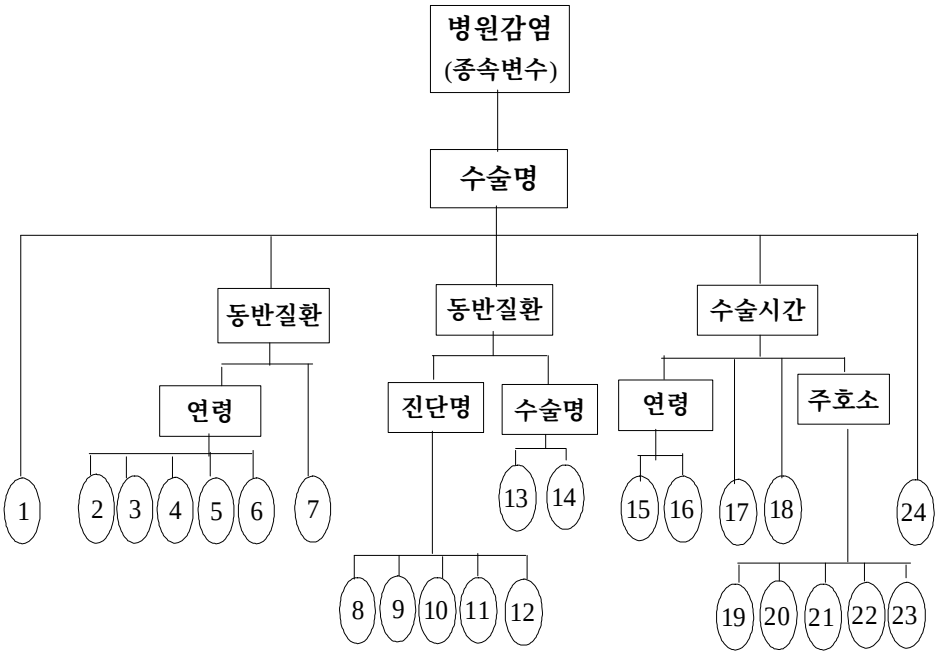


그림 4 . 병원감염 발생에 관한 의사결정도

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 췌장 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 병

원감염발생 확률이 가장 높으며, 31-40세일 때가 가장 낮았다.

⑦번 계층은 ②번 ~⑥번 계층과 동일한 수술을 받은 환자군으로서 동반 질환이 없는 경우이며 병원감염 발생확률은 14.3%로 나왔다.

⑧번~⑫번 계층을 결정하는 주요 규칙은 수술명, 동반질환유무, 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·영양 및 대사질환일 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 산욕기 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.

⑮번~⑯번 계층을 결정하는 주요 규칙은 수술명, 수술시간, 연령으로 나타났다. 코·입 및 인두 수술로서 혀 수술, 편도 및 아데노이드 수술, 임파계 수술, 소화기계 수술로서 충수돌기 수술, 비뇨생식기계 수술로서 신장수술, 전립선 및 정액낭 수술, 난소 수술, 자궁 절개 및 절제술, 근골격계 수술로서 안면골 및 관절수술, 그 외의 골 수술, 골절 및 탈골 정복술, 근·건·건막 및 활액낭 수술, 기타 골·관절 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~⑱번 계층은 수술명이 ⑮번~⑯번계층과 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.

⑲번~㉓번 계층을 결정하는 주요 규칙은 수술명, 수술시간, 주호소로 나타났다. 수술명은 ⑮번~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 시행받은 환자군으로 병원감염 발생확률은 0%였다.

3) 병원감염 발생의 규칙

표 4와 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 점수화하였다.

규칙 1의 경우 후두 및 기관 수술을 받은 환자군이 병원감염발생 확률은 100%이며, 이 규칙의 추론을 위한 일련의 과정을 보면 다음과 같다.

