

검진자료를 활용한 건강점수 산출 방법에 관한 연구

이찬중^{1*} · 조인산² · 우현기³ · 임유리⁴ · 박철형⁵

The method of calculating health scores using examination data

Chanjung Lee^{1*} · Brian Jo² · Hyunki Woo³ · Yoori Im⁴ · Chul Hyung Park⁵

^{1*}Member, Data Science Group, EvidNet, Seongnam, 13488 Korea

²CEO, EvidNet, Seongnam, 13488 Korea

³Group Leader, Data Science Group, EvidNet, Seongnam, 13488 Korea

⁴Leader, Data Science Group, EvidNet, Seongnam, 13488 Korea

⁵MD, Department of Biomedical Informatics, Ajou University Hospital, Suwon, 16535 Korea

요 약

본 연구는 국민 건강 검진 결과를 활용하여 건강점수 산출하는 새로운 방법을 제시하고 해당 결과의 타당성을 확인한 연구다. 건강점수(Health Score, 'HS')는 건강 검진 항목에 대한 임상적 기준과 통계적 장치를 적용하여 개인 건강 정도를 점수화한 지표로 특정 검진 항목이 건강변화량에 미치는 전체 크기와 건강변화량에 주는 영향력, 바이오 마커가 정상 범위에서 벗어난 비율에 따른 가중치 변화율, 표준화 값으로 계산된다. 지표에 대한 검증은 선정된 질환에 대하여 질환 발생군과 미발생군 간의 HS를 비교하고, HS 순으로 나눈 구간에서 질환 발생률을 비교하여 타당성 검증을 시도하였다. 그 결과, 모든 선정 질환에서 질환의 유무에 따라 평균 HS 차이는 유의하였고, HS가 감소함에 따라 선정 질환의 발생률이 증가하는 경향을 보였다.

ABSTRACT

This study confirmed the validity of the health score(HS) calculation methodology and results. HS is an index that scores the degree of personal health by applying clinical standards and statistical function to health check-up items. It's calculated by the total size of the biomarker on the health change, the influence on the health change, the weight change rate according to the degree of outlier, and the standardized value. In order to confirm the validity of the results, several diseases were selected and HS was compared between the disease occurrence group and the disease absence group. And by segmenting the ranked HS, the disease incidence rate was compared. As a result, in all selected diseases, the difference in the mean of HS was significant according to the presence or absence of disease, and the incidence of selected diseases showed a tendency to increase as HS decreased.

키워드 : 건강 점수, 건강변화량, 건강 검진, 방법론

Keywords : Health Score, Health change, Health examination, Methodology

Received 19 October 2022, Revised 4 November 2022, Accepted 19 November 2022

* Corresponding Author Chanjung Lee(E-mail: lcjpaul@evidnet.co.kr, Tel:+82-31-605-5331)

Member, Data Science Group, EvidNet, Seongnam, 13488 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.12.1777>

print ISSN: 2234-4772 online ISSN: 2288-4165

© 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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

OECD 보건통계(Health Statistics) 주요 결과에 따르면, 2000년에서 2019년까지 약 20년간 우리나라 국민 기대수명은 7.3세 증가(76.0세 → 83.3세)하였고, 건강수명 역시 5.7세(67.4세 → 73.1세) 증가하였다[1]. 국민수명이 늘어남에 따라 의학적 기조는 질병 치료와 예후 관리에서 질병 예방과 건강 관리로 옮겨갔고, 현재 본인의 건강 상태를 확인하고 건강한 삶을 오랫동안 유지하고자 하는 욕구도 커졌다. 이에 건강 척도로 흔히 쓰였던 “사회적 나이(Chronological Age, ‘CA’)” 대신 실제 본인의 건강 정도를 보여주는 건강 지표를 개발하려는 연구가 꾸준히 시도되었다[2-4].

그 중 가장 활발히 연구되는 건강 지표는 “생물학적인 나이(Biological age, ‘BA’)”다[5-7]. BA는 선천적 요인, 환경적 요인 및 노화 상태가 반영된 건강 지표로 다양한 산출 방법과 검증, 활용 가능성에 대한 연구가 일반적이다[8-9]. BA 추정 연구로는 건강 일반인 대상자 자료를 활용하여 BA를 추정하는 다중 선형 회귀(Multi linear Regression, MLR) 모형 연구가 중점적이었으며[10-12], PCA(Principal component analysis)를 적용하여 BA를 추정하는 연구도 나오고 있다[12-13]. 그 외 건강 지표로 ‘문동주’는 “건강 지수(Health Index)”는 건강 상태의 다양한 측면에 따른 지수를 산출하고자 심혈관 지수, 스트레스 지수, 비만 지수, 관리지수의 평균값으로 정의하고 [14], ‘오필재’는 건강상태를 나타낼 수 있는 지표인 “건강 사고(health accident)”를 정의하고, 건강 사고 발생 가능성을 추정할 수 있는 모형으로 점수를 환산한 값인 “건강 수준”을 건강 상태 측정 지표로 제시한 바 있다[15].

