

# 진료 품질 향상을 위한 환자 데이터 맞춤형 분석 프로세스 개발: 외국인 환자를 중심으로

노을희\* · 김유정\* · 박상찬\*\*†

\*경희대학교 일반대학원 의료경영학과

\*\*경희대학교 경영대학

## Development of customized patient data analysis process for quality of care improvement : focused on foreign patients

Eul Hee Roh\* · Yoo Jung Kim\* · Sang Chan Park\*\*†

\*Department of Healthcare Management, Graduate School, Kyung Hee University

\*\*Department of Business Administration, Kyung Hee University

### ABSTRACT

**Purpose:** The purpose of this study was to find meaningful patient groups of disease using foreign patients data and analyze implemented test of the patient groups.

**Methods:** The data was collected by foreign patients' EMR data of K university hospital. The author proposed tree-form patients' characteristic diagram through statistical methods that association rule, proportion test, clustering using prescription information and questionnaire information.

**Results:** This study's analysis process was applied high blood data and diabetes data. Analysis showed other characteristic of meaningful patient groups in high blood and diabetes. In high blood, test implementation rate of patient group showed the differences. And in diabetes, test implementation rate of patient group and implemented test list showed differences.

**Conclusion:** The result of this study can play a role as basic data that can be clinical testing standard in preventive aspect. Eventually, 5 dimensions of SERVQUAL will be improved by this study's process.

**Key Words :** Quality of Care, Electronic Medical Record, Standard, Association Rule, Tree

● Received 15 June 2018, 1st revised 27 June, accepted 28 June 2018

† Corresponding Author(sangchan@khu.ac.kr)

© 2018, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

※ 연구비수혜: 2017년도 산업통상자원부의 제원으로 한국산업기술평가관리원의 지원을 받음(No. 10045403).

## 1. 서 론

기존 의학에서는 동일 진단을 받은 환자에게 동일한 치료법을 적용하는 평균적 환자에 맞춰진 One-Size-Fits-All 방법을 주로 활용했다(Jung, 2015). 하지만 이제는 의료 지식의 누적과 함께 다양한 임상 데이터의 축적도 함께 이뤄지면서 환자별 특이성을 고려하여 최적의 치료방법을 제시하는 맞춤 의료의 중요성이 커지고 있다. 맞춤 의료에서는 환자별 특이성을 고려하는 것이 중요하기 때문에 축적된 임상 데이터로부터 환자가 어떤 특이한 집단에 속하는지 판단하는 것이 기반 정보로 활용될 수 있다(Lee, 2010). 또한 기존에는 환자의 질병이 발병했을 때 그 질병을 치료하는 쪽에 초점을 맞추었지만, 이제는 유병 환자들의 특성을 미리 파악하여 질병을 예방하는 예방의료 쪽에 초점을 맞춰야 할 필요가 있다. 실제로 빅데이터, 인공지능 등과 같은 새로운 기술의 등장으로 전 세계 보건의료 패러다임은 ‘치료 중심’에서 ‘예측 및 예방 중심’으로 변화하고 있다.

이러한 의료 패러다임의 변화와 함께 외국으로 이동하여 현지의 의료기관 등을 통해 질병을 치료하거나 건강을 증진하는 의료관광의 시장규모도 지속적으로 커지고 있다. 국내에서도 2009년 외국인환자 유치가 허가된 이후로 매년 외국인환자 수가 증가하고 있으며, 2016년 한 해 동안 약 36만 명의 외국인이 국내 의료기관을 방문했다. 인종에 따른 치료 효과의 차이가 입증된 다양한 연구들이 외국인과 한국인의 치료 방법에 차이를 두어야한다는 점을 시사하고 있지만, 국내에서는 아직 외국인 환자 치료 시 참고할 수 있는 증례, 진료지침 등이 부족한 실정이다.

의료는 질병을 치료하는 과정에서 다 분야 간의 참여와 협력이 요구되는 특성을 가지고 있지만, 치료 과정에서 관련 전문직 간의 의사소통 문제 및 단편적이고 일관성 없는 환자관리로 인해 불필요한 검사나 처치의 시행, 지연 혹은 중단 등의 상황이 발생하게 되어 환자관리의 질이 떨어지는 경우가 발생한다(Graybeal, 1993). 이러한 상황을 방지하기 위해서는 검사·시행과 관련된 표준화된 문서가 필요한데, 그 중 하나가 특정 환자나 질환의 치료과정을 관리함에 있어서 목표를 미리 설정하고, 최대한 효율적으로 달성하기 위해 수행하는 의료행위를 일정한 시간적 틀에 따라 적절하게 지정해놓은 환자관리계획인 표준진료지침(Clinical Pathway)이다(Kim, 2015). 그러나 전 세계적으로 의료 환경이 급속도로 변화하고 있고, 의료 기술도 빠르게 발달하고 있기 때문에 진단·치료와 관련된 표준 문서도 시대에 부응하여 지속적으로 새롭게 변화시켜 나가야 한다. 의료 분야에서는 환자의 건강검진 정보와 질병 정보, 의사의 진료 정보와 치료 결과, 환자의 유전자 정보 등의 방대한 양의 정보가 발생하는데(Jung, 2015), 이러한 데이터들을 활용하여 진료 표준을 도출하면 분석한 데이터에 따라 다른 결과를 보여주기 때문에 의료 환경의 변화에 맞춰서 지속적으로 업데이트된 자료를 제공할 수 있다.