IF 수술명 EQUALS 후두 및 기관수술

THEN

NODE: 2

N: 31

**병원감염 발생확률**: 100%

**병원감염 미발생확률**: 0%

**규칙** 8**의 경우, 수술명이 뇌수술 등 중의 하나이고, 동반질환이 있으며, 연령이** 61**세 이상이면 병원감염발생확률이** 61.5%**이다. 이 규칙의 추론을 위한 일련의 과정은 다음과 같다.**

IF **연령** EQUALS 61**세 이상**

AND **동반질환** EQUALS **유**

AND **수술명** IS ONE OF : **뇌수술 등**

THEN

NODE: 17

N: 26

**병원감염 발생확률**: 61.5%

**병원감염 미발생확률**: 38.5%

표 4. 병원감염발생의 규칙

| 예측 집단 | 규칙 | 수술명 | 동반 질환 | 수술시간 | 연령 | 진단명 | 주호소 | 예측분류에 대한 신뢰도(%) |
|---|---|---|---|---|---|---|---|---|
| 병 원 감 염 발생 그룹 | 1 | 후두 및 기관수술 | - | - | - | - | - | 100.0 |
| | 2 | [1)]타액선수술 등 | 유 | - | - | [6)]신경계질환 등 | - | 100.0 |
| | 3 | [2)]혀수술 등 | - | 3시간 이상 | - | - | 신경계 및 감각계 | 100.0 |
| | 4 | 장관의 절개, 절제 및 문합술 | 무 | - | - | - | - | 100.0 |
| | 5 | [3)]뇌수술 등 | 유 | - | 41-50 | - | - | 94.1 |
| | 6 | 타액선수술 등 | 유 | - | - | 내분비,영양 및 대사질환 | - | 75.0 |
| | 7 | 뇌수술 등 | 유 | - | 21-30,51-60 | - | - | 68.5 |
| | 8 | 뇌수술 등 | 유 | - | 61- | - | - | 61.5 |
| | 9 | 혀수술 등 | 무 | 3시간이상 | - | - | 소화기계 | 60.0 |
| 병원 감염 미 발생 그룹 | 1 | [4)]안검수술 등 | 무 | - | - | - | - | 100.0 |
| | 2 | 혀수술 등 | - | 3시간이상 | - | - | 비뇨기계 | 100.0 |
| | 3 | 혀수술 등 | - | 1시간이내 | -60 | - | - | 99.5 |
| | 4 | [5)]장관수술을 제외한 수술 | 무 | - | - | - | - | 97.0 |
| | 5 | 혀수술 등 | - | 1-2시간 | - | - | - | 89.3 |
| | 6 | 타액선수술 등 | 유 | - | - | [7)]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 | - | - | [8)]감염성 및 기생충질환 등 | - | 72.0 |
| | 13 | 뇌수술등 | 유 | - | 31-40 | - | - | 61.5 |
| | 14 | 뇌수술 등 | 유 | - | -20 | - | - | 53.8 |
| | 15 | 타액선수술 등 | 유 | - | - | [9)]순환기계질환 등 | - | 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 절개,절제 및 문합술, 기타 장관수술, 직장 및 직장-S상결장 및 주위조직수술, 담낭 및 담관수술,

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

2) 혀수술 등 : 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

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

4) 안검수술 등 : 안검수술, 누관수술, 결막을 제외한 안과계수술, 귀수술, 혀·타액선·편도 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, 자궁경부수술, 기타 자궁부속기수술, 질 및 맹낭수술, 기타 내분비선수술, 제왕절개술, 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 및 염색체 이상

4) 병원감염발생 위험집단의 선정

병원감염 발생의 계층별 패턴과 규칙을 추출한 후, 병원감염 발생그룹을 세분화하여 다른 그룹에 비해 병원감염발생의 위험도가 높은 감염관리 대상을 찾아내기 위해 이익도표를 활용하였다. 표 6의 Node-by-Node 통계량을 보면, ‘Node’는 노드의 번호를 나타내고, ‘Node:n'은 마디에 속하는 개체의 수, ’Resp:n'은 마디에 속하는 목표범주의 개체의 수, 'Gain(%)'은 마디에 속하는 목표범주의 개체의 수/ 마디에 속하는 개체의 수, ‘Index(%)'는 특별한 마디에서의 목표범주의 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 병원감염발생 환자로서 이는 전체 병원감염발생 환자군 242명의 12.8%에 해당된다. 즉 규칙 1의 특성을 지닌 대상이 전체 병원감염 발생환자군 중 12.8%에 해당되는 것을 의미하며, ‘Gain(%)'가 ’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.

