빅데이터를 이용한 만성질환 유무에 따른 삶의 질에 미치는 영향

김민경[1] · 조영복[2]

[1]소노엠 기업부설연구소 · [2]대전대학교

The effect of Quality of Life by chronic disease using Bigdata

Min-kyoung Kim[1] · Young-bok Cho[2]

[1]SONOUM Inc · [2]Daejeon University

E-mail : minkyoungk79@gmail.com / ybcho@dju.ac.kr

요 약

본 연구는 빅데이터 플랫폼을 이용해 만성질환유무에 따른 개인적 요인과 지역사회요인이 삶의 질에 미치는 영향을 알아보는데 목적이 있다. 연구방법은 2017년 지역사회건강조사 자료와 통계청 자료를 보건소 단위로 매칭하였다. 연구결과 남자의 경우 나이가 어릴수록, 도시지역에 거주하는 경우 학력이 높은수록, 월가구소득이 많을수록, 경제활동을 하는 경우, 배우자가 있는 경우 삶의 질이 높았다. 지역사회 요인의 경우 인구밀도가 낮을수록, 고령인구비율이 낮을수록, 의료기관 종사의사주가 많을수록, 재정자주도가 높을수록 삶의 질이 높았다. 지역다음은 요약문 입니다.

ABSTRACT

The purpose of this study is to investigate the effect of personal factors and community factors on the quality of life based on the presence of chronic diseases based on the Big Data Platform.

The research methodology was the matching of the 2017 Community Health Survey data and the National Statistical Office data to the health center units. In the study, The higher the age, the higher the education level, the higher the monthly household income, the economic activity, the spouse, the higher the quality of life. In the case of community factors, the lower the population density, the lower the elderly population ratio, the more doctors engaged in medical institutions, the higher the financial independence, the higher the quality of life.

키워드

Chronic disease, Quality of Life, Multi-level Regression Analysis, Gig Date

Ⅰ. 서 론

4차 산업혁명이 도입되면서 빅데이터를 활용한 융합을 통해 보건·사회·경제 전반에 많은 변화가 나타나고 있다. 빅데이터란 기존 데이터베이스의 데이터 저장·관리·분석 능력을 초과하는 다양한 형식을 가진 대량의 데이터를 의미한다[1]. 다양한 분야에 많은 데이터들이 생산됨에 따라 특히 빅데이터를 활용한 보건의료에 많은 관심을 보이고 있다. 국민의 평균 수명 연장과 만성질환 유병률 증가로 인해 많은 의료비 지출에 대한 문제가 논의되면서 IT와 의료기술을 접목한 U-Health 도입도 추진하고 있다.

우리나라는 급격한 인구의 고령화, 생활습관의 변화, 환경오염 등으로 만성질환이 증가하고 있다. 만성질환은 3개월 이상 지속되는 증상으로 완치가 어려우며 장기간 관리가 필요한 질환이다[2].

2016년 기준 고혈압 환자 5,899천명, 당뇨병 환자 2,704천명을 전년도 대비 각각 3.3%, 7.1% 증가하여 매우 높은 유병률을 보이고 있어[3] 만성질환으로 인한 사망률이 전체 사망의 81%를 차지할 정도로 심각한 상황이고, 사망원인 상위 10위 중 7개가 만성질환이 차지하고 있다[4].

OECD(2010) 보고에 의하면, 만성질환은 전 세계적으로 장애와 사망의 주된 원인으로, 전세계인구의 60%가 만성질환으로 사망하고 있는 것으로 추정된다. 이렇게 만성질환자 증가로 인해 사회경제적 부담 또한 증가하고 있다. 건강보험자료에 의하면 고혈압 및 당뇨병으로 인한 진료비는 건강보험재정의 1, 2위를 차지하고 있다. 고혈압으로 인한 총 진료비는 2008년에 2조 998억원으로 2002년대비 2.5배 증가하였으며, 당뇨병으로 인한 총 진료비는 동 기간에 2.2배 증가한 1조 1,276억원에 달하였다. 만성질환은 대부분 완치되지 않는 경우가 많기 때문에 삶의 질에 영향을 미칠 것으로 본다. 또한 산업들이 융·복합된 많은 양의 데이터들이 생산되면서 삶의 질에 대한 욕구도 증가하고 있다. WHO에서 삶의 질은 한 개인이 살고 있는 문화권과 가치체계의 맥락 안에서 자신의 목표, 기대, 규범, 관심과 관련하여 인생에서 자신이 차지한 상태에 대한 개인적인 지각이라고 정의하였다. 영국의 한 연구에서는 삶의 질은 개인이 가지고 있는 특성과 지역주민이 생활하고 있는 환경적 특성의 상호작용에 의해 영향을 받는다고 하였다[5]. 따라서 삶의 질은 복합적으로 작용하기 때문에 삶의 질을 측정하기 위해서는 객관적, 주관적 지표가 모두 활용하는 것이 정확하다[1]. 기존의 연구들은 개인적 요인 중심으로 연구가 한정되어 왔다. 이에 본 연구는 전국을 대표하는 지역사회건강조사 자료와 통계청 자료를 이용하여 만성질환유무에 따른 삶의 질에 미치는 요인을 분석하고자 한다.

