

# 개인 식품섭취 선호도에 따른 당뇨병 발생 위험도 분석 모델

전소혜<sup>1</sup>, 김남현<sup>1\*</sup>

<sup>1</sup>연세대학교 의과대학 의학공학교실

## Diabetes Risk Analysis Model with Personalized Food Intake Preference

So-Hye Jeon<sup>1</sup> and Nam-Hyun Kim<sup>1\*</sup>

<sup>1</sup>Department of Medical Engineering, Yonsei University

**요 약** 최근 만성질환의 증대로 질병의 지속적인 관리에 대한 필요성이 대두되고 있으나 개인 맞춤형 식품섭취와 관련한 분석 연구는 미비하다. 본 연구에서 제시한 당뇨병 발생 위험도 분석 모델은 기존 상품 선호도 연구를 통해 검증된 유사도평가 방법인 피어슨 상관계수 산출을 이용하여 식품 선호도를 산출하고, 유병률 분석 연구 등에 활용된 로지스틱 회귀분석 방법을 이용하여 당뇨병발생 위험도를 제시하였다. 모델의 유의성 평가를 위해 당뇨 진단을 받은 환자 52명(남자: 22명, 여자: 30, 평균연령: 57(±13.2)세)과 비교군과 비슷한 연령을 조건으로 무작위 추출된 52명(남자: 17명, 여자: 35, 평균연령: 58(±9.4)세)의 t-검정을 통해 0.05 수준에서 유의함을 검증하였다. 이를 통해 식품섭취빈도에 따른 당뇨병 발생 위험도에 차이가 있음(p=0.046)을 확인할 수 있었다. 이 모델을 기반으로 분석된 정보는 스마트폰이나 웹을 이용한 개인 건강관리 시스템에 적용을 통해 새로운 개인 맞춤형 건강관리 방법으로 제시될 수 있다.

**Abstract** The need of continuous management for diseases came to the fore as a chronic disease has increased, however, research related to personalized food intake analysis are insufficient. In diabetes risk analysis model of this study, food preferences are calculated by Pearson correlation coefficient that is proven method to assess the similarity, and diabetes risk is computed as a Logistic regression that was used in prevalence studies. For the Significance evaluation of this model, it was verified through t-test at 0.05 level of 52 comparison subjects and 52 control subjects. Both groups were significantly independent (p=0.046 <0.05). This model is a new way to personalized health management, through the application to healthcare system based on web and mobile.

**Key Words** : Food Intake Frequency, Korean National Health and Nutrition Examination Survey (KNHANES), Pearson correlation coefficient, Similarity Weight

### 1. 서론

현대사회의 질병은 영양소 결핍이나 병원균에 의한 질환보다 생활습관 및 식이요인 등으로 생기는 만성질환이 큰 비중을 차지하고 있다. 그 중 당뇨병은 국내 사망원인 5순위로 사망률 21.5%로 높은 순위로 조사되었다[1]. 또한 당뇨병의 위험성은 심장마비, 뇌졸중, 신부전, 망막증, 신경합병증 등과 같은 합병증을 유발하기 때문에 지속적

인 관리가 중요하다고 알려져 있다[2]. 만성질환의 지속적인 관리를 위해 개인건강관리시스템이 등장하였고, 이를 이용한 건강관리방법을 다양하게 제시하고 있지만 개인의 맞춤형 식품섭취와 관련한 질병관리방법에 대한 연구는 미비하다. 특정 연령군 혹은 집단의 영양소 섭취 경향에 대한 연구나, 나트륨, 당분 같은 특정 영양소에 관한 연구를 통해 영양소가 질병발생이나 건강에 미치는 요인에 관한 연구들이 있다[3-5]. 그러나 영양소를 기반

\*Corresponding Author : Nam-Hyun Kim(Yonsei Univ.)