이에 본 연구에서는 외국인 환자 데이터를 활용하여 질병 별로 유의미한 특성을 가진 환자 그룹을 파악하고 해당 환자들이 시행한 검사를 분석하는 프로세스를 개발함으로써 예방 측면의 검사 시행 표준이 될 수 있는 기초자료를 제공하고자 한다. 세부적으로는 환자의 처방 정보, 문진 정보를 연관규칙, 비울검정, 클러스터링 등의 통계적 방법을 통해 분석하여 트리형 환자 특성 다이어그램을 제시하고자 한다.

## 2. 문헌고찰

### 2.1 환자 데이터 활용 연구

의료 분야에서는 EMR(Electronic Medical Record)의 도입으로 인해 환자와 관련된 데이터가 대량으로 축적되고 있으며, 이러한 데이터를 활용한 연구도 활발히 진행되고 있다. EMR에는 환자의 인적사항, 건강상태, 병력, 진단, 검사, 치료에 관한 모든 정보가 포함되어 있다. 기존 연구에서는 집단 간 위암 재발 여부에 영향을 미치는 변수 분석(Kwak, 2015), 균형증의 여부에 따른 급성신우신염 환자의 임상적 양상의 차이 분석(Hwang, 2009), 중증 뇌 손상으로 바비튜레이트 혼수요법을 적용한 환자의 임상적용결과 및 예후영향인자 분석(Kim, 2010) 등으로 데이터가 다양하게 활용되었다. 또한 환자의 생활습관, 행동적 특성 요인 중 특정 질병에 영향을 미치는 변수를 찾기 위해 문진을 통해 얻은 데이터를 활용한 연구들도 많았다. An(2017)은 유방 초음파를 받은 환자들의 문진표를 활용해서 유방암 위험인자를 분석했고, Oh(2011)은 건강문진표 전산자료로 상부 위장관 질환 군별로 영향요인과 소화기 증상이 각각 어떤 차이가 있는지를 파악했다. 이 외에도 2차 건강 검진자 판별을 위한 통계적 모형을 구축하거나(Choi, 2004), 임상검사결과와 교차분석하여 검진자료로서의 활용가능성을 평가한(Joe, 2002) 연구 등 다양한 방향으로 문진 데이터가 활용되었다.

### 2.2 의료분야에서의 연관규칙 활용 연구

대량의 데이터가 발생하는 의료 분야에서는 다양한 데이터 마이닝 기법이 활용되고 있으며, 대표적으로 베이저안 네트워크(Bayesian Network), 신경망(Neural Network), 의사결정나무(Decision Tree), 연관규칙, 회귀분석 등의 기법이 있다. 특히 베이저안 네트워크, 의사결정나무, 연관규칙과 같은 확률론은 임상 분야에서의 다양한 연구를 통해 유용성이 입증되고 있다. 본 연구에서는 연관규칙 분석을 활용한 연구를 중점적으로 살펴보았다. 연관규칙은 항목 집합으로 표현된 거래에서 각 항목간의 연관성을 반영하는 규칙이며, 연관규칙을 평가하는 기준에는 지지도(Support), 신뢰도(Confidence), 향상도(Lift) 등이 있다(Kim, 2008). 연관규칙은 주로 상병의 연관성을 분석하기 위해 활용되었으며, 뇌경색증과 동반되는 여러 질환들 사이의 연관성 분석(Lee, 2010), 심근경색을 일으키는 원인의 발견(Lee, 2013), 당뇨병과 다른 질병, 증상, 투약 등의 연관성 분석(Lakshmi, 2014) 등 다양한 질병에 적용되었다. 질병뿐만 아니라 전자의무기록 데이터를 기반으로 연관규칙마이닝을 이용하여 진료단계별 규칙을 추출하고(Bae, 2005), 추출, 분류, 정화 과정을 거친 분말을 활용하는 중국식 허브 치료 제품인 CHP의 공통처방패턴을 파악하는데 활용했다(Chu, 2015). 또한 의료데이터 특성을 고려해서 빈번한 항목과 빈번하지 않지만 의학적으로 의미 있는 항목들을 대상으로 연관 규칙을 구성하여 의료 전문가의 의사 결정에 도움을 주기 위한 시스템을 제안하기도 했다(Park, 2016).

### 3. 분석 방법

#### 3.1 활용 데이터

본 연구에서는 2011년 4월 27일부터 2017년 11월 24일까지 K 대학병원에 내원한 외국인 환자 중 IRB(Institutional Review Board, 임상시험심사위원회) 심의 확정을 받은 총 1,106명 환자의 EMR(Electronic Medical Record) 데이터를 활용했다. K 대학병원의 EMR 데이터는 크게 상병코드, 처방종류, 처방명 등으로 구성된 처방정보와 성별, 국가, 연령, 흡연습관, 음주습관, 가족력 등으로 구성된 문진정보로 구성되어 있다. 본 연구에서는 처방정보 중 차트번호, 환자명, 상병코드, 처방종류, 처방명 데이터를 활용했고, 문진정보 중 성별, 국가, 연령, 흡연 여부, 음주여부 데이터를 활용했다.

#### 3.2 분석 방법

본 연구에서는 환자정보 분석 프로세스를 개발하는 과정에서 연관규칙, 비유효검정, 클러스터링 등의 통계적 기법을 활용하였다. 본 연구의 분석 프로세스는 크게 환자 그룹 클러스터링, 환자 그룹 선정, 검사시행률 도출, 트리형 환자 특성 다이어그램 생성 네 단계로 구성되었다. 본 연구의 분석 방법을 연구 모형으로 나타내면 다음과 같다.