II. 관련연구

2.1 의료 빅데이터 현황

국내는 보건의료 데이터는 공공 영역과 민간 영역에 서 수집되고 있다. 공공 데이터는 보건복지부와 기타 부처가 관할하며 내용과 수집 방식을 고려하여 유전체 데이터, 청구·행정 데이터, 조사데이터로 구분할 수 있다. 민간 영역에서는 의료기관이 환자 진료 과정에서 수집한 임상데이터와 개인의 선택에 의해서 소셜 네트워크 서비스 또는 모바일 장치 등을 통해 수집되는 스트림데이터 등으로 구분될 수 있다. 높은 수준의 IT기술 활용으로 공공과 민간 영역 모두 상당한 수준의 규모와 다양성으로 데이터가 구축되고 있으나 국가적으로 보건의료 빅데이터를 구축하고 활용하는 데는 한계가 있다.

첫째, 기관별로 분산된 보건의료 데이터가 상호연계·통합되어 국가적으로 의미 있는 활용을 유도할 수 있는 법적, 기술적, 정책적 기전이 부족하다. 공공기관 내부의 데이터 통합은 추진되고 있으나 기관 간 데이터 연계는 법적으로 허용된 업무 수행을 위해서만 제한적으로 이뤄지고 있다. 특히 민간영역에서 보유하고 있는 의미 있는 임상데이터가 국가적으로 연계되어 활용되는 제도적·물리적 기반이 구축되어 있지 못하다.

둘째, 국가 단위 빅데이터 구축의 한계는 오픈데이터를 통한 새로운 가치 창출도 제한하고 있다. 새로운 가치를 창출하는 빅데이터는 데이터공개와 접근성 확대를 통해 기대할 수 있다. 개인정보 보호와 데이터 보안이 확인된 데이터는 제한적으로 이용된 일정 기간 이후에 익명화하여 공개함으로써 보건의료시장에서 최종 사용자인 국민의 편익을 높이는 다양한 서비스 상품 개발에 활용되어야 한다.

마지막으로, 유전체 데이터의 활용을 통한 질병발생 기전 등의 임상지식 창출도 제한적이다. 최근 보건의료 분야에서 빅데이터는 보다 근원적인 질병 발생 기전을 분석하기 위한 유전체 데이터의 활용성을 강조하고 있음에도 유전자 정보의 규모 및 내용과 기타 데이터와 연계하는 정보의 완결성 측면에서 활용 가치를 기대하기 어렵다. 질병관리본부, 국립암센터 등이 한국인 표준 게놈(400명) 및 호발질환 유전자 분석(2만여명) 자료, 암 통합오믹스 자료 등을 구축하고 있지만, 외국에 비해 소규모18)이고 기타 정보와의 연계 제한19)으로 정밀의료와 맞춤의료 시대에 대비한 투자와 발전이 필요하다.

2.2 빅데이터 플랫폼

빅데이터 플랫폼은 빅데이터에서 가치를 추출하기 위한 일련의 과정을 지원 하기 위한 프로세스를 규격화한 기술 · 서비스 모임으로 빅데이터의 특성인 Volume과 비정형, 실시간성으로 인해 데이터의 수집과 저장, 분석 등의 모든 영역에서 학문적인 측면보다는 서비스 중심적인 개념으로 표현하는 경향을 보이고 있다. 빅데이터 플랫폼은 데이터에 대해 수집→ 저장 → 처리 → 분석 → 시각화등을 통해 원시데이터(Raw Data)로 부터 Insight 및 가치(Value) 추출하기 위해 분석, 시각화 측면에서는 가치 추출을 위해서 전통적인 통계 분석에서는 인과관계를 최종적인 결과로서 제시한다면, 빅데이터에서는 연관관계 · 상관관계를 중심으로 시사점을 도출하려는 경향을 보이고 있다.