본 연구에서는 연령과 검진 수치를 활용하여 건강 노화 지표로 “건강 점수(Health Score)”를 산출하는 방법을 제시한다. 건강점수 산출은 건강 검진 수치와 연령과의 영향 정도를 밝히고, 해당 항목별 임상적 기준을 반영한 수치값을 표준화시켜 얻게 된다. 산출 결과는 선행연구[16]를 토대로 건강점수와 질환 발생과의 관계의 통계적 유의성을 검증한다. 본 연구의 차별성은 ‘건강정도’를 단일한 통계적 모형이나 주관적 정의에 근거하여 해석하고자 했던 기존 연구와 달리 개인 검진 수치와 연령간 영향정도를 밝히고 해당 검진 수치에 임상적 기준을 적용하여 건강 지표화한 방법론이라는 점에 있다. 실제 의료 빅데이터로부터 개발 및 검증이 이루어져 임상

적 객관성과 통계적 신뢰성을 획득하였다는 점 역시 본 연구 타당성을 입증한다.

II. 방법론

건강점수는 개인 검진 결과로부터 임상적 근거와 통계적 장치를 통해 본인의 건강 정도(건강변화량)를 상대적 점수로 나타낸 값이다. 건강점수 산출 방법은 다음과 같다. 먼저 연령과 높은 상관성을 보이는 검진 항목을 찾고 해당 항목이 연령에 미치는 정도를 건강변화량 가중치로 설정한다. 가중치는 개인별 검진 수치가 정상 범위에서 벗어난 비율만큼 보정되어 개인 검진 값에 곱해진다. 이렇게 산출된 값은 개인 건강 정도를 나타내는 수치인 “건강변화량”으로 정의되고, 건강변화량은 전체 연구 집단 내에서 min-max 정규화되어 개인 건강점수로 환산된다.

2.1. 데이터

본 연구는 1997년부터 2020년까지 아주대학교 병원 CDM에서 건강 검진을 받은 내원자 500,028명을 대상으로 검진일 기준 과거 1년 이내에 다음 기본 검진 항목이 있는 20-79세 성인남녀 55,929명을 연구대상자로 추출하였다. 연구에 활용한 변수는 연령과 성별을 포함하여 국내 일반 검진 항목 16개로 선정하였다.

Table. 1 Features Code

Feature	ID	name
Albumin	3024561	Albumin [Mass/volume] in Serum or Plasma
ALT	3006923	Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma
AST	3013721	Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma
BMI	3038553	Body mass index (BMI) [Ratio]
BUN	3013682	Urea nitrogen [Mass/volume] in Serum or Plasma
Cholesterol	3027114	Cholesterol [Mass/volume] in Serum or Plasma
Creatinine	3016723	Creatinine [Mass/volume] in Serum or Plasma
DBP	3012888	Diastolic blood pressure

Feature	ID	name
Glucose Fasting	3037110	Fasting glucose [Mass/volume] in Serum or Plasma
Hb	3027484	Hemoglobin [Mass/volume] in Blood by calculation
HDL	3007070	Cholesterol in HDL [Mass/volume] in Serum or Plasma
LDL	3028437	Cholesterol in LDL [Mass/volume] in Serum or Plasma
SBP	3004249	Systolic blood pressure
Triglyceride	3022192	Triglyceride [Mass/volume] in Serum or Plasma

2.2. 변수와 가중치

일반적으로 검진 수치들은 성별과 연령에 따라 증감하는 경향을 보인다. 본 연구에서는 성별에 따라 연령과 검진 수치 간 상관관계수가 0.7 이상인 항목을 표2. 와 같이 선정하고 해당 항목들이 연령에 미치는 가중치를 표3. 에 제시하였다.

Table. 2 The Selected Features

Sex	Features
Female(8)	Albumin, BUN, AST, SBP, Creatinine, HDL, Triglyceride, Glucose_Fasting
Male(6)	Albumin, Glucose_Fasting, BUN, SBP, HDL, Hb

산출된 항목에 대한 검증은 연령에 대하여 표2. 의 변수를 독립변수로 하는 회귀모형을 적합한 결과, 남성은 $RMSE \leq 1.5, R^2 \geq 0.98$, 여성은 $RMSE \leq 1.6, R^2 \geq 0.99$ 로 높은 설명력을 보였다.