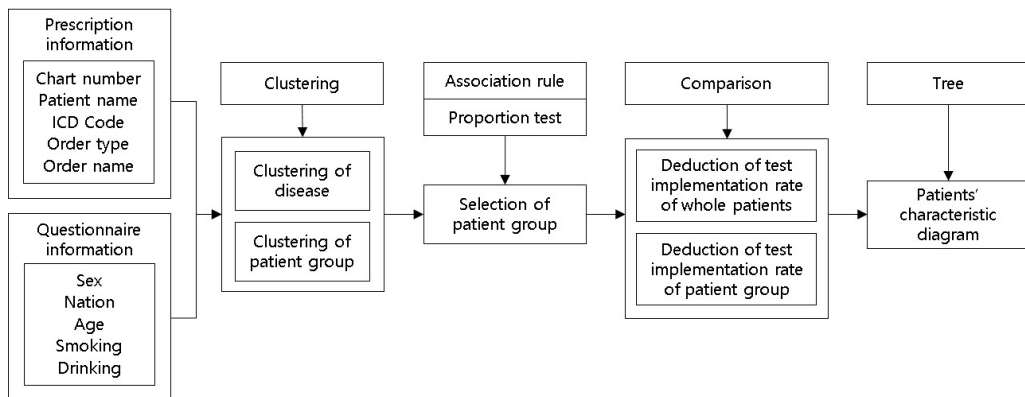


Figure 1. Research Model

##### 3.2.1 환자 그룹 클러스터링

본 단계는 전체 환자 데이터를 질병 및 환자 그룹 별로 클러스터링을 하는 과정이다. 본 연구에서는 환자 그룹을 전체 환자 중 인구통계학적 요인(국가, 성별, 연령)과 라이프스타일 요인(음주여부, 흡연여부) 등과 같은 개인의 특성과 관련된 변수들의 조합과 일치하는 환자들의 그룹으로 정의했다. 본 연구에서는 분석 대상 환자 표본의 수가 적기 때문에 국가 변수는 환자들의 국적이 가장 많은 것으로 나타난 러시아와 그 외 국가로 나누고, 연령은 50세 미만과 50세 이상으로 나눠서 환자 그룹을 분류했다. 사용하는 변수의 개수에 따라서 클러스터링 해야 하는 환자 그룹의 수가 달라지며, 예를 들어 변수가 3개일 때는  $(2 \times {}_3C_1) + (2 \times 2 \times {}_3C_2) + (2 \times 2 \times 2 \times {}_3C_3)$ 의 식을 따라 26개의 환자 그룹으로 분류된다. 본 연구에서는 5개의 변수를 활용하였고, 이에 따라 전체 환자를 총 242개의 환자 그룹으로 분류했다.

**Table 1.** Example of Variable Combination

variable combination	example of combination	
1 variable	male	female
	russia	other countries
2 variables	male, smoking	male, non-smoking
	female, under 50	female, over 50
3 variables	male, over 50, drinking	male, over 50, non-drinking
	russia, smoking, drinking	russia, non-smoking, drinking
4 variables	female, other countries, under 50, smoking	female, other countries, over 50, smoking
	male, under 50, non-smoking, drinking	male, under 50, non-smoking, non-drinking
5 variables	female, russia, under 50, smoking, drinking	male, russia, under 50, smoking, drinking
	female, other countries, over 50, non-smoking, drinking	male, other countries, over 50, non-smoking, drinking

### 3.2.2 연관 규칙을 활용한 환자 그룹 선정

본 연구에서는 연관규칙의 신뢰도, 지지도, 향상도 개념을 활용해서 전체 환자 그룹 중 유의한 환자 그룹을 선정했다. 본 연구에서 신뢰도는 전체 환자 중 특정 조건에 해당하는 환자 중에서 특정 질병 환자 중 특정 조건에 해당하는 환자 수를 나타낸 값을 의미한다. 또한 지지도는 전체 환자 중에서 특정 질병에 해당하는 환자 수를 나타낸 값이다. 앞서 언급한 신뢰도를 지지도로 나눈 값이 향상도를 의미하며, 일반적으로 향상도가 1에 가까우면 대상 품목 간 상호 독립적인 관계이고, 1보다 크면 양의 연관성, 1보다 작으면 음의 연관성을 갖는다(Yoo, 2016). 이에 본 연구에서는 향상도가 1이상으로 나타난 환자 그룹은 유의하다고 판단했다. 본 연구에서 향상도가 1보다 크다는 것은 조건을 고려했을 때의 유병률과 조건을 고려하지 않았을 때의 유병률에 차이가 있다는 것을 의미한다. 또한 향상도가 1보다 크다는 것이 유병률 향상에 유의한지를 확인하기 위해 추가로 단일표본비율검정을 실시했다. 비율검정은 비율에 대한 통계적 차이의 유의성을 검정하는 것으로 Z분포를 가정하며, 비율검정 공식을 본 연구에 대입하여 Z값을 도출하는 공식으로 나타내면 다음과 같다.

$$Z = \frac{\text{지지도} - \left(3 \times \sqrt{\frac{\text{지지도}(1 - \text{지지도})}{\text{전체환자수}}}\right)}{\left(3 \times \sqrt{\frac{\text{지지도}(1 - \text{지지도})}{\text{전체환자수}}}\right)}$$

본 연구에서는 정규분포에서 99.74%의 확률을 나타내는 구간이 평균을 중심으로 ±3σ 사이에 있으므로 P0값으로 3을 넣어서 계산식을 도출했다. 위의 공식을 통해 도출한 Z값의 p값이 유의수준 0.05보다 작게 나오는 향상도를 구해서 특정 향상도 이상의 값을 나타내는 변수 조합을 유의한 환자 그룹으로 선정하였다.