다음으로 규칙 14는 노드 31에 해당되고 이 때의 관찰된 개체의 수는 22명이고, 전체 대상자 1,523명의 1.4%에 해당되며, 관찰 된 개체수 22명 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 5).

따라서 병원감염발생 위험집단으로서 인덱스 점수가 높은 대상집단에 우선적인 질관리 전략이 수립되어진다면 다른 대상집단보다 상대적으로 높은 이득을 얻을 수 있음을 알 수 있었다.

표 5. **병원감염발생 환자군의 이익도표**(Node-by-Node **통계량**)

| 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 |
| : | | | | | | | |

또한 병원감염발생 위험집단의 관리대상범위를 설정하기 위해 병원감염 발생 위험집단에 대한 최적의 세분화가 필요했다. 이익도표의 누적 통계량은 병원감염관리의 대상집단 선정시 의사결정에 필요한 정보를 제공해주고 있다. 예를 들어 가장 병원감염 발생률이 높은 마디 2 만을 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 24 를 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 감염관리대상집단으로 관리하면 전체 병원감염환자군의 23.1%의 병원감염 관리효과를 가져올 수 있기 때문이다(표 6)(그림 5).

표 6 . 병원감염발생 환자군의 이익도표( 누적통계량)

| 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 |
| : | | | | | | | | |

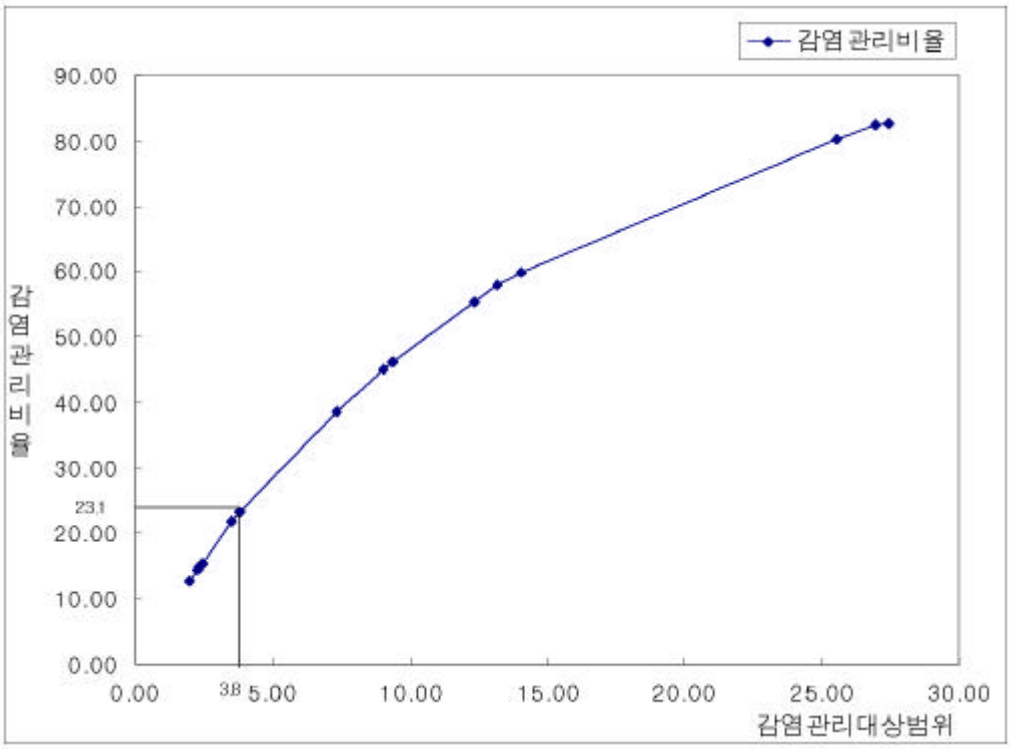


**그림 5. 감염관리 대상범위에 따른 감염관리비율**

### 5) 데이터마이닝 기법을 통해 도출된 병원감염발생규칙에 대한 타당성 검증

지금까지 나무구조의 끝마디에 대한 평가로서 이익도표에 의한 병원감염 발생위험집단을 선정해 보았다. 나무구조 전체에 대한 평가로서 실제 범주(Actual Category)가 의사결정나무의 추론규칙에 의해 예측된 범주(Predicted Category)를 비교하여 측정함으로써 데이터마이닝 결과인 병원감염 발생규칙의 타당성검증을 실시하였다(표 7).

