

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/316232010>

Development of a Model to Estimate Body Fat Percentage using Decision-Tree Analysis

Article · November 2016

CITATIONS

0

READS

81

4 authors, including:



Junbae Mun

Middle Tennessee State University

19 PUBLICATIONS 10 CITATIONS

[SEE PROFILE](#)



Heontae Kim

University of Mississippi

28 PUBLICATIONS 24 CITATIONS

[SEE PROFILE](#)



Minsoo Kang

University of Mississippi

212 PUBLICATIONS 2,581 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Health Psychology [View project](#)



Feasibility of theory-based sedentary behavior intervention [View project](#)

의사결정나무 분석을 이용한 체지방률 추정모형 개발

Development of a Model to Estimate Body Fat Percentage using Decision-Tree Analysis

문준배 · 김헌태 · 류승호 · 강민수* 미들테네시주립대학교

Junbae Mun · Heontae Kim · Seungho Ryu · Minsoo Kang Middle Tennessee State Univ.

요약

체지방률을 추정하는 방법으로 체지방률 추정모형이 널리 사용되고 있다. 그러나 대부분의 체지방률 추정모형은 복잡한 추정식을 포함하고 있어 일상생활에서 사용하기가 쉽지 않다. 따라서, 본 연구의 목적은 의사결정나무 분석방법을 이용하여 누구나 쉽게 사용할 수 있는 체지방률 추정모형을 개발 및 검증하는 것이었다. 이를 위해 국민건강영양조사 2009년도 데이터($n = 7,161$)를 추정모형 개발을 위한 데이터(80%)와 검증을 위한 데이터(20%)로 나누었다. 개발 단계에서 이중에너지 방사선 측정법(DEXA)으로 측정된 체지방률과 인체계측치 및 건강관련 행태 변인을 이용하여 다수의 추정모형을 개발하였고, 그중 가장 간결하고 낮은 추정 위험도를 가진 모형을 선택하였다. 선택된 추정모형의 정확성은 R^2 과 평균 제곱근 오차(RMSE)에 의해 평가되었다. 선택된 추정모형은 4개의 변인(성별, BMI, 허리둘레, 나이)을 포함하였고, 총 21개의 끝마디와 5단계 깊이를 가졌다(추정 위험도 = 13.98). 검증 단계에서 추정된 체지방률은 DEXA에 의해 측정된 체지방률과 비교시 73%의 설명력과 4.03%의 평균 제곱근 오차를 보였다. 개발된 체지방률 추정모형은 기존의 회귀분석을 바탕으로 개발된 추정모형들과 비교시 대등하거나 높은 정확성을 보였을 뿐만 아니라 복잡한 계산 없이 나무구조로 모형화된 순서도를 따라 매우 쉽게 체지방률을 추정할 수 있었다.

Abstract

Body fat percentage(BF%) prediction models have been widely used; however, most of the prediction models rely on complex equations that may be impractical for everyday use. Therefore, decision tree(DT) analysis was applied to develop a user-friendly model to estimate BF%. Data(age ≥ 19 years; $n = 7,161$) from the 2009 Korean National Health and Nutrition Examination Survey (KNHANES) were divided into 80% training and 20% testing splits. Multiple DTs were constructed using dual-energy x-ray absorptiometry(DEXA) measured BF% and anthropometrics and demographics variables in training split. A parsimony DT with the lowest risk estimate was selected. The accuracy of the chosen DT was evaluated by R^2 and root mean square error(RMSE) against BF% measured by DEXA in testing split. The best model for estimation of BF% was a DT consisting of 5 levels and 4 predictors(risk estimate = 13.98; # of terminal nodes = 21). BF% estimated by the DT accounted for 73% of the variation in criterion-measured BF% with a RMSE of 4.03%. The new model using DT analysis to estimate BF% had moderate R^2 and small RMSE in a national representative sample of Korean adults. In addition, BF% can be estimated much easier by following the DT map (i.e., flowchart) compared with other equation models.

Key words: body fat percentage, decision tree analysis, KNHANES

* Minsoo.Kang@mtsu.edu

서론

신체구성(체지방, 체지방, 체지방률 등)은 영양 및 건강 상태에 대한 정보를 나타내는 중요한 지표로 영양학, 보건 및 의학, 운동과학 등 여러 분야에서 널리 활용되고 있다. 비정상적인 신체구성은 건강과 체력에 부정적 영향을 미친다(Institute of Medicine [IOM], 2012; World Health Organization [WHO], 2000). 특히, 건강에 위험을 초래할 정도의 과도한 체지방이 축적된 상태로 정의되는 비만은 고혈압, 당뇨, 암, 심혈관질환 등과 같이 잘 알려진 만성질환의 주요 위험 인자이다(WHO, 2000). 신체구성 요소 중 체지방률은 체중에서 체지방량이 차지하는 비율을 의미하며, 비만 정도를 나타내는 주요 지표로 활용되고 있다(Heyward, 2001; Lear, Humphries, Kohli, & Birmingham, 2007; Romero-Corral et al., 2007). 정확한 체지방률 측정 방법 개발은 비만 및 비만과 관련된 주요 만성질환의 예방과 치료를 위해 중요하다.