Table. 3 Weights(β) of Features

Feature	Female	Male
Albumin	-0.377	-0.608
AST	0.355	
BUN	0.615	0.254
Creatinine	0.020	
Glucose_Fasting	0.379	0.507
Hb		-0.364
HDL	-0.278	-0.203
SBP	0.344	0.366
Triglyceride	0.114	

검진 항목에 대하여 연령을 추정하는 일반화 식은 수

식(1) 과 같다. \hat{Y} 은 연령 추정치, Y 은 연령, X_i 은 i 번째 검진 결과값, β 는 i 번째 검진 항목 가중치($Weight_i$), α 는 절편, E 는 오차다.

$$Y = \sum_{i=1}^n \beta_i X_i + \alpha + E \rightarrow \hat{Y} = \sum_{i=1}^n \beta_i X_i + \alpha \quad (1)$$

다만, 위 식은 단일한 수치를 제공하지만 해당값은 개인 건강 정도나 임상적 근거에 대한 어떠한 정보도 담고 있지 않다. 예컨대, SBP(연령과 양의 상관관계)가 정상치보다 높고 AST(연령과 양의 상관관계)가 정상치보다 낮은 남성의 경우, 고혈압이 있음에도 수치상 SBP가 높은 만큼 낮은 AST가 상쇄되어 어떠한 건강 위험 정보도 주기 어려울 것이다. 따라서 검진 수치를 수리적인 계산을 넘어 임상적 관점에서 건강 기준이 고려된 수식이 필요하다.

2.3. 정상 범위와 이상치 비율

질병관리청 국가건강정보포털에서는 각 검진 항목에 대한 정상/위험 범위를 제공한다[17]. 위험 범위는 관측치가 위험 최소값(outlier range min)보다 작을 경우 위험 최소값으로 치환되고 위험 최대값(outlier range max)보다 클 경우, 최대값으로 치환되도록 하였다. 한편, 일정 구간까지는 검진 수치가 줄어들거나 증가하더라도 건강 정도에 이상이 없지만 검진 수치가 위험 구간에 속한다면 해당 대상자에게 가중치를 더 크게 적용하여 건강변화량을 계산하는 장치가 필요하므로 건강 점수 산출 방법은 검진 수치가 정상치에서 벗어난 만큼 해당 검진 항목 가중치가 보정되도록 설계하였다.

$Rate_i$ 는 Normal Range(NR) 과 Outlier Range (OR [min|max])과의 최대거리를 1로 두고 수치가 해당 구간 내에 위치한 비율을 구하여 산출한다. 즉, $Rate_i$ 는 X_i 값 위치에 따라 다음 3가지 경우로 계산된다.

첫 번째, $NR_{min} \leq X_i \leq NR_{max}$ 은 $f(X_i) = 0$ 이다.

두 번째, $NR_{max} \leq X_i$ 경우에는 수식 (2)와 같으며, 개념적 내용은 그림1. 을 참고할 수 있다.

$$f(X_i) = \frac{X_i - NR_{max}}{OR_{max} - NR_{max}} \quad (2)$$

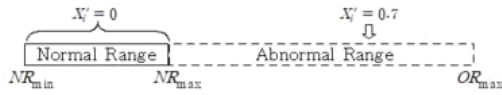


Fig. 1 Output Concept: $NR_{max} \leq X_i$

세 번째, $X_i \leq NR_{min}$ 경우에는 수식 (3)와 같으며, 개념적 내용은 그림2. 를 참고할 수 있다.

$$f(X_i) = \frac{NR_{min} - X_i}{NR_{min} - OR_{min}} \quad (3)$$



Fig. 2 Output Concept: $X_i \leq NR_{min}$

이때, $Weight_i > 0$ 변수는 가중치를 증가시키는 변수로 수식 (4)을 활용한다.

$$Rate_i = 1 + f(X_i) \quad (4)$$

$Weight_i < 0$ 변수는 가중치를 감소시키는 변수이다.

$$Rate_i = 1 - f(X_i) \quad (5)$$

표준화에 사용하는 평균과 표준편차는 변수값 전체 또는 집단별 평균과 표준편차를 사용하여 변환한다.

$$Z_i = \frac{X_i - \bar{X}}{\sigma(X)} \quad (6)$$

2.4. 건강 변화량

건강변화량은 산출은 다음의 수식(7)을 따른다.

$$\text{건강변화량} = \rho \times \sum_{i=1}^n Weight_i \times Rate_i \times Z_i \quad (7)$$

ρ 은 검진변수가 건강변화량 변화 크기(가중치)를 결정하는 값이고, $Weight_i$ 은 i 번째 검진변수의 가중치, $Rate_i$ 은 검진변수가 정상 범위로부터 벗어난 정도를 비율로 표현하여 $Weight_i$ 를 조정하는 값이며, Z_i 은 i 번째 검진변수의 표준화 값이다.