Tel: +82-2-2228-1915 email: knh@yuhs.ac

Received October 10, 2013

Revised (1st October 23, 2013, 2nd November 4, 2013)

Accepted November 7, 2013

으로 한 연구들의 경우, 단일 식품의 섭취보다는 다양한 식품으로 구성된 음식을 개인 선호도와 습관에 따라 선택하여 섭취하는 식사의 특징으로 인해 실생활 적용에 직접적인 방향을 제시하지 못한다[6]. 이러한 문제를 보완하기 위하여 식사패턴을 분석하는 연구들이 제안되었다. 식사패턴에 대한 분석으로 식사의 질 평가를 하거나, 식품섭취빈도의 요인분석을 통해 질병과 관련된 식품의 요인을 분석하는 연구들도 진행되고 있고, 식품섭취빈도의 패턴에 따른 당뇨병의 상관관계 연구 또한 진행되고 있다[7-10]. 이러한 연구에서 일반적인 당뇨병 유병율을 제시하여 질병관리에 대한 필요성에 대해 고찰하고 있으나, 이 또한 개인의 특성에 맞춘 맞춤형 서비스를 제공할 수 없다.

본 연구에서는 개인 식품 선호도 분석을 통한 개인화된 서비스를 제공하기 위한 방법으로 63가지 식품섭취빈도 값의 유사도에 따른 당뇨병 발생 위험도 분석 모델을 개발하였다. 국가통계 원시자료에 기반으로 한 분석 참조 데이터와의 식품섭취빈도 유사도 가중치를 구하고, 이에 따른 당뇨병 발병의 위험도를 평가하여 개인의 식품섭취빈도와 당뇨병의 관련성을 분석하고자 한다.

## 2. 관련연구

### 2.1 식품섭취빈도

식생활과 건강의 관련성에 대한 연구의 접근으로 영양 섭취 실태 및 식품섭취에 관한 연구가 진행되어왔다. 특정 연령군이나 집단의 영양소 섭취의 경향성에 대한 연구와 특정 영양소 분석을 목적으로 식사의 질에 대한 평가 연구들이 진행되어왔다[3,4]. 이러한 연구가 개별 영양소의 영양상태에 대한 연구는 가능하지만 음식섭취를 통한 실제 영양상태 분석에 한계가 있다[5]. 그래서 식품섭취패턴의 분석을 통한 음식과 식품의 영향을 분석하는 것이 필요하다[11].

식품섭취패턴에 대한 분석 중 식품섭취빈도를 통한 분석은 1947년 Burke[12]가 개발한 식이역사법에 포함된 식품섭취조사표가 구조화 되어 개인의 절대적인 섭취량을 측정하기 보다는 개인간 차이를 이용하여 식이섭취를 비교하기 위해 활용된다. 또한 식품섭취빈도조사는 개인의 장기간의 일상생활의 식이를 잘 반영하는 도구로 역학연구에 사용되고 있다[13]. 식품섭취빈도를 활용한 패턴분석의 역학연구를 당뇨병의 유병율을 분석하는 연구[10] 통해 공공보건학적 연구에 치중되어 있다. 하지만 아직 이를 이용한 개인적 분석이 가능한 연구는 거의 없

다.

식품섭취빈도 연구의 국내의 자료구조는 국민건강영양조사에서 수집된 식품섭취빈도조사 원시자료를 기반으로 한다. 국민건강영양조사의 식품섭취빈도조사는 식품섭취내용을 24시간 회상법을 통해 식품섭취 횟수를 63가지 항목의 식품에 대한 식품섭취조사를 실시하였다. 식품섭취빈도조사에서 활용된 63가지 식품 항목은 곡류, 두류, 서류, 육류, 난류, 생선류, 채소류, 해조류, 과일류 등 다양한 식품군의 조사를 가능하게 하였다[14].