체지방률은 수중체중밀도법(Hydrodensitometry), 이중에너지 방사선 측정법(dual energy X-ray absorptiometry [DEXA]), 컴퓨터 단층 촬영법(computed tomography [CT]), 자기공명 단층 촬영법(magnetic resonance imaging [MRI]) 등의 정확한 측정이 가능한 준거 측정방법과 생체전기저항 분석방법(bioelectrical impedance analysis [BIA]), 피부두껍기법(skin folder), 체질량지수(body mass index [BMI]), 허리둘레 등과 같이 실용적인 현장 측정방법을 이용하여 측정할 수 있다(Wagner & Heyward, 1999). 준거 측정방법은 정확한 측정이 가능한 장점이 있는 반면, 현장 측정방법들에 비해 상대적으로 많은 시간, 높은 비용, 복잡한 기술 등을 필요로 하는 단점이 있다(Lukaski, 1987; Wagner & Heyward, 1999). 반대로 현장 측정방법들은 간단하고 사용하기 쉬운 장점을 갖고 있으나, 정확성이 준거 측정방법에 비해 상대적으로 낮은 단점이 있다(Lukaski, 1987; Wagner & Heyward, 1999).

BMI 및 허리둘레와 같은 인체계측치는 측정의 간편성과 실용성 때문에 임상 및 보건 현장과 대규모 관찰 연구에서 비만 측정을 위해 널리 사용되고 있다(Healy, Matthews, Dunstan, Winkler, & Owen, 2011; Healy et

al., 2008; Maher, Mire, Harrington, Staiano, & Katzmarzyk, 2013; Norgan, 2005; Van Dyck et al., 2015). BMI 및 허리둘레는 체지방량을 직접 측정하지는 않지만, 체지방량과 상관관계가 있기 때문에 체지방률을 간접적으로 측정하는 대체 측정방법으로 알려져 있다(Flegal et al., 2009; WHO, 2000). 하지만, 선행연구에서 BMI 및 허리둘레와 체지방률과의 관계는 나이, 성별 및 인종에 의해 달라지는 것으로 보고되었다(Flegal et al., 2009; Gallagher, Visser, Sepulveda, Pierson, Harris, & Heymsfield, 1996; Heo, Faith, Pietrobelli, & Heymsfield, 2012; Lear et al., 2007). 그러므로, 나이, 성별 및 인종에 대한 고려없이 간단한 인체계측치만으로 비만을 측정하는 것은 논란의 여지가 있다(De Lorenzo, Deurenberg, Pietrantuono, Di Daniele, Cervelli, & Andreoli, 2003; Ho-Pham, Campbell, & Nguyen, 2011; Lear et al., 2007; Seidell & Flegal, 1997).

이러한 이유로, 체지방률 측정방법의 대안으로 인체계측치(예, BMI, 허리둘레 등)와 사회인구학적 변인(예, 나이, 성별, 인종 등)에 근거한 체지방률 추정모형이 사용되고 있다(Cui, Truesdale, Cai, & Stevens, 2014; Gómez-Ambrosi et al., 2012; Kanellakis & Manios, 2012; Kim, Park, Park, Kim, & Kim, 2011; Norgan, 2005). 체지방률 추정모형은 BMI 및 허리둘레와 같이 단순한 측정방법 보다 정확한 체지방률을 추정할 수 있고(조정환, 2007; Cui, Truesdale, Cai, & Stevens, 2014; Gallagher et al., 2000; Zanovec et al., 2012), 준거 측정방법 보다 사용하기 간편하며 비용과 시간적 측면에서 실용적인 것이 장점이다(Gómez-Ambrosi, 2012; Norman, 2005). 그러나, 대부분의 체지방률 추정모형은 회귀분석 방법을 바탕으로 한 복잡한 추정식을 기반으로 하고 있기 때문에(조정환, 2007; Cui, Truesdale, Cai, & Stevens, 2014; Zanovec, Wang, & O'Neil, 2012) 일반인이 체지방률 추정모형을 일상생활에서 사용하기 어려운 실정이다. 또한, 추정모형에 포함되는 피부두께와 같은 인체계측치는 일반인이 측정하기 어려울 뿐만 아니라 대규모 관찰연구에서도 사용이 제한될 수 있다. 그러므로, 정확하면서도 대규모 관찰연구 또는 임상 및 보건 현장 그리고 일상생활 현장에서 누구나 쉽게 사용

할 수 있는 새로운 체지방률 추정모형 개발에 대한 추가 연구가 필요하다.

데이터 마이닝(Data Mining) 기법 중 하나인 의사결정나무 분석(decision tree analysis)은 기존의 데이터 내에 존재하는 모든 의사결정 규칙을 탐색하여 찾아낸 최적의 의사결정 규칙을 나무구조로 모형화 하는 기법으로, 분류와 예측을 위한 모형을 개발하는데 효과적인 방법이다(최종후, 한상태, 강현철, 김은석, 김미경 및 이성진, 2002; Ragan & Kang, 2005). 의사결정나무 분석은 선형성(linearity), 정규성(normality), 등분산성(equal variance) 등의 가정이 요구되지 않는 비모수적인 방법이며, 가능한 모든 교호작용(interaction) 효과를 분석할 수 있어 두개 이상의 예측변인이 어떻게 목표변인에 영향을 미치는지 쉽게 확인할 수 있는 장점이 있다(최종후 등, 2002). 의사결정나무 분석에 의해 형성된 나무구조 모형은 잠재적 사용자가 모형을 이해하기 쉽고, 예측변인이 목표변인에 미치는 영향을 쉽게 서열화 할 수 있으며, 새로운 데이터에 적용하기 쉬운 특성을 갖는다(최종후 등, 2002). 이러한 특성을 갖는 의사결정나무 분석 방법을 이용하여 체지방률 추정모형을 개발한다면, 대규모 관찰연구, 임상과 보건 현장 및 일상생활에서 누구나 쉽게 활용이 가능한 추정모형을 개발할 수 있을 것으로 사료된다(최종후, 2002; Ragan & Kang, 2005; Weatherby, Kang, Shapshak, Chiappelli, & McCoy, 2006).