2.5. 건강점수(Health Score, HS)

최종적으로 계산되는 HS는 min-max 정규화 방식을 활용하여 수식(8)과 같이 산출되고 0 ~ 100까지의 범위를 가진다.

$$HS = \left(1 - \frac{\text{건강변화량} - \text{건강변화량}_{min}}{\text{건강변화량}_{max} - \text{건강변화량}_{min}} \right) \times \text{최대점수} \quad (8)$$

III. 검증

3.1. 기술통계

아주대병원 수검자의 검진 항목별 기술 통계표는 표 4. 와 같다.

Table. 4 Descriptive statistics

Features	Total(N=55,929)	Female(N=22,923)	Male(N=33,006)
Stats	Mean±sd	Mean±sd	Mean±sd
Age	45.66±14.26	47.27±15.22	44.53±13.45
Albumin	4.62±0.32	4.53±0.3	4.69±0.32
ALT	24.32±14.74	18.56±11.37	28.31±15.47
AST	23.68±9.15	21.61±8.26	25.11±9.46
BMI	23.87±2.97	22.85±3.01	24.58±2.72
BUN	13.23±4.11	12.63±4.24	13.65±3.97
Cholesterol	183.73±34.96	183.52±35.16	183.88±34.82
Creatinine	0.91±0.21	0.76±0.17	1.01±0.17
DBP	75.65±9.33	72.89±9.26	77.57±8.89
Glucose_Fasting	99.27±19.89	97.03±19.1	100.83±20.28
Hb	14.26±1.5	13.04±1.1	15.11±1.1
HDL	55.18±13.45	61.03±13.71	51.11±11.65
LDL	105.24±29.47	103.44±29.19	106.48±29.61
SBP	123.51±13.44	121.12±14.69	125.16±12.23
Triglyceride	123.92±70.71	103.26±56.68	138.26±75.76

여성의 평균 HS는 87.5(±2.7), 남성의 경우 84.5(±3.1)로 여성이 남성에 비해 평균 HS가 3점 높았으며, 표준편차도 0.4점 더 적은 것으로 나타났다.

그림3. 을 보면, 동일 연령대의 여성이 남성에 비해 평균 HS가 높고, HS 최소값과 최대값 모두 높음을 알 수 있다. 이는 평균적으로 동일 연령대 여성이 남성에 비해 바이오마커가 정상 범위에 가까이 분포한다는 것을 의미하며 이러한 현상은 모든 연령대에서 나타났다.

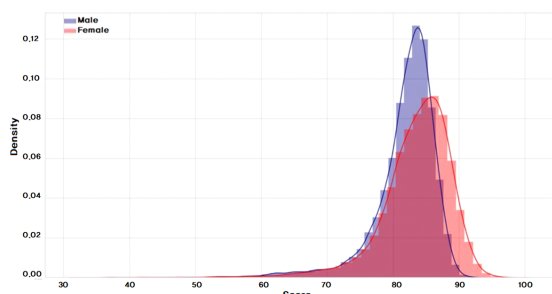


Fig. 3 Histogram: HS Distribution by sex

3.2. 타당성 검토

HS의 타당성을 확인하기 위해 질환별 평균 HS를 구하고 평균 HS가 낮을수록 질환 발생률이 높는지 확인하였다. 관심 질환은 만성질환 4가지(고혈압 [Hypertension, 'HT'], 고지혈증 [Hyperlipidemia, 'HL'], 당뇨병 [Diabetes, 'DM'], 심뇌혈관질환 [Cardiovascular-Cerebrovascular Disease, 'CVD'])와 검진 항목과 관련된 주요 장기 질환 3가지(간[Liver Disease, 'LiD'], 신장 [Kidney Disease, 'KD'], 폐[Lung Disease, 'LuD'])를 선정하였으며, OMOP-CDM 코드로 표5. 과 같이 정의하였다. 관심 질환 대상자는 검진일 이전 1년에서 검진일 이후 1년까지 기록이 존재하는 경우 대상자로 포함하였다.