병원감염의 발생여부를 예측했을 때 실제 병원감염으로 분류된 환자를 병원감염으로 예측한 감수성(Sensitivity), 병원감염 미발생으로 분류된 환자를 병원감염 미발생으로 예측한 특이성(Speciticity), 그리고 전체적으로 올바로 예측한 환자의 합을 전체 환자수로 나눈 예측도(Correct Prediction Rate)를 나타냈다. 감수성은 실제 병원감염 발생환자를 병원감염 발생환자로 예측하는 비율로 82.15%였고, 특이성은 실제 병원감염 미발생환자를 병원감염 미발생환자로 예측하는 비율로 83.3%였다. 따라서 전체 예측도는 82.79%의 비율을 보였다.

표 7. 병원감염발생 환자의 위험도표(예측도) 단위 : %

| 실제범주 | 예측된 범주 | | | 정분류율 (예측도) |
|---|---|---|---|---|
| | 감염발생 | 감염미발생 | 계 | |
| 감염발생 | 626(82.15) a | 135(17.71) | 762(100.00) | 82.79% |
| 감염미발생 | 126(16.53) | 635(83.33) b | 762(100.00) | |

a: Sensitivity b: Speciticity

## 4. 연관성 규칙을 활용한 병원감염 발생환자의 합병 질환 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 8).

표 8. 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 |
| 처치에 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 |

## 5. 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 이익도표를 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 기준으로하여 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 절개절제 및 분리술, 관절의 절개 및 절제술, 관절재건술중 어느 하나를 시행받았고 동반질환이 있으며 질환명이 신경계 질환, 주산기 병태,기타 증상, 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 3시간이상인 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 절개절제 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. 종격동 수술, 간수술, 췌장수술 중 어느 하나를 시행받았고 동반질환이 있으며, 연령이 21-30세, 51-60세인 경우 였다.

여덟째, 수술명이 뇌 및 두개골 뇌막수술, 척수수술, 흉막. 늑막. 종격동 수술, 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.

# Ⅴ. 고 찰

최근 의료기관의 전산화로 대용량의 데이터가 컴퓨터에 축적되고 있고, 의료기관들이 변화하는 의료환경에서 경쟁력을 확보하는 수단으로서 데이터마이닝 등 정보기술을 활용한 효율적인 CQI 활동이 요구되고 있는 현 시점에서, 본 연구는 질관리의 일환으로 병원감염관리의 현황파악을 위해 퇴원요약 DB로 SPC기법의 하나인 불량률 관리도를 적용하여 질적 문제 중 병원감염 발생률의 변동을 분석한 결과, 연도별로 관리한계선을 벗어나 있음을 알 수 있었으며, 이에 대한 심층분석을 위해 데이터마이닝 기법 중 의사결정나무분석 및 연관성 규칙을 적용하였다.

의사결정나무분석으로 추출된 24개의 규칙 중 예측분류에 대한 신뢰도가 50%이상인 9개의 규칙을 선정하여 병원감염관리의 지침으로 제시하였다. 또한 이익도표의 노드별 통계량을 통해 병원감염발생의 위험대상군을 선정하는 데 필요한 정보를 제공받을 수 있었으며, 누적 통계량을 통해 최적의 병원감염발생 위험집단을 설정할 수 있었으며, 설정된 범위에 따른 병원감염관리 효과도 미리 예측할 수 있어서 효율적인 CQI활동을 가능하게 했다.