그러므로, 본 연구의 목적은 의사결정나무 분석방법을 이용하여 성인을 위한 체지방 추정모형을 개발하는 것이다. 이를 위해 준거 측정방법인 DEXA에 의해 측정된 체지방률을 바탕으로 체지방률 추정모형을 개발하였고, 독립적인 표본에서 개발된 추정모형의 정확성을 확인하였다.

연구방법

1. 연구대상

본 연구에서는 보건복지부의 질병관리본부에서 실시하고 있는 국민건강영양조사 제4기(2007~2009) 중

2009년 자료를 활용하였다. 국민건강영양조사는 대한민국에 거주하는 모든 국민을 모집단으로 하며, 매년 3단계 층화집락표본추출방법과 순환표본조사 방법을 통해 전 국민을 대표할 수 있는 표본을 추출하여 건강 설문조사, 영양조사, 검진조사 등을 실시하는 단면조사(cross-sectional survey)이다. 본 연구는 2차 자료분석 연구이므로 생명연구윤리위원회 심의는 생략되었다.

총 10,533명의 참가자 중 성인 7,197명(만 19세 이상)을 대상으로 체지방률을 측정하였다. 임신 중인 여성, 1주일 이내에 조영제를 이용한 검진(contrast agent-based examination)을 받은 인원, 몸무게 136kg 및 신장 196cm 초과 인원은 건강상 위험과 부정확한 측정 가능성 때문에 DEXA측정에서 제외 되었다. 추가적으로 주요 예측 변인으로 선별된 인체계측치(신장, 체중, BMI, 허리둘레)와 사회인구학적 변인(나이, 성별)에 대한 자료가 누락된 인원 36명이 본 연구에서 제외 되었다. 따라서, 최종 연구대상은 성인 7,161명이었다.

2 측정변인

1) 인체계측 및 체지방률 측정

인체계측 및 체지방률 측정은 국민건강영양조사 검진 지침서의 절차에 따라 측정되었다. 조사 참가자는 검진용 가운을 착용하였고, 검사에 방해가 될 수 있는 신발, 양말, 장식품 및 귀금속류 등은 제거되었다. 신장은 신장계를 이용하여 0.1cm 단위로 측정되었고, 체중은 디지털 체중계를 이용하여 0.1kg 단위로 측정되었다. BMI(kg/m²)는 측정된 신장과 체중을 바탕으로 산출되었다. 허리둘레는 참가자의 측면(mid-axillary line)에서 마지막 늑골 하단 및 장골 능선 상단 두 지점을 측정하여 두 지점의 중간 부위에서 바닥과 평행하게 줄자를 사용하여 0.1cm 단위로 측정하였다.

신체구성은 DEXA를 이용하여 측정하였다. 이를 위해 QDR4500W Discovery fan beam densitometer(Hologic Inc., Bedford, MA, USA)와 Hologic Discovery software(version 13.1)가 사용되었다. 체지방률은 총지방량을 몸무게로 나눈 값에 100을 곱하여 산출되었다. DEXA 측정과정 감독 및 측정결과에 대한 질적 검토는 대한골다공증학회에서 실시하였다.

2) 사회인구학적 변인 및 건강관련 행태 측정

나이, 성별(남, 여), 각 강도별 신체활동량, 신체활동 가이드라인 실천여부, 월평균 가구 총소득(만원), 교육 수준(중졸 이하, 고졸, 대졸 이상), 1일 에너지 섭취량(kcal), 1일 지방 섭취량(g), 음주 및 흡연 상태는 국민 건강영양조사 건강설문을 이용하여 참가자의 자기 보고법에 의해 측정되었다. 신체활동량은 최근 일주일 동안 실시한 격렬한 신체활동량(분), 중등도 신체활동량(분), 걷기 활동량(분)으로 각각 측정되었다. 신체활동 가이드라인 실천여부는 격렬한 신체활동(1회 20분, 주 3일 이상 실시), 중등도 신체활동(1회 30분, 주 5일 이상 실시), 걷기 신체활동(1회 30분, 주 5일 이상 실시) 중 하나를 만족했을 경우 실천한 것으로 간주하였다. 음주상태는 주당 음주 횟수에 따라서 비음주(평생 비음주 또는 지난 1년간 비음주), 가벼운 음주(주 1회 이하 음주) 그리고 과음주(주 2회 이상 음주) 그룹으로 분류하였고, 흡연 상태 또한 하루 흡연양에 따라 비흡연(평생 비흡연 또는 현재 비흡연), 가벼운 흡연(하루 20개비 이하 흡연), 및 과흡연(하루 20개비 이상 흡연) 그룹으로 분류하였다. 1일 에너지 섭취량과 지방 섭취량은 24 시간 회상조사방법으로 측정하였다.

3. 자료처리방법

총 7,161명을 포함하는 최종 데이터를 단순무작위추출법에 의해 8:2의 비율로 체지방률 추정모형 개발을 위한 데이터($n = 5,728$)와 추정모형의 정확성을 검증하기 위한 데이터($n = 1,433$)로 나누었다. 이를 위해 복합 표본설계 데이터에서 단순무작위로 샘플을 추출하기에 적합한 SAS SURVEYSELECT procedure를 사용하였다.

목표변인은 DEXA로 측정된 체지방률이었으며, 잠재적 예측변인은 4개의 인체계측치(체중, 신장, BMI, 허리둘레), 4개의 사회인구학적 변인(나이, 성별, 교육 수준, 월평균 가구 총소득), 8개의 건강관련 행태 변인(격렬한 신체활동, 중등도 신체활동, 걷기 신체활동, 신체활동 실천여부, 음주상태, 흡연상태, 1일 에너지 섭취량, 1일 지방 섭취량) 등 총 16개였다. 잠재적 예측변인은 국민건강영양조사에서 제공하는 변인들 중 이전에 개발된 추정식 모형(조정환, 2007; Cui et al., 2014;

Zanovec et al., 2012)과 비만에 영향을 미치는 사회인구학적 및 건강관련 행태 변인들(Healy et al., 2011; Maher et al., 2013; Seidell & Flegal, 1997)을 고려하여 대규모 관찰연구 및 현장에서 어렵지 않게 측정 가능하거나 또는 통상적으로 함께 측정되는 변인들로 선별하였다.