Table. 5 OMOP-CDM Code

No	Disease	Code
1	LiD (22)	192240,193256,194984,194990,196463,196625,198964,200762,200763,201612,201901,439674,4015842,4059290,4059297,4064161,4211974,4245975,4296554,4340383,4340385,4247079
2	HL(4)	432867,437530,437827,438720
3	HT(5)	319826,320128,443771,4028741,4289933
4	DM(9)	201820,201826,440360,443238,443412,400857,6,4099216,4193704,40482801
5	KD(4)	193782,443597,443611,46271022
6	CVD (56)	134057,198177,312327,314054,314059,314658,314666,315286,315558,315564,316139,316437,317576,317585,318443,319835,320739,320744,321042,321318,373503,378774,381591,439846,440417,443454,443962,763015,4028073,4049659,4057008,4068155,4089462,4091901,4103295,4108814,4110185,4124683,4124693,4141360,4148906,4153380,4154290,4163710,4189343,4232697,4270024,4275423,4332246,4353709,37016924,40479625,42535425,43530683,43530727,44784217

No	Disease	Code
7	LuD (15)	253954,255573,255848,256622,256722,257004,257315,257907,261325,261880,4078925,4119786,4187218,4197819

3.2.1. 질환별 건강점수 비교

표6.과 표7. 은 정의한 관심 질환 대상자의 평균 HS를 성별에 따라 비교한 결과다. 여성 에서는 질환-미발생군이 평균 HS가 86.18점으로 가장 높았으며, 다음으로 고지혈증, 간 질환, 폐 질환, 심뇌혈관 질환, 고혈압, 당뇨병, 신장 질환 순으로 나타났다. 질환-미발생군과 각 질환군 간 평균 HS 비교 결과(independent T-test), 유의수준 1% 하에서 질환-미발생군의 평균 HS가 나머지 모든 질환군 평균 HS 보다 작은 것으로 나타났다.

Table. 6 The mean of HS by Disease: Female

No	Disease	N	Proportion (%)	Health Score				p-value
				Mean	SD	95% LB	95% UB	
1	None	9,144	40%	86.18	3.79	86.1	86.26	-
2	LiD	1,441	6%	79.64	6.95	79.28	80	<0.001
3	HL	2,401	10%	80.13	5.43	79.91	80.35	<0.001
4	HT	2,351	10%	77.71	7.26	77.42	78.01	<0.001
5	DM	1,368	6%	75.76	7.73	75.35	76.17	<0.001
6	KD	241	1%	66.6	13.24	64.93	68.27	<0.001
7	CVD	2,104	9%	78.23	7.19	77.93	78.54	<0.001
8	LiD	412	2%	78.94	8.4	78.13	79.75	<0.001

Female: HS T-test between None-Disease group and Each Disease group

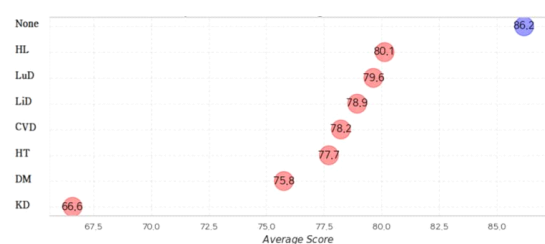


Fig. 4 The mean of HS by Disease: Female

남성의 경우, 질환-미발생군의 평균 HS는 83.43점으로 가장 높았으며, 다음으로 간 질환, 고지혈증, 고혈압, 심뇌혈관 질환, 폐 질환, 당뇨병, 신장 질환 순으로 나타났다 결과는 표7. 과 같다. 질환-미발생군과 각 질환군 간 평균 HS 비교 결과(independent T-test), 유의수준 1%

하에서 질환-미발생군의 평균 HS가 나머지 모든 질환군 평균 HS 보다 작은 것으로 나타났다.

Table. 7 The mean of HS by Disease: Male

No	Disease	N	Proportion (%)	Health Score				p-value
				Mean	SD	95% LB	95% UB	
1	None	19,305	58%	83.43	3.21	83.38	83.47	-
2	LiD	2,632	8%	79.41	5.65	79.2	79.63	<0.001
3	HL	2,437	7%	78.95	5.71	78.72	79.17	<0.001
4	HT	3,158	10%	76.65	6.54	76.42	76.87	<0.001
5	DM	2,063	6%	73.69	7.44	73.37	74.01	<0.001
6	KD	320	1%	68.8	9.48	67.76	69.84	<0.001
7	CVD	3,794	11%	76.61	6.84	76.39	76.83	<0.001
8	LiD	726	2%	75.72	7.81	75.15	76.29	<0.001

Male: HS T-test between None-Disease group and Each Disease group



Fig. 5 The mean of HS by Disease: Male

3.2.2. 건강점수 구간별 발생률 비교

그림6.에서 여성의 질환 발생 비율을 확인한 결과, HS등급이 올라갈수록 질환 발병률은 감소하고, 질환 발생자수는 28%에서 90%까지 증가함을 확인할 수 있다.