### 2.2 유사도 가중치(Similarity Weight)

유사도 가중치에 관한 연구는 경영학이나 마케팅 분야에서 고객의 선호도에 따른 다음 소비를 유추하기 위한 연구로 진행되어왔다[15-17]. 유사도 계산방법으로 피어슨 상관계수(Pearson Correlation), 코사인 계수(Cosine coefficient), 벡터 내적의 곱 (Inner product), 카이 제곱 검정(Chi-square Test) 등 다양한 방법이 있으나, 최근에는 대용량 데이터의 분석에 뛰어난 성능의 피어슨 상관계수를 주로 사용한다고 알려져 있다[18].

피어슨 상관계수를 이용한 유사도 가중치 연구는 GroupLens의 독자 선호도의 유사도에 따른 뉴스 제안 시스템에서 처음 사용되었고[19], 이후 다양한 영역의 선호도에 따른 유사도 측정을 위한 연구들에서 사용되고 있다. 개인의 감성 및 상황을 고려한 음악 추천시스템[15]과 고객의 선호도에 따른 상품추천 시스템[17], 메이크업 추천 시스템[20]등 다양한 분야에서 활용하고 있다. 최근 이러한 유사도 분석을 이용한 통증을 예측하는 연구를 통해 유사한 증상에 따른 신규환자의 통증도를 예측할 수 있는 연구 또한 진행되고 있다[21]. 이렇게 개인의 상태 변수에 따른 유사도를 피어슨 상관계수로 계산하여, 해당값을 유사도 가중치로 산정할 수 있다.

본 논문에서는 개인의 식품섭취빈도를 국민건강영양 조사를 통해 얻은 식품섭취빈도 데이터 구조와 동일하게 입력받아, 개인 데이터와 분석 참조 데이터를 피어슨 상관계수를 통해 유사도 가중치를 산출하고, 그에 따른 개인의 당뇨병 발생 위험도를 평가하였다.

## 3. 연구방법

### 3.1 식품섭취빈도 데이터 및 분석 파라미터

#### 3.1.1 데이터 셋(Data Set)

질병관리본부에서 공개한 2007~2011년 국민건강영양조사 원시자료를 기반으로 데이터를 구성하여 개인 식품

섭취빈도를 분석할 수 있는 모델에 활용하였다. 국민건강영양조사는 매년 실시하는 국가 통계자료로서 국민의 대표성을 가질 수 있는 표본을 추출하여 국민 건강의 지표를 제공하는 자료로 활용된다. 이러한 표본으로 추출된 데이터를 기반으로 개인의 식품섭취빈도에 따른 질병 발생 위험도를 분석하는 자료로 사용하고자 한다. 대상자는 만 12세 이상의 국민건강영양조사 영양조사에 참여한 대상으로 결측 대상자를 제외한 제4기 1차년도(2007)의 2,928명, 2차년도(2008)의 5,383명, 3차년도(2009)의 6,000명, 제5기 1차년도(2010)의 4,326명, 2차년도(2011)의 4,902명의 자료를 통합하여 23,539명의 데이터를 포함한다.

국민건강영양조사의 원시자료에서 섭취빈도 평가를 위한 변수는 63항목의 식품섭취빈도와 당뇨병 유병여부를 확인하기 위한 당뇨병 유병여부변수(HE\_DM)를 이용하였다.

변수 구성은 식품섭취빈도의 결과 63가지 식품에 대해서 1주 1회에 1점을 기준으로 하여 「1일 3회」 21점, 「1일 2회」 14점, 「1일 1회」 7점, 「1주 4~6회」 5점, 「1주 2~3회」 2.5점, 「1주 1회」 1점, 「1달 2~3회」 0.55점, 「1달 1회」 0.22점, 「1년 1~6회」 0.07점, 「거의없음」 0점으로 10개 단계로 나누어 점수화 하여 식품의 섭취빈도를 계량화 하였고, 당뇨병 유병여부 변수(HE\_DM)의 1: 정상, 3: 당뇨병으로 진단을 분류하였다.