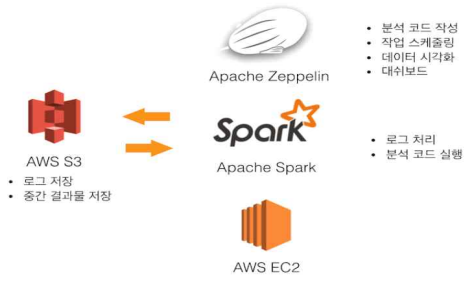

그림 1. 빅데이터 분석 시스템 아키텍처

빅데이터를 다룰 때 가장 많이 쓰는 기술은 Hadoop MapReduce와 연관 기술인 Hive이다. Hadoop은 클러스터 컴퓨팅 프레임웍으로 컴퓨터를 여러 대 연결하면 대수에 따라서 데이터 처리 성

능이 스케일되는 기술이다. MapReduce는 슈퍼컴퓨터 없이도 서버를 여러대 연결하여 빅데이터 분석을 가능하게 해준 혁신적인 기술이다. Apache Spark는 Hadoop MapReduce와 비슷한 목적을 해결하기 위한 클러스터 컴퓨팅 프레임웍으로, 메모리를 활용한 아주 빠른 데이터 처리가 특징이다. 또한, 함수형 프로그래밍이 가능한 언어인 Scala를 사용하여 코드가 매우 간단하며, interactive shell을 사용할 수 있다.

III. 빅데이터 분석을 통한 삶의 질 측정

3.1 데이터 분석 환경

만성질환자의 삶의 질에 미치는 요인을 파악하기 위해 2017년 지역사회건강조사자료를 바탕으로 통계청 자료를 시군구 지역으로 매칭하여 데이터베이스를 구축하였다. 지역사회건강조사 자료는 2008년부터 질병관리본부 주관으로 매년 실시하는 전국 표본조사로 254개 보건소 관할 지역에서 수행된다. 통계청 자료는 우리나라를 대표하는 자료로 국민 전체의 지역특성을 파악하는데 대표적인 자료에 해당된다. 그림 2는 위에서 설명한 지역사회건강조사자료와 통계청 자료 분석을 위해 통합하고 분석기준을 도식화한 것이다.

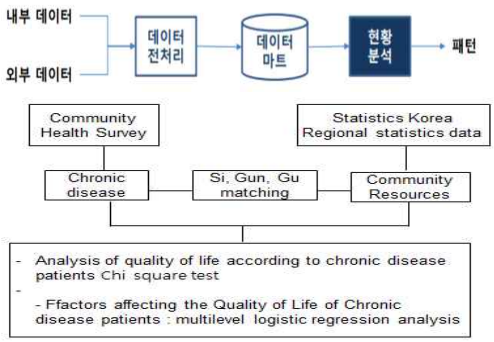

그림 2. 데이터 통합을 위한 필드 선택

3.2 자료분석 환경

삶의 질 측정도구는 EuroQol Group 이 개발한 지표로 EQ-5D를 사용하였다. EQ-5D는 광범위한 건강상태 및 치료의 평가에 이용할 수 있어 국가간, 지역간 비교를 위해 널리 사용되고 있다[6]. EQ-5D는 운동능력(mobility), 자기관리(self-care), 일상활동(usual activity), 통증 및 불편(pain/discomfort), 불안 및 우울(anxiety/depression) 5가지 영역을 EQ-5D index 하나의 값으로 전환하기 위해 한국인을 대상으로 한 질병관리본부의 가중치를 이용하였다[7]. 1 - (0.05 + 0.096*M2 + 0.418*M3 + 0.046*SC2 + 0.136*SC3 + 0.051*UA2 + 0.208*UA3 + 0.037*PD2 + 0.151*PD3 + 0.043*AD2 + 0.158*AD3 + 0.05*N3)

데이터 통합후 분석을 위한 데이터 분석 하둡 환경은 그림3에서와 같다.