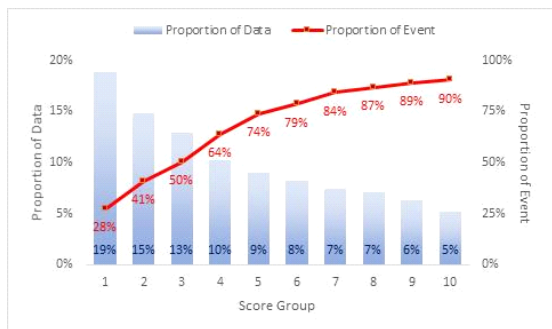


Fig. 6 The Incidence rate by HS level: Female

그림7.에서 남성의 질환 발생 비율을 확인한 결과, HS등급이 올라갈수록 질환 발병률은 감소하고, 질환 발생자수는 19%에서 86%까지 증가함을 확인할 수 있다.

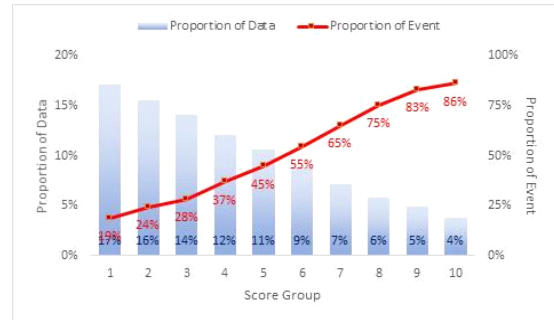


Fig. 7 The Incidence rate by HS level: Male

3.2.3. 건강점수 등급별 관심 질환 발생률 비교

표8.과 표9.는 여성과 남성에 대하여 관심 질환 발병률을 확인한 결과다. HS 등급이 높을수록 질환 발생률이 높은 것으로 나타났고 신장질환은 다른 질환에 비해 상대적으로 HS 등급에 따른 질환 발생률 변화가 적은 것으로 나타났다. 남녀 간 편차는 있지만 여성은 남성에 비해 HS 등급이 높을수록 고지혈증과 심뇌혈관질환 발생률이 높았고 남성은 여성에 비해 간질환, 고혈압, 당뇨에 대한 발생률이 높은 것으로 나타났다.

Table. 8 The disease incidence rates by HS level : Female

HS level	LiD	HL	HT	DM	KD	CVD	LuD
1	1.0%	1.4%	0.5%	0.2%	0.0%	0.9%	0.5%
2	2.7%	2.7%	1.2%	0.3%	0.0%	1.5%	0.8%
3	3.9%	5.9%	3.1%	1.0%	0.1%	3.2%	0.9%
4	6.7%	10.0%	6.4%	2.0%	0.0%	7.4%	1.5%
5	7.6%	14.1%	9.7%	3.6%	0.5%	9.5%	1.9%
6	7.3%	16.5%	13.7%	6.2%	0.4%	12.2%	2.8%
7	9.5%	19.7%	18.9%	9.6%	1.1%	14.7%	2.9%
8	10.7%	21.3%	25.4%	13.9%	0.9%	20.5%	2.5%
9	13.1%	21.8%	28.4%	20.4%	2.6%	25.0%	3.3%
10	17.8%	20.3%	37.8%	33.8%	12.2%	31.7%	6.0%

Table. 9 The disease incidence rates by HS level : Male

HS level	LiD	HL	HT	DM	KD	CVD	LuD
1	2.4%	2.3%	0.9%	0.1%	0.0%	1.3%	0.3%
2	4.4%	3.8%	2.0%	0.5%	0.0%	2.1%	0.6%
3	5.7%	4.4%	3.3%	0.8%	0.1%	4.5%	0.8%
4	7.8%	6.0%	5.8%	1.8%	0.2%	6.3%	1.4%
5	9.1%	7.7%	7.4%	2.6%	0.2%	9.5%	1.5%
6	11.9%	10.3%	12.5%	5.9%	0.6%	16.6%	2.3%
7	11.7%	13.0%	19.7%	11.1%	1.1%	24.8%	3.3%
8	14.0%	13.7%	25.8%	17.8%	1.7%	26.3%	6.4%
9	17.0%	17.4%	32.4%	29.0%	2.8%	38.4%	7.6%
10	16.5%	19.3%	41.1%	46.8%	14.3%	49.7%	11.9%

IV. 결 론

본 건강 점수(HS)는 CA에 영향을 미치는 검진항목의 가중치를 구하고 해당 검진 수치가 정상범위로부터 벗어난 정도에 따라 가중치가 보정된 값으로 산출된다. 해당 방법론은 CA군(5 세단위)과 CA군별 평균 HS 간에 음의 상관관계가 있음을 확인하고 선정 질환에 대하여 질환 발생군이 미발생군에 비해 평균 HS가 유의하게 높음을 검증함으로써 그 타당성을 입증하였다.