의사결정나무 형성을 위해 가장 널리 활용되고 있는 CHAID(chi-squared automatic interaction detection)와 CART(classification and regression trees) 알고리즘(algorithm)을 적용하였다(Ragan & Kang, 2005). CHAID 알고리즘은 F 통계량을 최적의 분리기준으로 사용하였고, 유의 수준은 .05로 지정하였다. CART 알고리즘은 분산(variance)의 감소량을 최적의 분리기준으로 사용하였고, 최소한의 향상도는 .0001로 지정하였다. 간결한 모형 형성을 위해 5단계 깊이 이내(maximum tree depth), 부모 마디(parent node) 최소 100명, 자식 마디(child node) 최소 50명 기준을 적용하였다. 위에서 언급된 방법 및 기준들을 적용하여, 개발 데이터에서 다수의 의사결정나무 모형을 형성하였고, 형성된 모형 중 가장 간단하면서도 낮은 위험 추정치(risk estimate)를 갖는 의사결정나무 모형을 최적의 체지방률 추정모형으로 선택하였다.

최종 선택된 체지방률 추정모형은 검증 데이터를 이용하여 정확성을 확인하였다. 추정모형의 정확성은 회귀분석의 결과로 제공되는 R^2 와 평균 제곱근 오차(root means square error = $\sqrt{\sum(Y-Y')^2/N}$, $Y = \text{DEXA 측정 체지방률}$, $Y' = \text{추정모형 추정 체지방률}$, $N = \text{표본수}$)에 의해 평가되었다. 이를 위해 체지방률 추정모형에 의해 추정된 체지방률을 예측변인으로 하고 DEXA에 의해 측정된 체지방률은 목표변인으로 하여 단순 회귀분석을 수행하였다. 이와 더불어, 추정된 체지방률의 편향성(systematic bias) 여부를 확인하기 위해 단순 회귀분석의 기울기와 y절편이 각각 1, 0과 유의한 차이가 있는지 확인하였다. 추가적으로 추정모형에 의해 추정된 체지방률과 DEXA에 의해 측정된 체지방률간의 평균 차이를 검증하였다.

의사결정나무 분석은 SPSS v.20.0을 이용하여 수행하였고, 이외 모든 자료분석은 복합표본설계(다단계 층화집락확률추출법)를 고려하기 위하여 SAS v.9.3 SURVEYSELECT, SURVEYFREQ, SURVEYMEANS, SURVEYREG 절차를 이용하였다.

결과

연구 대상자의 일반적 특성은 <표 1>과 같다. 전체 연구대상자의 성별 비율은 대략 50:50 이었고, 평균 나이는 약 45세 이었다. 평균 BMI와 체지방률은 각각 23.62kg/m²와 27.05 ± 0.16% 이었다. 전반적으로 단순 무작위 추출법에 의해 나누어진 개발 데이터(n = 5,728)

와 검증 데이터(n = 1,433)의 인체계측치, 사회통계학적 변인 및 건강행태 변인의 평균 및 비율은 전체적으로 유사하였다($p < .05$).

형성된 다수의 체지방률 추정모형들 중 가장 정확한 모형과 간결하고 낮은 위험 추정치를 가진 대표적 4개의 모형은 <표 2>와 같다. 이중 최종 선택된 모형은 4개의 변인(성별, 나이, BMI, 허리둘레)과 CART 알고리

표 1. 연구대상자 특성

	전체 데이터 (N = 7,161)	개발 데이터 (n = 5,728)	검증 데이터 (n = 1,433)
나이(years)	44.78 ± .38	44.94 ± .40	44.15 ± .57
성별(%)			
남자	49.76 ± .54	50.51 ± .65	46.79 ± 1.28
여자	50.24 ± .54	49.49 ± .65	53.21 ± 1.28
신장(cm)	163.80 ± .16	163.86 ± .18	163.57 ± .28
체중(kg)	63.61 ± .17	63.68 ± .19	63.36 ± .40
BMI(kg/m ²)	23.62 ± .05	23.63 ± .06	23.59 ± .11
허리둘레(cm)	80.72 ± .23	80.82 ± .24	80.32 ± .41
체지방률(%)	27.05 ± .16	26.99 ± .16	27.31 ± .27
신체활동 실천여부(%)			
실천	44.05 ± .77	44.01 ± .82	44.22 ± 1.42
미실천	55.95 ± .77	55.99 ± .82	55.78 ± 1.42
신체활동량(분)			
격렬한 신체활동	107.96 ± 2.31	108.58 ± 2.67	105.75 ± 4.05
중등도 신체활동	106.48 ± 2.28	107.64 ± 2.67	102.04 ± 5.24
건기 신체활동	82.27 ± 1.50	83.14 ± 1.75	78.82 ± 2.42
월평균 가구 소득(만원)	339.69 ± 15.65	338.66 ± 16.70	343.79 ± 24.83
교육수준(%)			
중졸 이하	30.31 ± 1.01	30.78 ± 1.10	28.46 ± 1.52
고졸	39.97 ± 1.02	39.94 ± 1.04	40.95 ± 1.71
대졸 이상	29.71 ± 1.17	29.28 ± 1.22	31.44 ± 1.73
음주상태(%)			
비음주	24.40 ± .72	24.50 ± .80	24.02 ± 1.34
가벼운 음주	52.53 ± .82	52.21 ± .91	53.79 ± 1.49
과음주	23.07 ± .62	23.29 ± .68	22.19 ± 1.23
흡연상태(%)			
비흡연	73.01 ± .65	72.43 ± .77	75.33 ± 1.25
가벼운 흡연	15.48 ± .05	15.90 ± .59	13.80 ± 1.08
과흡연	11.51 ± .49	11.67 ± .55	10.88 ± .90
1일 에너지 섭취량(kcal)	1918.36 ± 13.32	1916.99 ± 15.42	1923.85 ± 29.55
1일 지방 섭취량(g)	38.62 ± .55	38.63 ± .64	38.58 ± .93

주) N 또는 n 은 복합표본설계를 고려하기 전의 표본수; 모든 값들은 평균 ± 표준 오차(standard error of mean) 또는 퍼센트 ± 표준 오차(standard error of percent)로 표현됨.