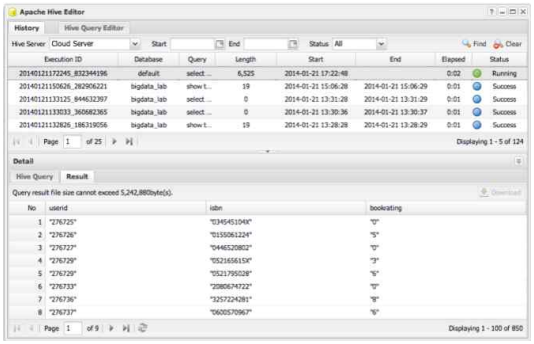

그림 3. 하둡에서 데이터 분석

만성질환 유무에 따라 인구사회학적 특성 차이를 보기 위하여 카이제곱검정을 실행하였고, 삶의 질과 지역사회요인 파악은 상관분석을 이용하여 통계적으로 유의한 영향을 미치는 요인만을 선별하였다. 최종적으로 만성질환 유무에 따른 삶의 질에 영향을 미치는 요인을 파악하기 위하여 다수준 회귀분석(multilevel regression analysis)을 실행하였다. 수집된 자료는 SAS 9.4 통계 프로그램을 이용하여 분석하였다.

level 1: 개인적 요인으로 성별, 연령, 지역, 교육수준, 월 가구소득, 세대타입, 경제활동여부, 배우지 유무

level 2: 지역사회요인으로 인구밀도, 고령인구비율, 의료기관종사 의사 수, 재정자주도, 문화기반 시설 수, 체육시설 수로 다음 식1과 같이 정리할 수 있다.

$$Y=a+b1X1+b2X2+b3X3+b4X4+\cdots+bnXn \quad (1)$$

Y : 종속변수, X : 독립변수, a,b : 회귀계수

IV. 결과

이 논문에서는 빅데이터 플랫폼 spark를 이용해 분산 환경에서 지역사회 건강 조사표와 통계청 자료를 연계한 만성질환 유무에 따른 삶의 질을 분석하였다. 분석 결과 성별에 따라 남자의 경우 삶의 질이 높고 만성질환이 있는 경우 평균 0.91, 없는 경우 0.96으로 만성질환이 없는 경우 삶의 질이 유의하게 높았다. 나이가 적을수록 삶의 질이 높았고, 도시지역에 거주하는 경우, 학력이 높을수록, 수입이 많을수록, 경제활동을 하는 경우, 배우자가 있는 경우 삶의 질이 높았다. 세대유형별로는 2세대가 삶의 질이 가장 높았고, 3세대, 1세대 순이었다.

다수준 분석결과 남자의 경우 여성보다 0.5배 삶의 질이 높았고, 20대가 80대보다 8배 삶의 질이 높았다. 도시지역에 거주하는 경우 0.3배, 무학보다 대학교 이상 교육수준인 경우 6.5배, 월가구소득이 100만원 미만보다 500만원 이상의 경우 2.9배, 경제활동을 하는 경우 3.5배, 배우자가 있