본 연구는 검진결과 값을 임상적 범위 구분없이 단순 수치로 간주하여 건강지표를 산출하였던 기존 시도들과는 다르게, 건강 점수 산출에 활용할 항목을 통계적인 검증을 통해 선별하고 검진 정상/비정상 범위와 같은 임상적인 개념을 수식화한 연구라는 점에서 그 차별성을 가진다. 또한 산출된 결과를 실제 의료 데이터에서 얻은 질환 발병과의 관계를 확인함으로써, 그 타당성을 평가하였다는 점에서 의의가 있다.

해당 건강 점수 산출 방법은 현재 본인의 건강 상태를 반영한 종합적인 건강 지표와 함께 질환 예방에 참고할 수 있는 지표로도 활용되리라 기대한다. 추후 해당 방법론과 상관성을 보이는 질환을 찾거나 검진 항목(바이오마커)과 샘플 사이즈를 확장한다면 보다 정밀한 모형으로 개발될 수 있을 것이다.

REFERENCES

[1] OECD Health Statistics 2019 Definitions, Sources and Methods. (OECD) OECD Health Statistics 2019 Definitions, Sources and Methods [Internet]. Available: http://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_STAT.

[2] T. Sentell, W. Zhang, J. Davis, K. K. Baker, and K. L. Braun, "The Influence of Community and Individual Health Literacy on Self-Reported Health Status," *Journal of General Internal Medicine*, vol. 29, no. 2, pp. 298-304, Feb. 2014.

[3] S. Schütte, J. -F. Chastang, A. Parent-Thirion, G. Vermeulen, and I. Niedhammer, "Association between socio-demographic, psychosocial, material and occupational factors and self-reported health among workers in Europe," *Journal of Public Health*, vol. 36, no. 2, pp. 194-204, Jun. 2014.

[4] C. Schmeer, A. Kretz, D. Wengerodt, M. Stojiljkovic, and O. W. Witte, "Dissecting Aging and Senescence—Current Concepts and Open Lessons," *Cells*, vol. 8, no. 11, pp. 1446, Nov. 2019.

[5] L. Sagers, L. Melas-Kyriazi, C. J. Patel, and A. K. Manrai, "Prediction of chronological and biological age from laboratory data," *Aging (Albany NY)*, vol. 12, no. 9, pp. 7626-7638, May 2020.

[6] J. Yoo, Y. Kim, E. R. Cho, and S. H. Jee, "Biological age as a useful index to predict seventeen-year survival and mortality in Koreans," *Bio Med Central Geriatrics*, vol. 17, no. 1, pp. 7, Jan. 2017.

[7] K. L. S. Husted, M. Fogelstrøm, P. Hulst, A. Brink-Kjaer, K. Å. Henneberg, H. B. D. Sorensen, F. Dela, and J. W. Helge. "A Biological Age Model Designed for Health Promotion Interventions: Protocol for an Interdisciplinary Study for Model Development," *JMIR Research Protocol*, vol. 9, no. 10, pp. e19209, Oct. 2020.

[8] X. Li, "Biological ages : correlations, genetic determinants, and health outcomes," Ph. D. dissertation, Karolinska Institute, Dept of Medical Epidemiology and Biostatistics, MA, 2021.

[9] G. A. Borkan and A. H. Norris, "Assessment of biological age using a profile of physical parameters," *Journal of Gerontology*, vol. 35, no. 2, pp. 177-184, Mar. 1980.

[10] M. E. Levine and E. M. Crimmins, "Is 60 the New 50? Examining Changes in Biological Age Over the Past Two Decades," *Demography*, vol. 55 no. 2, pp. 387-402, Apr. 2018

[11] T. L. Dubina, V. A. Dyundikova, and E. V. Zhuk,

“Biological age and its estimation. II. Assessment of biological age of albino rats by multiple regression analysis,” *Experimental Gerontology*, vol. 18, no. 1, pp. 5-18, 1983.