표 2. 체지방률 추정모형 요약

입력된 예측변수	알고리즘	형성된 의사결정나무 모형의 예측변수	추정 위험도
나이 성별 체중 신장 BMI 허리둘레 신체활동 가이드라인 실천 여부 격렬한 신체활동량 중등도 신체활동량 걷기 신체활동량 교육수준 월평균 가구소득 음주 상태 흡연 상태 1일 에너지 섭취량 1일 지방 섭취량 (총 16개 변수)	CART	성별, BMI, 허리둘레, 나이, 흡연 상태, 1일 에너지섭취량, 격렬한 신체활동량(총 7개 변수)	Risk Estimate = 13.86 Std. Error = .30 R ² = .76
		1 단계: 성별	
		2 단계: BMI, 허리둘레	
		3 단계: BMI, 허리둘레	
		4 단계: BMI, 허리둘레, 1일 에너지 섭취량	
	CHAID	5 단계: BMI, 허리둘레, 나이, 1일 에너지 섭취량, 흡연 상태, 격렬한 신체활동량	Risk Estimate = 13.95 Std. Error = .31 R ² = .76
		성별, BMI, 허리둘레, 나이, 신장, 흡연 상태, 음주상태, 1일 에너지섭취량, 월평균 가구소득, 교육수준, 중등도 신체활동량, 신체활동 실천여부(총 12개 변수)	
		1 단계: 성별	
		2 단계: BMI, 허리둘레	
		3 단계: BMI, 허리둘레, 나이, 흡연, 신장, 1일 에너지 섭취량, 월평균 가구소득	
나이 성별 BMI 허리둘레 (총 4개 변수)	CART	4 단계: BMI, 신장, 신체활동 실천여부, 나이, 교육수준, 음주 상태, 중등도 신체활동량	Risk Estimate = 13.98 Std. Error = .30 R ² = .76
		5 단계: 나이, 흡연 상태	
		성별, BMI, 허리둘레, 나이(총 4개 변수)	
		1 단계: 성별	
		2 단계: BMI, 허리둘레	
	CHAID	3 단계: BMI, 허리둘레, 나이	Risk Estimate = 14.41 Std. Error = .32 R ² = .75
		성별, BMI, 허리둘레, 나이(총 4개 변수)	
		1 단계: 성별	
		2 단계: BMI, 허리둘레	
		3 단계: BMI, 허리둘레, 나이	
		4 단계: 나이	

주) BMI = 체질량지수

음을 통해 형성된 체지방률 추정모형이었다(표 2 하단에서 두 번째 모형). 모형의 위험 추정치는 13.98이었고, DEXA에 의해 측정된 체지방률의 변산도(variability)는 추정모형에 의해 76% 설명 되었다. 추정모형은 끝 마디(terminal node) 21개와 5단계의 깊이를 가졌다. 나무구조로 모형화된 체지방률 추정모형은 <그림 1>에 묘사 되었다. 최종 선택된 모형은 가장 높은 정확성을 보인 모형(표 2 상단에서 첫 번째 모형; 위험 추정치 = 13.86)과 비교시 위험 추정치는 .12 높았지만, 그 크기는 매우 근소하였으며 설명력의 차이는 거의 없었다(R² = .76). 또한, 예측변인은 7개(성별, 나이, BMI, 허리둘레,

1일 에너지 섭취량, 격렬한 신체활동량)에서 4개로 줄었으며, 선택된 예측변인들은 현장에서 측정이 용이한 나이, 성별, BMI 및 허리둘레였다.

최종 선택된 체지방률 추정모형은 정확성 검증을 위해 검증 데이터에 적용되었고, 결과는 <표 3, 4>와 같다. 체지방률 추정모형은 준거 측정방법인 DEXA에 비해 체지방률을 과소 추정하는 것으로 나타났지만, 유의한 차이는 없었다(평균차 = -0.17; 95% 신뢰구간 = -0.53 ~ .18). 추정모형과 DEXA에 의해 측정된 체지방률간의 단순 회귀분석 결과, 기울기는 1과 다르지 않았고(95% 신뢰구간 = .97 ~ 1.06), Y 절편은 0과 다르지

표 3. DEXA와 선택된 모형에 의해 추정된 체지방률

	평균	표준오차	95% 신뢰도 구간	
			하한 값	상한 값
DEXA 측정 체지방률(%)	27.31	.27	26.78	27.85
모형에 의해 추정된 체지방률(%)	27.14	.18	26.79	27.49
추정오차 (%)	-0.17	.18	-0.53	.18

표 4. DEXA와 선택된 모형에 의해 추정된 체지방률간 단순 회귀분석

	평균	표준오차	95% 신뢰도 구간	
			하한 값	상한 값
Y 절편(intercept)	-0.32	.58	-1.47	0.83
기울기(slope)	1.02	.02	.97	1.06