Table. 1 Socio-demographic characteristics of the subject

Factor		without chronic disease		with chronic disease	
		N(%)	M±SD	N(%)	M±SD
Gender	Men	64,913(47.4)	0.96±0.10	36,050(41.8)	0.91±0.15
	Women	72,028(52.6)	0.95±0.10	50,227(58.2)	0.83±0.18
Age	19-29	24,105(17.6)	0.98±0.06	810(0.9)	0.95±0.10
	30-39	30,534(22.3)	0.97±0.06	2,863(3.3)	0.95±0.09
	40-49	34,214(25.0)	0.97±0.07	8,679(10.1)	0.95±0.10
	50-59	25,160(18.4)	0.96±0.09	19,450(22.5)	0.92±0.12
	60-69	11,995(8.8)	0.93±0.11	23,001(26.7)	0.88±0.15
	70+	10,933(8.0)	0.84±0.19	31,474(36.5)	0.78±0.20
Region	urban	85,625(62.5)	0.96±0.08	41,222(47.8)	0.88±0.16
	rural	51,316(37.5)	0.95±0.11	45,055(52.2)	0.84±0.18
Education	Ineducation	6,478(4.7)	0.93±0.19	19,093(22.2)	0.75±0.20
	Elementary school	11,689(8.5)	0.91±0.15	22,142(25.7)	0.84±0.17
	Middle school	11,572(8.5)	0.94±0.11	12,714(14.8)	0.89±0.15
	High school	55,163(40.3)	0.97±0.08	20,079(23.3)	0.92±0.12
	Over college	51,896(37.9)	0.97±0.06	12,140(14.1)	0.95±0.10
Monthly Family Income (unit 1, 000 won)	Under 999won	17,485(12.9)	0.89±0.17	31,682(37.1)	0.79±0.20
	1,000-2,999won	49,100(36.3)	0.96±0.09	29,906(35.0)	0.89±0.15
	3,000-4,999won	43,303(32.0)	0.97±0.07	15,085(17.7)	0.92±0.13
	Over 5,000won	25,284(18.7)	0.97±0.06	8,658(10.1)	0.93±0.12
Type of Household	One generation	40,006(29.2)	0.94±0.12	49,411(57.3)	0.84±0.18
	Two generation	84,000(61.3)	0.97±0.08	29,137(33.8)	0.90±0.15
	Three generation	12,933(9.4)	0.96±0.09	7,724(9.0)	0.85±0.18
Economic activity	Yes	93,436(68.3)	0.97±0.07	44,632(51.8)	0.91±0.12
	No	43,435(31.7)	0.93±0.14	41,551(48.2)	0.80±0.20
Spouse	Yes	92,390(67.5)	0.96±0.09	60,698(70.4)	0.89±0.16
	No	44,439(32.5)	0.95±0.10	25,531(29.6)	0.80±0.19

는 경우 0.1배 삶의 질이 높았다. 지역사회 요인의 경우 인구밀도가 낮을수록, 고령인구비율이 낮을수록, 의료기관종사 의사수가 많을수록, 재정자주도가 높을수록 삶의 질이 높았다.

Acknowledgement

This research was supported by the CHUNGBUK TECHNOPARK, Korea, under the (Development of Prediction and Diagnosis System for Pediatric Adolescents Using Iris based Image Mining) support program (No.20180186)

Ⅳ. 결 론

본 연구는 개인요인과 지역사회요인을 통하여 만성질환여부에 따른 분석데이터는 2014년 지역사회건강조사 자료의 개인적 요인과 통계청의 지역사회요인을 통해 삶의 질에 미치는 영향요인을 알아보고자 하였다. 개인요인은 그 개인이 갖고 있는 유전적 요인, 사회경제적 요인이 건강관련 삶의 질을 결정하는 중요한 변수로 작용하였고, 지역사회 환경과 주민이 거주하고 있는 사회경제적 요인 또한 간접적으로 삶의 질에 영향을 미친 것으로 나타났다. 연구의 제한점으로는 지역사회건강조사가 단면연구이므로 개인적요인과 지역사회요인에 따른 인과관계를 정확히 알기에는 다소 어려움이 있다. 향후 빅데이터의 발전으로 삶의 질을 종단적으로 파악할 수 있는 연구가 필요하다.

References

[1] M. K. Kim, Y. B. Cho,“ An analysis of Factors Affecting Quality of Life through the analysis of Public Health Big Data”, *Journal of the Korea Institute of Information and Communication Engineering*, Vol 22, No.6, pp. 835-841, Jun. 2018.

[2] Korean society and trends:Chronic disease trend and management, pp. 208-215, 2010.

[3] National Health Insurance Corporation 2016 Medical Statistics by Region, 2017.

[4] Korea Centers for Disease Control and Prevention: http://www.cdc.go.kr/CDC/notice/

[5] Bowing, A., Gabriel, Z., Dykes, J., Evans, O., Fleissing, A., Banister, D., Sutton, S., “*Let’s ask them: A national survey of definition of quality of life and its enhancement among people aged 65 and over*”, International Journal of Aging and Human Development, Vol 56, No. 4, pp. 269-306, 2003.

[6] Anna, N., Tosteson, A., “Preference-based health outcome measures on low back pain”, *Spine*, Vol 25, No. 24, pp. 3161-3166, 2000.

[7] South Korean time trade-off values for EQ-5D health states : http://www.cdc.go.kr/CDC/info/