- [12] L. Jia, W Zhang, and X. Chen, “Common methods of biological age estimation,” *Clin Interv Aging*, vol. 12, pp. 759-772, May 2017.
- [13] J. Park, B. Cho, H. Kwon, and C. Lee, “Developing a biological age assessment equation using principal component analysis and clinical biomarkers of aging in Korean men,” *Archives of Gerontology and Geriatrics*, vol. 49, no. 1, pp. 7-12, Jul-Aug. 2009.
- [14] D. Moon, K. Kim, H. Kang, Y. Yang, S. Lee, and J. Lee. “A Suggestion on Evaluating Personal Health State: Health Index,” *Journal of biomedical Engineering Research*, vol. 29, no. 5, pp. 400-407, Oct. 2008.
- [15] P. Oh, H. Kim, and H. Kwon, “A method for evaluating and scoring of health status,” *The Korean journal of applied statistics*, vol. 33, no.3, pp. 239-256, Jun. 2020.
- [16] Y. G. Kang, E. Suh, J. W. Lee, D. W. Kim, K. H. Cho, and C. Y. Bae, “Biological age as a health index for mortality and major age-related disease incidence in Koreans: National Health Insurance Service - Health screening 11-year follow-up study,” *Clinical Interventions in Aging*, vol. 13, pp. 429-436, Mar. 2018.
- [17] The Korean Academy of Medical Sciences. The National Health Information Portal [Internet]. Available: <https://health.kdca.go.kr/healthinfo/biz/health/gnrlzHealthInfo/gnrlzHealthInfo/gnrlzHealthInfoMain.do>.

Supplementary table

Table. 1 HS statistics : Female

CA	N	Mean	SD	Quartile					
				min	20%	40%	60%	80%	max
[20, 25)	684	87.5	2.7	71.9	85.7	87	88.3	89.7	96
[25, 30)	3,187	87.6	2.9	60.6	85.8	87.2	88.3	89.8	98.7
[30, 35)	2,563	87.1	3.2	59.8	85.1	86.7	87.9	89.5	97.1
[35, 40)	1,618	86.2	3.4	57.8	84.3	85.7	87.1	88.6	97.5
[40, 45)	2,166	85.2	3.6	47.8	83.1	84.8	86.1	87.6	96.1
[45, 50)	2,162	83.7	4.8	42.1	81.6	83.6	85.1	86.8	96.8
[50, 55)	2,501	82.3	4.9	37.2	80.2	82.3	83.9	85.6	93.7
[55, 60)	2,316	81.5	4.6	43.3	79.2	81.3	82.8	84.6	93.3
[60, 65)	2,073	80.1	5.6	37.2	77.8	80.3	82	83.7	93.2
[65, 70)	1,740	79.1	5.6	33.3	76.7	79.4	81.1	82.8	91.1
[70, 75)	1,230	77.8	6	37.1	75	78.3	80.1	82	89.4
[75, 80)	683	77.2	5.7	36.2	74	77.4	79.4	81.1	88.3

Table. 2 HS statistics : Male

CA	N	Mean	SD	Quartile					
				min	20%	40%	60%	80%	max
[20, 25)	275	84.5	3.1	66.5	81.3	82.8	83.5	84.3	85
[25, 30)	3,701	84.8	2.5	64.4	81.9	83.1	83.9	84.5	85
[30, 35)	5,525	84.1	2.7	60.9	81.2	82.4	83.2	83.8	84.3
[35, 40)	5,192	83.4	3	53.1	80.4	81.6	82.5	83.1	83.7
[40, 45)	3,645	82.6	3.7	52.2	79.2	80.8	81.7	82.4	83
[45, 50)	3,198	81.6	3.9	52.3	77.4	79.6	80.7	81.5	82.2
[50, 55)	3,414	80.8	4.6	46.3	76.1	78.6	80	80.9	81.6
[55, 60)	2,811	79.9	4.9	52.8	74.2	77.4	79	80	80.8
[60, 65)	1,963	78.4	5.5	42.1	72	75.9	77.3	78.5	79.5
[65, 70)	1,495	77.2	5.7	47.9	70	74.4	76.3	77.4	78.4
[70, 75)	1,077	75.5	6.2	32.5	66.8	72.4	74.5	75.9	76.9
[75, 80)	710	74.7	6.1	46	66	70.9	73.3	74.8	76.2



이찬중(Chanjung Lee)

에비드넷 데이터사이언스그룹 매니저
※관심분야: 헬스케어, CDM, 의료 통계, AI



조인산(Brian Jo)

에비드넷 대표이사
※관심분야: 보건의료 빅데이터, 의료 데이터분석, 임상 역학, AI



우현기(Hyunki Woo)

에비드넷 데이터사이언스그룹 그룹장
※관심분야: 보건의료 빅데이터, 의료 데이터분석, 임상 역학, AI



임유리(Yoori Im)

에비드넷 데이터사이언스그룹 리더
※관심분야: 헬스케어, 의료 데이터분석, 임상 역학



박철형(Chul Hyung Park)

아주대학교 의과대학 졸업
아주대학교 의과대학 의료정보학교실 Ph.D. course
※관심분야: 헬스케어, CDM, 빅데이터, 의료통계, AI