### 3.1.2 개인식품섭취빈도 (PFF, Personalized Food Frequency)

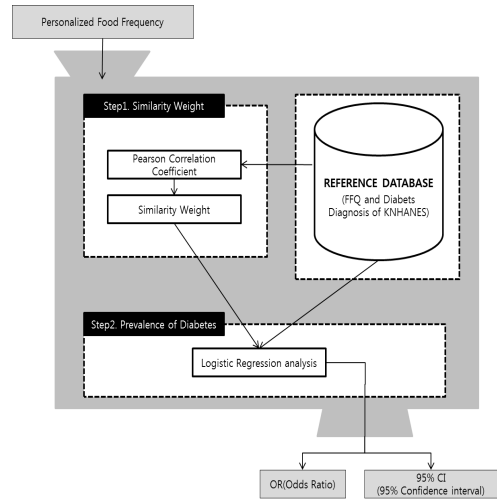
모델에 사용되는 PFF는 분석을 요청하는 개인의 식품섭취빈도 데이터이다. 개인이 63가지 식품섭취빈도를 입력하여 분석할 수 있다. 식품섭취빈도는 일상생활의 섭취패턴을 분석하기 위한 데이터이므로, 특별한 상황(여행, 질환 등)으로 인한 식품 섭취기간은 제외하였다.

### 3.1.3 분석 참조 데이터 (Reference Database)

PFF와의 유사도 가중치를 계산하는 참고자료이다. 분석 참조 데이터의 각 레코드별 유사도 가중치 산정을 통해 유사도 가중치가 높은 대상자부터 정렬하여 유사도 가중치에 따른 당뇨병 질병발생 위험도를 산출할 수 있도록 하는 데이터이다. 데이터는 국민건강영양조사 제4기 2차년도(2008)부터 제5기 2차년도(2011)까지 4개년도 데이터를 분석 참조 데이터로 구성하여 총 20,611명의 데이터와 PFF가 비교 분석 될 수 있도록 하였다.

## 3.2 당뇨병발생위험도 분석 모델 개발

Fig. 1은 당뇨병발생 위험도 분석을 위한 모델 프로세스이다. 입력된 개인 식품섭취빈도를 분석 참조 데이터와 식품섭취빈도 유사도가중치를 산출하여 사분위수로 그룹화 한뒤 로지스틱 회귀분석을 통해 당뇨병발생 위험도를 산출한다.



[Fig. 1] Diabetes Risk Analysis Model

### 3.2.1 식품섭취빈도 유사도 가중치

국민건강영양조사 원시자료의 식품섭취빈도로 구성된 분석 참조 데이터 각각과 개인을 통해 얻은 식품섭취빈도 자료의 유사 정도를 유사도 가중치를 통해 계량적으로 산출할 수 있다. 유사도 가중치 산정을 위해 피어슨 상관계수(Pearson Correlation Coefficient)를 이용하였다. 다음의 식 (1)은 선호도 유사 정도를 나타내는 유사도 가중치인 피어슨 상관계수 식이다.

$$r_{UJ} = \frac{\sum (U - \bar{U})(J - \bar{J})}{\sqrt{\sum (U - \bar{U})^2 \sum (J - \bar{J})^2}} \quad (1)$$

$r_{UJ}$ 는 개인  $J$ 와 분석 참조 데이터 내 임의의 대상자  $U$ 의 식품섭취빈도의 유사도가 산정된 값이며, 피어슨 상관계수이다. 이를 두 대상간의 유사도 가중치로 이용한다.  $U$ 는 분석 참조 데이터 내 임의의 대상자의 식품섭취빈도이고,  $\bar{U}$ 는 국가통계 자료 내 임의의 대상자의 식품섭취빈도 평균이다.  $J$ 는 평가하고자 하는 개인의 식품섭취빈도이고,  $\bar{J}$ 는 개인의 식품섭취빈도의 평균이다. 피어슨 상관계수 식을 이용하여  $J$ 와  $U_n (n = 1 \dots 20,611)$

의 유사도 가중치를 산정할 수 있다. 분류를 통해 유사도 가중치의 값에 따른 정렬을 통해 유사도 순위를 정할 수 있다.