### 3.2.3 환자 그룹 검사 시행률 도출

본 단계에서는 특정 질병 전체 환자의 검사 시행률과 앞서 선정한 각 환자 그룹의 검사 시행률을 계산해서 비교하는 과정을 진행했다. 기존 연구의 검사 시행률에 대한 의미를 살펴보면 Hong(2009)은 당화혈색소 검사 시행률을

전체 연구대상자 중 1년에 1회 이상 검사를 받은 경험이 있는 환자의 비율로 측정했다. 또한 건강보험심사평가원에서 매년 실시하는 천식 적정성평가 보고서에서는 폐기능검사 시행률을 천식 환자 중 폐기능검사를 시행한 환자의 비율로 정의했다. 본 연구에서도 기존 연구와 동일한 맥락으로 검사 시행률을 특정 그룹에 포함된 전체 환자 중 검사를 시행한 환자의 비율로 계산했다.

### 3.2.4 트리형 환자 특성 다이어그램 생성

마지막 단계에서는 이전 단계에서 선정한 환자 그룹들을 branch로 하고, 환자 그룹의 특성 요인들을 node로 구성한 트리형 환자 특성 다이어그램을 생성하였다. 트리 구조는 방대한 환자-질병 데이터를 컴팩트하게 관리할 수 있다는 장점이 있다. 본 연구에서는 node의 순서를 임의의 중요도에 따라 성별, 국가, 연령, 흡연여부, 음주여부 순으로 구성했다. 각 branch에는 그룹에 포함된 환자의 수, 향상도 값이 표시되어 있다. 본 연구에서는 환자의 수가 많고 node의 수가 적으며 향상도가 높게 나타난 branch는 대표성이 높은 것으로 판단하고, 질병의 유병 가능성이 높은 특성을 가진 그룹을 의미한다고 보았다. 반대로 환자의 수는 적고 node의 수는 많지만 향상도가 낮게 나타난 branch는 대표성이 낮은 것으로 판단했다.

## 4. 분석 결과

본 연구에서 개발한 분석 프로세스를 K 대학병원의 총 1,106명의 외국인 환자 EMR 데이터 중 해당하는 환자 수가 많은 질병인 고혈압과 당뇨병 데이터에 각각 적용하여 분석을 시행했다. 분석 결과, 고혈압과 당뇨병의 유의미한 환자 그룹의 특성은 다르게 나타났으며, 고혈압에서는 환자 그룹 별로 검사 시행률에서 차이를 보였고, 당뇨병에서는 환자 그룹별 검사 시행률뿐만 아니라 시행한 검사 항목에서도 차이가 나타났다.

### 4.1 고혈압 환자 데이터 분석 결과

47명의 고혈압 환자 데이터에 적용하여 분석을 시행한 결과, 242개의 환자 그룹 중 유의미한 환자 그룹은 7개로 나타났다. 연관규칙을 활용해 각 환자 그룹의 향상도를 구하고, 비율 검정을 통해 향상도가 3.96 이상을 나타내는 환자 그룹이 유의미하다는 것을 파악했다. 선정된 고혈압 환자 그룹은 아래 그림과 같다.

Table 1. Selected Patient Group of High Blood

Group's characteristic					High blood patient number	Whole patient number	Confidence	Support	Lift
Sex	Nation	Age	Smoking	Drinking					
Male	Russia	Over 50	No	Yes	3	15	0.20	0.042	4.70
Male	-	Over 50	No	Yes	4	23	0.17	0.042	4.09
-	Russia	Over 50	No	Yes	4	22	0.18	0.042	4.27
-	Russia	Over 50	No	-	27	136	0.19	0.042	4.67
-	Russia	Over 50	-	Yes	4	9	0.44	0.042	10.4
-	Russia	Over 50	-	No	24	133	0.18	0.042	4.24
-	Other	Over 50	-	Yes	2	8	0.25	0.042	5.88

다음 단계로 고혈압을 진단받은 전체 환자들이 시행한 73개 검사를 기준으로 앞서 선정한 각 환자 그룹의 검사 시행률을 계산 및 비교하는 과정을 수행했다. 본 논문에는 고혈압 전체 환자를 기준으로 검사 시행률이 높게 나타난 20개 검사의 시행률을 표로 나타냈다. 제시된 표를 보면 동일한 검사에서 고혈압 전체 환자의 검사 시행률과 각 환자 그룹의 검사 시행률이 차이를 보인다는 것을 알 수 있다.

**Table 3.** Test Implementation Rate of High Blood

Test name	Implementation rate of whole patients	Patient group 1	Patient group 2	Patient group 3	Patient group 4	Patient group 5	Patient group 6	Patient group 7
Hb	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
Hct	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
MCH	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
MCHC	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
MCV	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
PLT	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
RBC	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
WBC	63.83%	100%	100%	100%	62.96%	100%	58.33%	100%
ALT(GPT)	61.70%	100%	100%	100%	62.96%	100%	58.33%	100%
AST(SOT)	61.70%	100%	100%	100%	62.96%	100%	58.33%	100%
ALP	59.57%	100%	100%	100%	62.96%	100%	58.33%	100%
Cr	59.57%	100%	100%	100%	59.26%	100%	54.17%	100%
BUN	57.45%	100%	100%	100%	59.26%	100%	54.17%	100%
FBS	57.45%	100%	100%	100%	55.56%	100%	50%	100%
Basophil	55.32%	66.67%	50%	75.00%	55.56%	75.00%	54.17%	50%
Eosinophil	55.32%	66.67%	50%	75.00%	55.56%	75.00%	54.17%	50%
K	55.32%	100%	100%	100%	55.56%	100%	50.00%	100%
Lymphocyte	55.32%	66.67%	50%	75%	55.56%	75%	54.17%	50%
Monocyte	55.32%	66.67%	50%	75%	55.56%	75%	54.17%	50%
Na	55.32%	100%	100%	100%	55.56%	100%	50.00%	100%

마지막으로는 위의 7개 환자 그룹을 branch로 구성해서 트리형 환자 특성 다이어그램을 생성했다. 고혈압 환자 특성 다이어그램에서는 러시아, 50세 이상, 비 흡연의 특성을 가지고 있는 네 번째 환자 그룹이 가장 대표성이 높은 그룹인 것으로 나타났다.