병원감염을 설명하는 중요한 변수로서 수술명, 동반질환유무, 수술시간, 진단명, 연령, 주호소가 나왔다. 동반질환, 수술시간, 진단명, 연령은 기존의 감염발생규칙과 동일하게 나타났다. 동반질환이 있는 경우 병원감염 발생확률이 높았다. 이는 정희진 등(1999)의 보고에 의하면 병원감염발생이 환자의 기저질환별 차이가 있다는 것과 동일한 의미이다. 수술시간의 경우 수술시간이 길수록 병원감염 발생확률이 높았다. 김정순(1988)의 보

고에 의하면 특히 창상감염의 경우 수술시간이 1분 내지 30분 걸리는 수술에 비해 2시간 30분 이상 걸리는 수술은 병원감염 발생위험이 20배 높다는 것과 일치했다. 연령의 경우 뇌수술 등 및 혀 수술 등의 수술을 받은 경우 연령별로 병원감염 발생확률에 차이가 있는 것으로 나왔다. 이는 Turck 등(1981)의 보고에 의하면, 요로감염의 경우 60세 이상의 고령인 경우 병원감염위험이 높다고 했다. 환자의 건강상태를 나타내는 연령은 병원감염규칙 설정시 의미가 있다고 하겠다. 진단명 역시 동반질환과 마찬가지로 기저질환별 병원감염 발생확률에 차이가 있다는 것과 동일한 의미이다. 주호소는 환자가 내원하게 된 주된 이유로서 역시 진단명과 관련되는 변수이며 환자의 중증도를 설명해 주므로 병원감염발생 규칙 설정시 의미가 있다. 수술명은 병원감염발생의 주요 변수로서 후두 및 기관 수술, 뇌수술 등, 타액선 수술 등, 혀 수술 등, 안검수술 등의 다섯 그룹으로 분류되었으며, 현재 수술명에 따라 병원감염 발생률의 차이가 있다는 보고는 없었다. 그러나 수술명 자체가 병원감염발생의 직접적인 원인은 아닐지라도 이것을 통하여 환자의 특성 및 중증도를 파악할 수 있었으며, 중환자실 치료 등으로 인한 병원감염발생을 예측할 수 있었다.

감염환자의 주진단명과 합병진단간의 상병간의 연관성규칙을 적용한 결과 요로감염과 원인균과의 지지도가 가장 높았고, 수술부위의 파괴와 창상감염과도 높은 지지도를 보였으며 병원감염 중 요로감염이 가장 많으며 대부분 도뇨관과 관련된 것이라는 것을 알 수 있었다. 이것은 배영순(1995)의 연구에서도 요로감염이 전체 감염의 40%를 차지하고 이중 80% 이상이 도뇨관과 관련된다는 것과 거의 일치된 결과이다.

또한 Stephen등(1998)은 데이터마이닝 기법 중 연관성 규칙을 적용하여

녹농균(Pseudomonas Aeruginosa)과 항균제에 대한 병원감염발생 데이터를 분석하여 새로운 패턴을 발견했으나, 본 연구에서는 전체 퇴원환자의 퇴원분석 DB를 대상으로 의사결정나무기법을 적용하여 수술명을 중심으로 한 병원감염 발생규칙을 도출하여 병원감염발생 위험환자군의 예측을 통해 병원전체 감염관리전략수립에서의 의사결정에 필요한 정보를 제공해 줄 수 있다.

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, 항생제 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.

앞으로 병원감염발생의 변수로서 환자의 의식상태, 입원병실의 종류, 상재 요도카테타, 기계적 호흡, 면역억제제 사용 등을 충분히 반영하여야 할 것이고, 데이터마이닝은 작업의 많은 부분이 자료의 수집, 추출 그리고 정

제 등에 투입되기 때문에 이러한 작업을 단순화시킬 수 있는 데이터웨어하우스 및 데이터마트를 구축해서 다양한 원천으로부터의 양질의 변수를 확보하여 정보기술을 활용한다면 효율적인 CQI 활동에 많은 도움이 될 것이다.

# Ⅵ. 결 론

퇴원요약 DB를 이용하여 데이터마이닝 기법의 CQI 활동에의 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 의사결정나무기법을 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 병원감염발생규칙을 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, 요로감염과 그 원인균을 나타내는 상병과의 지지도가 '4.08'로서 일반환자보다 4배나 높은 연관성을 보였고, 뇌내출혈의 진단명을 가진 환자가 도뇨관으로 인한 요로감염을 동반할 경우 지지도가'1.46'으로서 일반환자보다 1.4배 높은 연관성을 보였다. 수술부위의 파괴라는 진단명을 가진 사람이 수술부위의 창상감염이라는 진단명을 동반할 경우 신뢰도 '75'로서 수술부위의 파괴는 수술부위의 창상감염을 가져오기 쉽다는 것을 알 수 있었다.