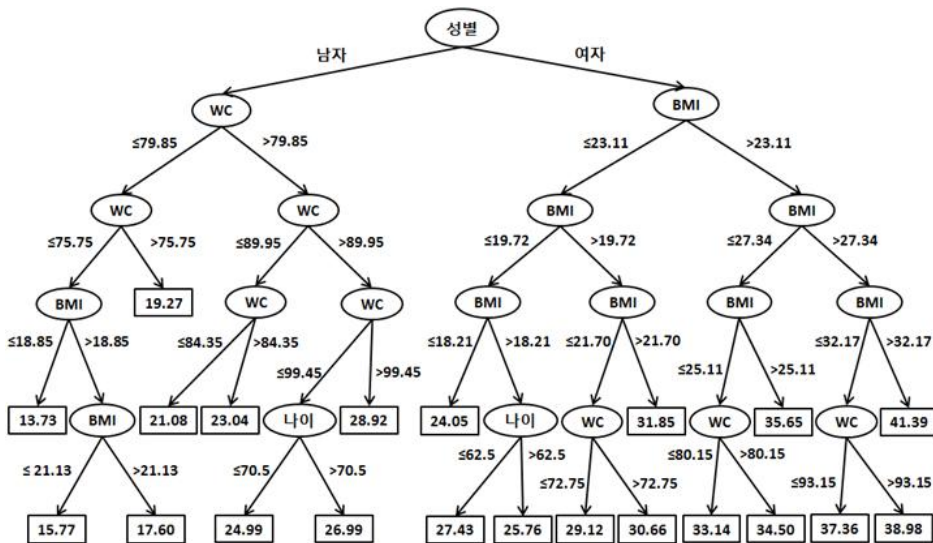


그림 1. 최종 선택된 체지방 추정모형 (WC: 허리둘레, BMI: 체질량지수)

않았다(95% 신뢰구간 = -1.47 ~ .83). 즉, 전반적으로 최종 선택된 모형에 의해 추정된 체지방률은 DEXA에 의해 추정된 체지방률을 편향성 없이 추정하였다. 추정모형에 의해 추정된 체지방률은 DEXA에 의해 측정된 체지방률 변산도(variability)의 73%를 설명했고($R^2 = .73$), 평균 제곱근 오차는 4.03%로 크지 않았다.

논의

본 연구의 목적은 대규모 관찰연구와 임상 및 보건 현장 뿐만 아니라 일반인이 일상생활에서 간편하게 사

용할 수 있는 체지방률 추정모형을 개발하고 검증하는 것이었다. 개발된 체지방률 추정모형은 <그림 1>에서 제시한 바와 같이 나무구조로 모형화 되어 성별, 나이, BMI 및 허리둘레 측정값을 알면, 추정모형에 대한 특별한 전문적 지식과 계산도구 없이도 매우 쉽고 간단하게 체지방률을 추정할 수 있다. CART 알고리즘에 의해 형성된 추정모형은 2개의 자식 마디로만 분리되기 때문에 뿌리 마디(1단계) 부터 끝 마디(4 또는 5단계)로 이르기까지, 각 단계에서 성별, 나이, BMI 및 허리둘레에 대한 분리기준에 “예” 또는 “아니요”로 답변하면 끝마디에 이르게 되고, 끝마디에 표시된 체지방률은 피측정자의 추정된 체지방률 값이 된다. 예를 들어, 피측정자

가 34세 남성(BMI 26.02kg/m², 허리둘레 96.60cm)이라고 가정하고 체지방률 추정해 보자. 첫 번째 질문은 “당신의 성별은 무엇인가요?”이고, 피측정자가 남자기 때문에 나무구조화된 모형을 따라 왼쪽으로 내려간다. 두 번째 질문은 “당신의 허리둘레는 79.85cm 보다 큰가요?”이고, 피측정자의 허리둘레는 96.60cm 이므로 오른쪽으로 내려간다. 세 번째 질문은 “당신의 허리둘레는 89.95cm 보다 큰가요?”이고, 피측정자의 허리둘레는 96.60cm 이므로 오른쪽으로 내려간다. 네 번째 질문은 “당신의 허리둘레는 99.45cm 보다 큰가요?”이고, 피측정자의 허리둘레는 96.60cm이므로 왼쪽으로 내려간다. 마지막 질문은 “당신의 나이는 70.5세 이상인가요?”이고, 피측정자의 나이는 34세이므로 왼쪽으로 내려간다. 마침내, 끝마디에 이르렀고 추정된 체지방률은 24.99%이다. 개발된 체지방률 추정모형은 그림 형태로 출력하여 가정 또는 임상 및 보건 관련 현장에 배치되어 사용할 수 있을 뿐만 아니라 추정모형의 알고리즘을 스마트폰 및 테블릿 PC 그리고 개인용 컴퓨터에 구현하여 개인 건강관리, 임상 및 보건 관련 현장, 대규모 관찰연구 등의 다양한 분야에서 활용이 가능할 것으로 사료된다.

본 연구에서 개발된 체지방률 추정모형은 위에서 설명된 것과 같이 사용이 쉽고 간편할 뿐만 아니라 합당한 정확성도 가지고 있는 것으로 판명되었다. ASEP(The American Society of Exercise Physiologists)는 사용자가 정확성이 검증된 체지방률 추정모형을 선택할 수 있도록 가이드라인을 제시하였다(Heyward, 2001). 가이드라인은 1) 준거 측정방법에 의해 측정된 체지방률을 기준 체지방률로 사용하여 개발된 모형, 2) 단순 무작위 추출법에 의해 모집된 100명 초과 인원을 기반으로 개발된 모형, 3) 준거 측정방법에 의해 측정된 체지방률과 추정식 모형에 의해 추정된 체지방률의 상관관계가 .80 초과하는 모형, 4) 수용 가능한 추정오차 또는 표준오차($\leq 4.5\%$)를 가진 모형, 5) 개발된 집단 외 표본에서 교차 타당성이 검증된 모형이다. 본 연구의 체지방률 추정모형은 5가지 기준을 모두 만족하였다. 1) 준거 측정방법인 DEXA에 의해 측정된 체지방률을 기준 체지방률로 사용, 2) 전 국민을 대표하는 표본 중 단순 무작위 추출법에 의해 선택된 5,728명의 성