아래 제시된 고혈압 환자 특성 다이어그램을 보면 포함된 환자 수가 가장 많은 상위 2개 branch는 4, 2번 순으로 나타났고, node는 1번부터 4번까지의 branch가 3개로 가장 적은 수로 구성되었고, 향상도가 높은 상위 2개 branch는 4, 1번 순으로 나타났다. 결론적으로 3개의 조건을 모두 충족하는 러시아, 50세 이상, 비 흡연자의 특성을 가지고 있는 4번 branch를 가장 대표성이 높은 환자 그룹으로 판단했다.

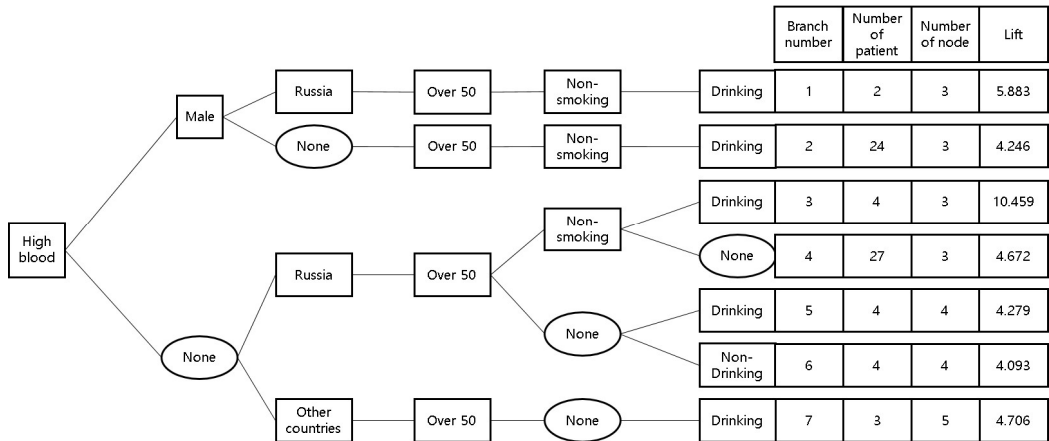


Figure 2. High Blood Patients' Characteristic Diagram

### 4.2 당뇨병 환자 데이터 분석 결과

34명의 당뇨병 환자 데이터에 적용하여 분석을 시행한 결과, 242개의 환자 그룹 중 유의미한 환자 그룹은 15개로 나타났다. 고혈압 데이터와 동일한 통계적 방법을 통해 향상도가 2.95 이상을 나타내는 환자 그룹이 유의미하다는 것을 파악했다. 선정된 당뇨병 환자 그룹은 아래 그림과 같다.

Table 4. Selected Patient Group of Diabetes

Group's characteristic					Diabetes patient number	Whole patient number	Confidence	Support	Lift
Sex	Nation	Age	Smoking	Drinking					
Male	Russia	Over 50	Yes	No	2	15	0.133	0.031	4.33
Male	Russia	Over 50	Yes	-	3	30	0.100	0.031	3.25
Male	Russia	-	Yes	No	2	20	0.100	0.031	3.25
Male	Other	Over 50	No	No	5	39	0.128	0.031	4.17
Male	Other	Over 50	No	-	5	47	0.106	0.031	3.46
Male	Other	Over 50	-	No	6	48	0.125	0.031	4.06
Male	-	Over 50	Yes	No	3	24	0.125	0.031	4.06
-	Russia	Over 50	Yes	Yes	2	17	0.117	0.031	3.82
-	Russia	Over 50	Yes	No	2	19	0.105	0.031	3.42
-	Russia	Over 50	Yes	-	4	6	0.666	0.031	21.68
-	Russia	Over 50	-	Yes	3	9	0.333	0.031	10.84
-	Other	Under 50	Yes	-	2	9	0.222	0.031	7.22
-	Other	Under 50	-	Yes	2	12	0.166	0.031	5.42
-	Other	Over 50	-	No	10	99	0.101	0.031	3.28
-	-	Over 50	Yes	No	3	33	0.090	0.031	2.95



다음 단계로 당뇨병을 진단받은 전체 환자들이 시행한 71개 검사를 기준으로 앞서 선정된 각 환자 그룹의 검사 시행률을 계산 및 비교하는 과정을 수행했다. 당뇨병 전체 환자를 기준으로 검사 시행률이 높게 나타난 10개 검사의 시행률을 표로 제시했으며, 환자 그룹별 검사 시행률뿐만 아니라 시행한 검사 항목에서도 차이를 보였다.