### 3.2.2 당뇨 발생 위험도

식품섭취빈도 유사도 가중치에 따른 당뇨병 발생 위험을 평가는 개인의  $r_{Uj}$ 의 유사도가중치에 따라 4분위수로  $U$ 의 유사도 가중치 순위 그룹을 구성하여 당뇨병 유병 여부를 로지스틱회귀분석(Logistic Regression analysis)을 이용하여 오즈비(Odds Ratio, OR)를 산출하고, 95% 신뢰구간(95% Confidence interval, 95% CI)으로 유의성을 검증하여 산정할 수 있다. 당뇨 발생 위험도 평가를 위한 로지스틱 회귀분석은 MySQL 데이터베이스로 구성된 데이터와 연계하여 R 3.0을 이용하여 통계 처리하였다.

## 3.3 모델 유의성 검증

### 3.3.1 연구대상

한국인 국민건강영양조사 제4기 1차년도(2007) 자료를 이용하여 모델 성능평가를 위한 샘플을 추출하였다. 이 중 식품섭취빈도 결측 대상자를 제외하였고 기존의 당뇨 유병으로 인한 수정된 식품섭취 성향을 보이는 대상자를 제외하기 위해 검진설문을 통해 과거 당뇨병 여부, 현재 당뇨병 여부, 당뇨병 의사진단 여부가 없는 대상으로 연구를 수행하였다. 기존 당뇨병 이력이 없으나 검진조사 결과 당뇨병으로 진단된 52명의 비교군으로 추출하였다. 대조군은 당뇨병 이력이 없고 검진조사에서도 당뇨병 진단을 받지 않은 52명의 샘플을 비교군과 비슷한 연령을 조건으로 무작위 추출하여 분석하였다. 두 그룹의 성별과 연령은 표1과 같고, 두 그룹의 연령과 성별에 관한 독립성 여부 검증에서 두 집단의 독립성이 기각되어 연령과 성별의 영향을 배제할 수 있음을 확인하였다.

추출된 샘플을 본 연구의 모델에 적용하여 질병발생 위험도를 도출하였다. 두 그룹간의 질병발생 위험도에 대한 t-검정을 실시하여 유의성을 확인하였다.

[Table 1] General characteristics of diabetes and normal sample groups

	Diabetes (n=52)	Normal (n=52)	t-value	Sign.
Male <sup>1)</sup>	22(42.3)	17(32.7)	1.008	0.316
Female <sup>1)</sup>	30(57.5)	35(67.3)		
AGE <sup>2)</sup>	57.1±13.2	58.0±9.4	0.384	0.702

<sup>1)</sup> n (%), <sup>2)</sup> Mean±SD

### 3.3.2 자료분석 및 통계처리

당뇨병 발생 위험도 분석 모델을 통해 산출된 오즈비의 유의성 검정을 위한 실험의 통계처리는 IBM SPSS Statistics 21을 이용하여 분석하여 비교군과 대조군의 오즈비 분석 결과의 유의성 여부를 검증하였다. 모든 결과의 통계적 유의성은  $p < 0.05$  기준으로 검증하였다.

## 4. 결과

### 4.1 모델 유의성 검증 결과

본 연구의 모델에 의해 도출된 4분위수로 나뉜 그룹의 식품섭취빈도의 유사도에 따른 당뇨병 발생 위험도 산출 후 비교군과 대조군의 독립성에 대한 유의성 검정을 실시하였다. 표2와 같이 유사도 가중치에 따라 4분위수로 구분된 그룹 Q1, Q2, Q3, Q4는 유사도 가중치 순위가  $Q1 < Q2 < Q3 < Q4$ 이고, Q1은 로지스틱 회귀분석의 기준값이 된다. Q2의 오즈비 평균은 비교군에서 4.24, 대조군에서 6.11으로 대조군에서 더 높은 수치가 나왔으나 유의하지 않았다. Q3의 오즈비 평균은 비교군에서 6.36, 대조군에서 5.33으로 당뇨군에서 1.03배 높은 결과가 나왔으나 유의하지 않았다. 유사도가 가장 높은 그룹으로 분류된 Q4의 오즈비 평균은 비교군(당뇨군)에서 9.18로 당뇨 발생 위험도의 평균이 Q1에 비해 9.18배 높음을 나타내고, 대조군에서는 5.78배 높음으로 당뇨군에 비해 3.4배 낮은 당뇨발생위험도 평가수치로 유의한 결과를 보였다 ( $p=0.046$ ).