인을 대상으로 개발, 3) 검증 데이터에서 준거 측정방법에 의해 측정된 체지방률과 모형에 의해 추정된 체지방률의 상관관계는 .85, 4) 검증 데이터에서 수용 가능한 평균 제곱근 오차(4.03%; fairly good)를 보임, 5) 전 국민을 대표하는 표본 중 단순 무작위추출법에 의해 선택된 1,433명의 성인을 대상으로 정확성을 검증하였다.

게다가, 최종 선택된 추정모형은 기존에 개발된 추정식을 기반으로한 모형들과 비교시 대등하거나 높은 정확성을 보였다. Cui et al.(2014)는 인체계측치와 사회통계학적 변인을 바탕으로 개발된 26개의 체지방률 추정식 모형을 미국 국민건강영양조사 데이터를 이용하여 정확성을 검증하였다. 대부분의 추정식 모형은 DEXA에 의해 측정된 체지방률 변산도(variability)의 50% ~ 70%를 설명할 수 있었고, 3.5% ~ 4.5% 평균 제곱근 오차를 보였다. 본 연구의 추정모형은 73%의 설명력과 4.03%의 평균 제곱근 오차를 보였다. 회귀분석 방법을 바탕으로 개발된 이전의 추정식 모형은 BMI와 허리둘레 간의 높은 상관관계로 다중공선성(multi-collinearity)을 야기할 수 있기 때문에 두 인체계측치를 하나의 모형에 예측변인으로 동시 입력이 제한된다. 따라서, 통상 두 예측변인 중 하나를 선택하여 하나의 모형을 만들거나, 각각의 지표를 포함하는 2개의 모형을 개발하였다(Cui et al., 2014; Zanovec et al., 2012). 하지만, 본 연구의 추정모형은 성별에 따라 체지방률과 다른 상관관계를 보이는 BMI와 허리둘레를 동시에 한 모형에 예측변인으로 입력이 가능했고, 이들이 중복 또는 교차 사용되어 최적의 분리기준을 만들었기 때문에 추정식에 기반을 둔 대부분의 추정모형 보다 비교적 정확한 예측이 가능했을 것으로 사료된다.

본 추정모형에 포함된 예측변인과 그 기여도는 기존의 연구들과 유사한 결과를 보였다. Cui et al.(2014)에 따르면 남자 그룹에서는 허리둘레를, 여자 그룹에서는 BMI를 포함한 추정식 모형이 가장 좋은 정확성을 보였다고 보고하였다. 본 연구에서도 유사한 경향을 확인할 수 있었다. 1단계 분리기준은 성별(남자, 여자)이었고, 2단계 분리기준은 남자 그룹에서는 허리둘레($\leq 79.85\text{cm}$ 또는 $> 79.85\text{cm}$), 여자 그룹에서는 BMI($\leq 23.11\text{kg/m}^2$ 또는 $> 23.11\text{kg/m}^2$)가 최적의 분리변인 및 기준으로 선택되었다. 또한, Zanovec et al.(2012)의 연

구에 따르면 성별, BMI와 허리둘레, 나이, 인종 순으로 체지방률의 추정에 기여했다. 본 연구에서 최종 선택된 모형에 포함된 예측변인은 성별, BMI, 허리둘레, 나이였고, 최적 분리기준 순서는 성별(1단계), BMI와 허리둘레(2~4단계), 나이(5단계) 순이었다. 이러한 결과는 의사결정나무 분석이 회귀분석에서 목표변인을 예측하는 변인들의 선택 또는 예측변인들 간의 교호작용 효과에 대한 사전 정보를 얻는 유용한 방법이 될 수 있음을 시사한다.

본 연구의 추정모형은 여러 장점에도 불구하고 다음과 같은 제한점을 갖는다. 직접 측정이 아닌 추정모형의 내재적 한계 때문에 개인의 체지방률에 대한 매우 정확한 진단이 필요한 경우는 사용이 제한 될 수 있다. 특히, 극단적으로 체지방률이 높거나 낮은 성인의 경우 추정오차가 평균 제공근 오차 보다 커질 수 있다. 또한, 본 연구의 체지방률 추정모형은 단면연구 자료를 이용하여 개발 및 검증되었기 때문에 체지방률 변화 추이를 추적하는 목적으로 사용하기 위해서는 추가 검증 연구가 필요할 것이다.

결론

본 연구는 대한민국 성인을 대표할 수 있는 표본으로부터 의사결정나무 분석 방법을 이용하여 체지방률 추정모형을 개발한 연구이다. 개발된 체지방률 추정모형은 널리 활용되고 있는 회귀분석 방법을 바탕으로 개발된 추정모형들의 단점을 보완하면서 동시에 비교적 정확한 체지방률을 추정할 수 있었다. 본 연구에서 개발된 추정모형의 가장 큰 장점은 복잡한 추정식과 컴퓨터 및 계산기에 의존하지 않고도 누구나 쉽게 나무구조로 모형화된 흐름도를 따라 체지방률을 추정할 수 있다는 것이다. 앞으로 본 연구의 체지방률 추정모형의 정확성을 확인하기 위하여 다양한 표본과 방법을 통해 추가 검증 연구가 필요할 것으로 사료된다.

참고문헌

조정환(2007). BMI지수를 활용한 체지방량 추정: 교차 타당화 연구. **한국체육학회지**, 9(1), 75-85.