**Table 5.** Test Implementation Rate of Diabetes

Test name	Implementation rate of whole patients	Patient group 1	Patient group 2	Patient group 3	Patient group 4	Patient group 5	Patient group 6	Patient group 7
Cr	61.76%	0%	0%	0%	100%	100%	100%	33.33%
ALT(GPT)	58.82%	0%	0%	0%	100%	100%	100%	33.33%
BUN	58.82%	0%	0%	0%	100%	100%	100%	33.33%
AST(SOT)	55.88%	0%	0%	0%	100%	100%	100%	33.33%
FBS	55.88%	0%	0%	0%	100%	100%	100%	33.33%
Hb	52.94%	0%	0%	0%	60%	60.00%	66.67%	33.33%
HbA1c	52.94%	0%	0%	0%	100%	100%	83.33%	0%
Hct	52.94%	0%	0%	0%	60%	60%	66.67%	33.33%
MCH	52.94%	0%	0%	0%	60%	60%	66.67%	33.33%
MCHC	52.94%	0%	0%	0%	60%	60%	66.67%	33.33%
Test name	Patient group 8	Patient group 9	Patient group 10	Patient group 11	Patient group 12	Patient group 13	Patient group 14	Patient group 15
Cr	0%	0%	0%	33.33%	100%	100%	90%	33.33%
ALT(GPT)	0%	0%	0%	0%	100%	100%	90%	33.33%
BUN	0%	0%	0%	33.33%	100%	100%	80%	33.33%
AST(SOT)	0%	0%	0%	0%	100%	100%	80%	33.33%
FBS	0%	0%	0%	33.33%	50%	50%	80%	33.33%
Hb	0%	0%	0%	33.33%	100%	50%	70%	33.33%
HbA1c	0%	0%	0%	0%	0%	50%	80%	0%
Hct	0%	0%	0%	33.33%	100%	50%	70%	33.33%
MCH	0%	0%	0%	33.33%	100%	50%	70%	33.33%
MCHC	0%	0%	0%	33.33%	100%	50%	70%	33.33%

마지막으로는 위의 15개 환자 그룹을 branch로 구성해서 트리형 환자 특성 다이어그램을 생성했다. 아래 제시된 당뇨병 환자 특성 다이어그램을 보면 포함된 환자 수가 많은 상위 5개 branch는 6, 14, 4, 5, 10번 순으로 나타났고, node는 10번부터 15번까지의 branch가 3개로 가장 적은 수로 구성되었고, 향상도가 높은 상위 5개 branch는 10, 11, 12, 13, 1번 순으로 나타났다. 결론적으로 3개의 조건을 모두 충족하는 러시아, 50세 이상, 흡연자의 특성을 가지고 있는 10번 branch를 가장 대표성이 높은 환자 그룹으로 판단했다.

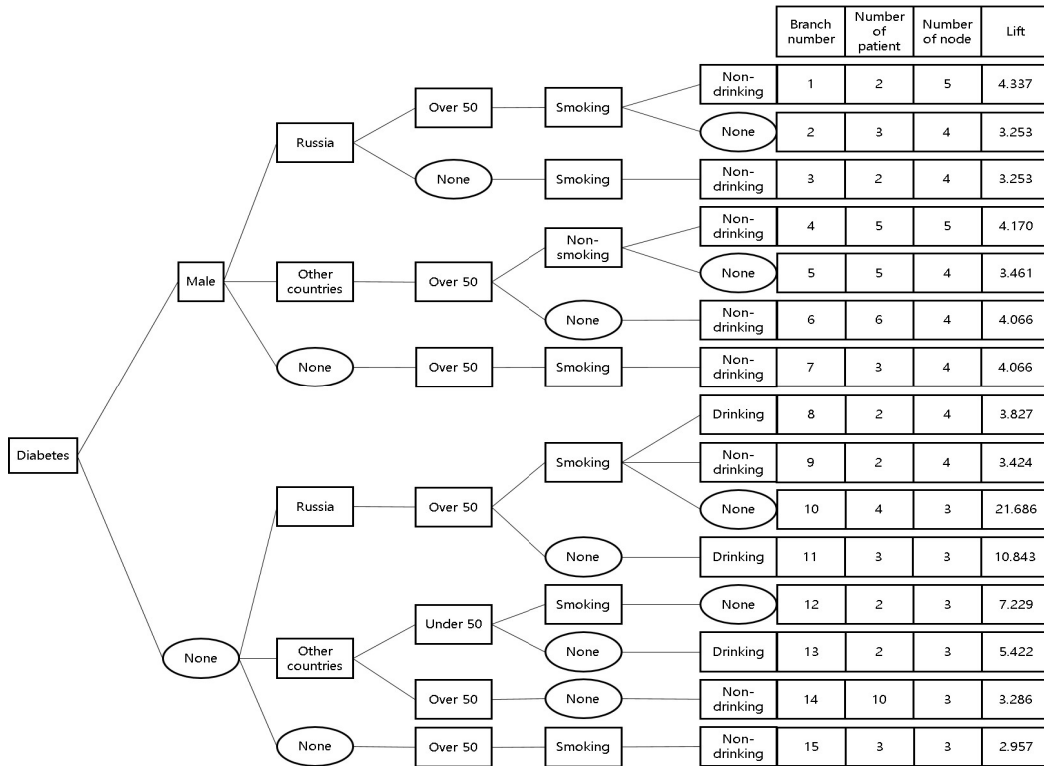


Figure 3. Diabetes Patients' Characteristic Diagram

## 5. 결 론

빅데이터, 인공지능 등과 같은 새로운 기술의 등장으로 의료 패러다임이 맞춤 의료, 예방 중심으로 변화하고 있고, 글로벌화로 인해 국가의 경계가 사라진 형태의 의료관광 산업이 활발해지고 있다. 국내 의료기관을 방문하는 외국인 환자 수도 점점 늘고 있고, 외국인 환자 진료 시 참고할 수 있는 충분한 자료의 필요성도 커지고 있는데, 이러한 상황에서 진료의 기준이 되는 표준진료지침이 참고 자료로 유용하게 활용될 수 있다. 또한 의료 환경이 급속도로 변화하고 의료 기술도 빠르게 발달하고 있기 때문에 진단·치료와 관련된 표준 문서도 시대에 부응하여 지속적으로 새롭게 변화시킬 필요가 있다.