데이터마이닝 기법으로 병원감염 발생규칙을 추출함으로서 병원감염의 전향적인 감시활동을 가능하게 했다. 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

배영순. 일 병원에서의 병원감염관리활동 사례연구. 한국의료QA학회, 1995

병원통계. 대한의무기록협회, 1999

안선경. CQI의 개념과 필요성. 한국의료QA학회 1999: 1-20

유승흠. 양질의 의료관리. 수문사, 1993

용환승. 데이터마이닝. 그린, 1998

이경미, 21세기 의무기록분야의 대응전략. 병원관리종합학술대회연제집, 1999; 15(1): 237-251

이선희. 국내 질 향상 활동 현황과 전망. 한국의료QA학회 1999; 51-56

이정화, 의무기록의 표준화를 위한 퇴원진료정보분석. 연세대학교 보건대학원 석사논문,1996

정현, Data Mining기법을 이용한 인공신경망 인력변수선정에 관한 연

구. 서울대학교 석사학위논문, 1999

정희진, 김우주, 최선주 등. 병원성 요로감염의 비용추정연구: 의료비용 및 재원기간에 미치는 영향에 대한 연구; 대한병원감염관리학회 추계학술대회, 1999

조우현, CQI의 개념과 필요성. 한국의료QA학회, 1999: 1-20

채영문. 보건정보관리 .수문사, 1994

최종후, 한상태, 강현철 등. SAS Enterprise Miner를 이용한 데이터마이닝; 기능과 사용법. 자유아카데미, 1999

최종후, 한상태, 강현철 등. Answer Tree를 이용한 데이터마이닝 의사결정나무분석, 1998

호승희, 채영문, 조경원 등, 의료분야의 지식경영을 위한 데이터마이닝 응용. 대한의료정보학회지 1999; 5(3): 169-179

홍준현. 의무기록관리학. 고문사, 1996

Cho JH. Reengineering of enterprise information analysis environment by data warehousing. SIGDM'98 1998; 1: 5-33

Choi JH, Han ST, Kang HC, et al. Data mining decision tree analysis using Answer Tree. Korea: SPSS Academy, 1998

Greg Rogers, Ellen Joyner. Mining your data for health care quality improvement. SAS Institute,Inc.,Cary,NC.,1998

Internetsite[http://www.cio.com/archive/100196_inws_content.html]

Internetsite[http://www.cio.seoul.kr/main4.html).

Kim SM, Lee DH, Nam DW, et al. Concept hierarchy control method for knowledge discovery in relational database. KESS

processing 1998; Fall: 91-96

Lawrence SB. Data Mining: Sophisticated forms of managed modeling through artificial intelligence. J of Health Care Finance 1997; 23(3): 20-36

Lee DH, Nam DW, Kim SM, et al. Object generation hierarchy consultation for association rule discovery in relation database. KESS processing 1998; Fall: 91-96

Lee GH, Na GH. Development and evaluation of Decision Tree Model for bankrupcy prediction, SIGDM'99 1999; 2: 179-187

Prather JC, Lobach DF, Goodwin LK, et al. Medical data mining: knowledge discovery in clinical data warehouse. Proc AMIA 1997; Fall: 101-5

Raymond GC, Robert CL. Quality with confidence in healthcare, 1997

Stephen EB, Alan PS, Hardin JM, et al. Association rules and data mining in hospital infection control and public health surveillance. J of AMIA 1998; 5(4): 373-81

Turck M, Stamm W. Nosocomial infection of the urinary tract . AmJ MED 1981; 70: 651-70

# Abstract

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

Mi Ohk Suhn

Graduate school of

Health Science and Management

Yonsei University

(Directed by Professor Hae Jong Lee, Ph.D.)

This study provides an application of datamining approach to CQI 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 prover 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.