최종후, 한상태, 강현철, 김은석, 김미경 및 이성진 (2002). **AnswerTree 3.0을 이용한 데이터마이닝 예측 및 활용**. 서울: SPSS 아카데미.

Cui, Z., Truesdale, K. P., Cai, J., & Stevens, J. (2014). Evaluation of anthropometric equations to assess body fat in adults: NHANES 1999-2004. *Medicine and science in sports and exercise*, 46, 1147-1158.

De Lorenzo, A., Deurenberg, P., Pietrantuono, M., Di Daniele, N., Cervelli, V., & Andreoli, A. (2003). How fat is obese? *Acta diabetologica*, 40, s254-s257.

Flegal, K. M., Shepherd, J. A., Looker, A. C., Graubard, B. I., Borrud, L. G., Ogden, C. L., et al. (2009). Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults. *The American journal of clinical nutrition*, 89, 500-508.

Gallagher, D., Visser, M., Sepulveda, D., Pierson, R. N., Harris, T., & Heymsfield, S. B. (1996). How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?. *American journal of epidemiology*, 143, 228-239.

Cómez-Ambrosi, J., Silva, C., Catalán, V., Rodríguez, A., Galofré, J. C., Escalada, J., et al. (2012). Clinical usefulness of a new equation for estimating body fat. *Diabetes Care*, 35, 383-388.

Healy, G. N., Matthews, C. E., Dunstan, D. W., Winkler, E. A., & Owen, N. (2011). Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003-06. *European heart journal*, 32, 590-597.

Healy, G. N., Wijndaele, K., Dunstan, D. W., Shaw, J. E., Salmon, J., Zimmet, P. Z., et al. (2008). Objectively measured sedentary time, physical activity, and metabolic risk the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). *Diabetes care*, 31, 369-371.

- Heo, M., Faith, M. S., Pietrobelli, A., & Heymsfield, S. B. (2012). Percentage of body fat cutoffs by sex, age, and race-ethnicity in the US adult population from NHANES 1999-2004. *The American journal of clinical nutrition*, 95, 594-602.
- Heyward, V. (2001). ASEP methods recommendation: body composition assessment. *Journal of Exercise Physiology*, 4, 1-12.
- Ho-Pham, L. T., Campbell, L. V., & Nguyen, T. V. (2011). More on body fat cutoff points. *Mayo Clinic Proceedings*, 86, 584-585.
- Kanellakis, S., & Manios, Y. (2012). Validation of five simple models estimating body fat in white postmenopausal women: use in clinical practice and research. *Obesity*, 20, 1329-1332.
- Kim, C. H., Park, H. S., Park, M., Kim, H., & Kim, C. (2011). Optimal cutoffs of percentage body fat for predicting obesity-related cardiovascular disease risk factors in Korean adults. *The American journal of clinical nutrition*, 94, 34-39.
- Institute of Medicine. (2012). *Fitness measures and health outcomes in youth*. Washington, DC: The National Academies Press.
- Lear, S. A., Humphries, K. H., Kohli, S., & Birmingham, C. L. (2007). The use of BMI and waist circumference as surrogates of body fat differs by ethnicity. *Obesity*, 15, 2817-2824.
- Lukaski, H. C. (1987). Methods for the assessment of human body composition: traditional and new. *The American journal of clinical nutrition*, 46, 537-556.
- Maher, C. A., Mire, E., Harrington, D. M., Staiano, A. E., & Katzmarzyk, P. T. (2013). The independent and combined associations of physical activity and sedentary behavior with obesity in adults: NHANES 2003-06. *Obesity*, 21, E730-E737.
- Norgan, N. G. (2005). Laboratory and field measurements of body composition. *Public health nutrition*, 8, 1108-1122.
- Ragan, B. G. & Kang, M. (2005). Construction of a Classification/Decision Tree. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 7, 61-75.
- Romero-Corral, A., Montori, V. M., Somers, V. K., Korinek, J., Thomas, R. J., Allison, T. G., ... & Lopez-Jimenez, F. (2006). Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *The Lancet*, 368, 666-678.
- Romero-Corral, A., Somers, V. K., Sierra-Johnson, J., Jensen, M. D., Thomas, R. J., Squires, R. W., et al. (2007). Diagnostic performance of body mass index to detect obesity in patients with coronary artery disease. *European heart journal*, 28, 2087-2093.
- Seidell, J. C. & Flegal, K. M. (1997). Assessing obesity: classification and epidemiology. *British medical bulletin*, 53, 238-252.
- Van Dyck, D., Cerin, E., De Bourdeaudhuij, I., Hinckson, E., Reis, R. S., Davey, R., et al. (2015). International study of objectively measured physical activity and sedentary time with body mass index and obesity: IPEN adult study. *International Journal of Obesity*, 39, 199-207.
- Wagner, D. R., & Heyward, V. H. (1999). Techniques of body composition assessment: a review of laboratory and field methods. *Research quarterly for exercise and sport*, 70, 135-149.
- Weatherby, N. L., Kang, M., Shapshak, P., Chiappelli, F., & McCoy, C. B. (2006). Screening women for Human Immunodeficiency Virus (HIV) infection using self-reported symptoms: A classification tree analysis. *Korean Journal of Measurement and Evaluation in Physical*

- Education and Sports Science*, 8(1), 1-17.
- World Health Organization. (2000). *Obesity: preventing and managing the global epidemic: report of a WHO consultation on obesity (WHO technical report series: 894)*. Geneva: WHO.
- Zanovec, M., Wang, J., & O'Neil, C. E. (2012). Development and Comparison of Two Field-Based Body Fat Prediction Equations: NHANES 1999-2004. *International Journal of Exercise Science*, 5, 5.