이에 본 연구에서는 K 대학병원의 외국인 환자 EMR 데이터를 활용하여 질병 별로 유의미한 특성을 가진 환자 그룹을 파악하고 해당 환자들이 시행한 검사를 분석하는 프로세스를 개발했다. 분석 프로세스 개발 과정에서 연관규칙, 비율검정, 클러스터링과 같은 통계적 기법이 활용되었다. 또한 프로세스는 크게 환자 그룹 클러스터링, 환자 그룹 선정, 검사시행률 도출, 트리형 환자 특성 다이어그램 생성 네 단계로 구성되었다. 본 연구의 프로세스는 분석한 데이터에 따라 다른 결과를 보여주기 때문에 의료 환경의 변화에 맞춰서 지속적으로 업데이트된 자료를 제공할 수 있고, 예방 측면의 검사 시행 표준이 될 수 있는 기초자료로서의 역할을 할 수 있다.

본 연구에서는 고혈압과 당뇨병 데이터에 각각 적용해서 분석을 시행했고, 두 질병의 분석 결과는 여러 부분에서 차이를 보였다. 고혈압에서는 향상도 3.96 이상의 유의미한 환자 그룹이 7개가 선정되었고, 러시아, 50세 이상, 비흡연자의 특성을 가지고 있는 네 번째 환자 그룹의 대표성이 가장 높은 것으로 판단했다. 당뇨병에서는 향상도가 2.95 이상으로 나타난 15개의 환자 그룹이 유의미한 것으로 나타났고, 러시아, 50세 이상, 흡연자의 특성을 보이는

열 번째 환자 그룹의 대표성이 가장 높은 것으로 판단했다. 또한 고혈압에서는 환자 그룹 별로 검사 시행률에서 차이를 보였고, 당뇨병에서는 환자 그룹별 검사 시행률뿐만 아니라 시행한 검사 항목도 환자 그룹마다 다른 것으로 나타났다.

본 연구의 분석 프로세스를 통해 도출되는 결과 자료는 향후 초진 외국인 환자의 특성과 기존의 환자 특성 다이어그램에 나타난 환자 그룹의 특성을 비교해서 초진 환자가 시행할 검사 항목과 검사의 우선순위를 제시하는데 활용할 수 있다. 비용 대비 효율성이 높은 검사들을 우선적으로 제시함으로써 환자가 불필요한 검사를 시행하는 것을 방지할 수 있고, 환자의 금전적 부담을 줄이는데 도움을 줄 수 있다. 또한 이 프로세스를 통해 궁극적으로 서비스 품질을 측정하기 위한 도구인 SERVQUAL의 5가지 차원(유형성, 신뢰성, 반응성, 확신성, 공감성)을 향상시킬 수 있다. 장기간에 걸쳐 분석하는 환자 데이터가 많아질수록 환자 특성에 따른 검사 시행 표준의 신뢰성은 높아지며, 이로 인해 의료진은 환자에게 정확한 진단·치료 방법을 제시할 수 있다. 프로세스를 활용하여 신속하게 환자 특성에 적합한 검사 세트를 추천할 수 있기 때문에 반응성도 향상될 수 있다. 또한 환자 그룹을 트리 형태로 클러스터링 해서 보여준다는 점에서 유형성도 높아지고, 환자 개인의 특성에 맞는 의료서비스를 제공함으로써 제공함으로써 환자의 공감성이 향상될 수 있다.

개인의 특성을 고려해서 환자 특성 다이어그램을 도출하는 본 연구의 프로세스는 향후 건강검진 프로그램의 검사 항목을 선정할 때에도 활용할 수 있다. 각급 병·의원에서 시행되는 기존의 종합검진 프로그램은 수검자의 개인 특성, 위험요인 등을 고려하지 않은 채 획일적인 검사 항목으로 구성되어 실시되는 경향이 있기 때문에 의료비 증가를 초래할 가능성이 있으며, 시행되는 각 검사 항목에 대한 의학적 근거도 논란의 대상이 되고 있다(Yang, 2006; Park, 2009). 특정 질병을 앓고 있는 기존 수검자의 특성을 분석해서 개인의 특성에 따라 시행 검사가 다르게 구성된 검진 프로그램을 개발하여 이러한 문제점을 보완한다면 건강검진의 효과와 효율성이 높아질 것이다.

본 연구는 외국인 환자 진료 품질의 향상을 위한 하나의 방안을 제시했다는 점에서 의의가 있지만, 일개 대학병원의 한정된 환자 데이터를 대상으로 분석했기 때문에 분석 결과가 편향되어 있을 가능성이 있고 환자 수의 부족으로 인해 연구 결과의 일반화가 어렵다. 또한 합병증을 가지고 있는 환자의 경우에는 합병증에 따라 시행하는 검사가 다르게 나타나기 때문에 이러한 환자들은 따로 분석을 시행해야 하지만, 본 연구에서 활용한 데이터의 합병증 환자 수가 부족하여 합병증에 대한 부분을 고려한 분석 단계까지 시행하지 못했다. 본 연구에서는 변수 기준을 설정할 때 임의적으로 연령 요인에서 50세를 기준으로 두 그룹으로 나누고, 국가 요인에서 러시아와 러시아가 아닌 국가의 두 개 그룹으로 나눠서 분석을 진행했는데, 이는 객관성이 떨어진다는 한계점을 가지고 있다. 또한 본 연구에서는 통계적 방법으로 접근했기 때문에 향후 실무적 활용 가능성에 대한 임상적 검증이 필요하고, 환자에게 개별검사가 아닌 세트 형태로 검사를 제시하기 위해 유사한 성격의 검사들을 그룹화 하는 과정에서도 임상적 도움이 필요할 것으로 생각된다.