[Table 2] T-test for 4 groups of odds ratio

	Quantile-1 (Q1)			Quantile-2 (Q2)			Quantile-3 (Q3)			Quantile-4 (Q4)		
	Mean of OR (±SD)	t	Sign.	Mean of OR (±SD)	t	Sign.	Mean of OR (±SD)	t	Sign.	Mean of OR (±SD)	t	Sign.
Diabetes	1 (REF.)	-	-	4.24(±4.33)	1.68	.097	6.36(±9.69)	-0.62	.534	9.18(±10.18)	-2.03	.046*
Normal	1 (REF.)			6.11(±6.80)			5.33(±6.83)			5.78(±6.52)		

\*:  $p < 0.05$

## 5. 고찰 및 결론

본 연구에서는 식품섭취빈도의 유사도 가중치에 따른 당뇨병 발병 위험도를 평가하여 개인의 식품섭취빈도와 당뇨병의 관련성을 분석할 수 있는 모델을 개발하여 평가하였다.

개인의 식품섭취빈도에 따른 당뇨발생 위험을 분석을 위해 분석에 기반이 되는 분석 참조 데이터의 구성은 2008년도부터 2011년도까지의 국민건강영양조사 자료를 기반으로 결측치를 제외한 원시데이터를 이용하여 국민건강영양조사 표본 설계 시 고려한 국민의 대표성 훼손을 최소화 하여 개인의 식품섭취빈도의 성향을 국민 전체와 비교하여 평가함을 목적으로 하였다. 개인의 식품섭취빈도를 분석용 데이터베이스 20,611명과 전수비교를 하여 개인과 각 분석 참조 데이터의 유사도 가중치를 산출한다. 유사도 가중치는 피어슨 상관계수를 이용하여 분석한다. 개인과 분석 참조 데이터 각각의 식품섭취빈도 상관계수를 산출하여 유사도가중치를 수치화 하여 유사도 가중치에 따라 정렬하여 순위를 갖도록 하였다. 유사도 가중치 산출에 이용된 피어슨 상관계수는 대량의 데이터 분석에 유용한 장점이 있고, 비교하고자 하는 데이터가 이미 정형화되어있어 개인의 식품섭취빈도 입력을 받아 분석하기에 적절한 유사도 가중치 산출방법이다.

피어슨 상관계수를 통해 산출된 유사도 가중치는 순위에 따른 분류는 사분위수로 구분하여 4개의 그룹으로 나누어 그룹 간 당뇨병 유병여부를 이용한 로지스틱 회귀 분석을 이용하여 당뇨병 발생 위험도를 산출을 통해 식품섭취빈도의 유사도가중치에 따른 당뇨병 발생 위험도를 평가하는 모델로 구성하였다.

본 연구의 모델의 유의성을 평가하기 위하여 2007년도 국민건강영양조사 자료 중 당뇨병 유병 여부에 따른 비교군과 대조군을 추출하여 모델을 통한 당뇨발생 위험도를 비교하였다. 2007년도 자료 중 비교군 52명은 2007년도 설문조사에서 기존의 당뇨병 유병 여부가 없으며 당뇨병 진단 경험이 없는 대상자 중에서 검진조사에서 당뇨병 진단을 받은 대상자로 기존에 당뇨병에 대한 문제의식을 갖지 않은 대상자로 선택하여, 분석 시 당뇨병 인식으로 인한 수정된 식품 섭취 경향을 갖고 있는 대상을 제외하고 선택하여 식품섭취빈도 비교에 외부영향을 최소화 하였다. 대조군 52명은 추출된 비교군 52명의 연령 평균과 유사한 대상으로 선택하여 연령으로 인한 당뇨병 발생 영향을 최소화 하였다. 두 집단의 104 케이스를 사분위수로 그룹화한 후 분석하여 당뇨병 발생 위험도를 산출하였다. 유사도가중치가 가장 높은 Q4그룹의 산출된 당뇨병 위험도는 비교군 (9.18±10.18)에서 대조군