## REFERENCES

- An, H., Yang, S. H., Im, C. I., and Lee, J. S. 2017. "Analysis of Risk Factors Affecting Breast Cancer Incidence: Breast Ultrasonography." *Journal of the Korean Society of Radiology* 11(4):289-295.
- Bae, I. H., Kim, J. S., and Kim, Y. N. 2005. "An Association Rule Mining Approach to Extract Clinical Pathways from EMR." *Proceedings of Korea Information Processing Society Spring Conference* 12(1):577-580.
- Choi, I. J., and Lee, S. S. 2004. "Development of Statistical Model for Distinction of Secondary Medical Examination Recipient using Health Questionnaire." *Journal of Basic Science* 14:1-15.

- Chu, S. M., Shih, W. T., Yang, Y. H., Chen, P. C., and Chu, Y. H. 2015. "Use of traditional Chinese medicine in patients with hyperlipidemia: A population-based study in Taiwan." *Journal of ethnopharmacology* 168:129–135.
- Graybeal, K. B., Gheen, M., and Mckenna, B. 1993. "Clinical pathways development :The Overlake model." *Nursing Management* 24(4):42–45.
- Hong, J. S., Kang, H. C., and Kim, J. K. 2009. "The Variation of HbA1c Examination Performance Rates among Diabetic Patients Using Ambulatory Care in South Korea." *Health Policy and Management* 19(1):49–61.
- Hwang, S. D., Park, K. S., Jeon, B. S., Kim, Y. J., Lee, S. H., Lee, K. S., Yoon, S. Y., and Lee, S. C. 2009. "Clinical Risk Factors for Bacteremia in Patients with Acute Pyelonephritis." *The Korean Journal of Nephrology* 28(5): 18–423.
- Joe, S. R., and Chang, C. G. 2002. "Evaluation of the Items of Questionnaire Based on the Result of Health Screening." *Dongduk journal of life science studies* 7:13–27.
- Jung Daniel, Kim, G.H., Park, J. W., Lee, H. S., Kim, H. J., Choi, H. Y., and Kim, J. H. 2015. "Prediction of Rehospitalization of Patients and Finding Causes of It with Data Mashup and Bigdata Analysis." *Entrue Journal of Information Technology* 14(3):133–149.
- Jung, K. C., Kim, S. G., Kim, S. H., and Lee, M. H. 2015. "Current States and Development Issues of Personalized Medicine based on Personal Genome." *STEPI Insight* 179:1–39.
- Kim, J. S. 2010. "Clinical Outcome and Analysis of Factors Affecting the Prognosis of Barbiturate Coma Therapy." *Journal of the Korean Data Analysis Society* 12(2):763–780.
- Kim, M. H., and Park, H. C. 2008. "Development of Intervening Association Rules and Macro Algorithm." *Journal of the Korean Data Analysis Society* 10(3):1549–1559.
- Kim, M. S., Bae, Y. E., Kim, J. Y., Jo, A. Yoon, H. S., Song, E. J., and Jeong, T. S. 2015. "Standard Clinical Procedure(Critical Pathway) on Dental Conscious Sedation for the Disabled Children." *Journal of The Korean Academy of Pediatric Dentistry* 42(3):226–232.
- Kwak, S. G., and Shin, I. H. 2015. "Clinical data analysis in retrospective study through equality adjustment between groups." *Journal of the Korean data & information science society* 26(6):1317–1325.
- Lakshmi, K. S., Kumar, G., and Santhosh. 2014. "Association Rule Extraction from Medical Transcripts of Diabetic Patients." 2014 Fifth International Conference 201–206.
- Lee, D. G., Ryu, K. S., Bashir, M., Bae, J. W., and Ryu, K. H. 2013. "Discovering Medical Knowledge using Association Rule Mining in Young Adults with Acute Myocardial Infarction." *JOURNAL OF MEDICAL SYSTEMS* 37(2):1–10.
- Lee, K. M., Hwang, K. S., and Kim, W. J. 2010. "A Data Cluster Specialization-based Patient Class Management Strategy for Personalized Medicine." *Proceedings of KIIS Fall Conference* 20(2):323–324.
- Lee, I. H., Shin, A. M., Son, C. S., Park, H. J., Kim, J. H., Park, S. Y., Choi, J. H., and Kim, Y. N. 2010. "Association Analysis of Comorbidity of Cerebral Infarction Using Data Mining." *The journal of Korean society of physical therapy* 22(1):75–81.
- Oh, D. N. 2011. "Comparative Study on Influencing Factors and Digestive Symptoms among Upper Gastrointestinal Disease Groups." *The Journal of the Korea Contents Association* 11(12):335–346.
- Ordonez, C. 2006. "Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction." *IEEE Journal of Biomedical and Health Informatics* 10(2):334–343.
- Park, H. S., Lee, M. S., Hwang, S. J., and Oh, S. Y. 2016. "Data Engineering : TF-IDF Based Association Rule Analysis System for Medical Data." *KIPS Transactions on Software and Data Engineering* 5(3):145–154
- Park, Y. S., Yoon, J. L., and Kim, J. S. 2009. "Evaluation of Evidence-Based Screening Programs for Cancers." *Korean journal of health promotion and disease prevention* 9(2):86–96.
- Yang, H. J., Lee, J. S., Kim, J. S., and Lee, J. K. 2006. "Evaluation of Scientific Evidence for Health Screening Tests Provided by Some Hospitals in Korea." *Korean Journal of Family Medicine* 27:723–732.
- Yoo, K. S., and Jin, S. H. 2016. "A Study on Improvement of Standardized Lift in Association Rules Analysis." *Journal of the Korean Data Analysis Society* 18(1):189–197.