(5.78±6.52)에 비해 평균이 높게 나왔고 두 집단에 대한 t-검정 결과 그 유의성을 확인할 수 있었다. 이러한 결과는 당뇨병 유병 대상자의 식품섭취빈도를 이용한 당뇨병 유병을 평가 연구[22]에서와 같이 당뇨병 질환자의 식품섭취 경향성에 따른 발병 위험도가 높아진다는 결과로, 기존의 연구와 일치하는 결과이다. 식품별 섭취빈도 평균의 분석을 통해 일반적인 결과보다 각 대상자의 개별분석 수행을 통해 개인 맞춤형 정보제공에 적합한 형태의 모델을 제시하였다. 개별적 분석을 통해 실제 당뇨군에서의 당뇨병 발생 위험도가 높게 평가된다는 것을 확인할 수 있어 식품선호도와 당뇨병의 관련성을 확인할 수 있다. 두 집단의 당뇨 발생 위험도의 크기의 차이가 유의하게 나타났으나, 대조군에서도 5.78 이라는 당뇨병 발생 위험도가 발생한 것을 확인할 수 있다. 이것은 대조군으로 추출된 대상자의 연령의 평균이 58세로 높은 연령대라는 것을 감안하여 식품섭취빈도 이외의 다양한 외부적 요인이 영향을 미친다는 고려할 수 있다. 하지만 비슷한 연령대에서 분석한 결과 식품섭취빈도에 따른 더 높은 당뇨발생 위험도가 발생한다는 것을 통해 식품섭취빈도가 당뇨발생위험도에 유의한 영향을 미친다는 것을 확인할 수 있다. 그 이외의 그룹에서는 유의한 결과가 나오지는 않았으나 Q3 그룹에서는 비교군에서 대조군에 비해 1.02배 높은 값을 볼 수 있었다. Q2의 경우는 대조군에서 비교군에 비해 더 큰 결과가 도출되어 대조군 그룹 중에서 Q2가 가장 높은 값을 보여 정상인 군에서 유사도 가중치가 낮은 집단에서 당뇨병 발생 위험도가 높게 나오는 경향을 확인할 수 있었다.

개인의 식품섭취빈도를 국가 통계데이터를 기반으로 비교하여 개인의 당뇨발생 위험도에 관한 분석하는 연구는 기존의 보건 역학 연구에서 국가 통계 데이터를 활용하던 것의 범위를 확대하여 활용할 수 있는 연구 방법이다. 보건 역학 연구는 특정 연령대의 질병 발생 위험도나 특정 집단의 질병 발생 위험도를 분석하여 공개함으로써 건강에 관한 포괄적인 통계 정보를 제공하여왔다. 하지만 본 연구를 통해 국가 통계 데이터를 특정 집단에 대한 분석이 아닌 개인에 관한 분석에 이용할 수 있는 방법을 제시함으로써 기 마련된 데이터를 통해 개인의 건강을 분석하여 개인의 성향에 따른 질병발생 위험도를 제공하는 연구임에 의의가 있다.

본 연구에서는 식품섭취빈도 변수 63가지를 훼손하지 않고 원형 그대로 비교하는 방법으로 분석을 시도하였다. 하지만 식품섭취빈도의 정제를 통해 더 유의한 결과 도출을 지속적으로 연구해야 할 것으로 사료되며 향후 더 유의한 검증을 위해 인구통계학적 정보 및 특정 질병과 관련됨이 밝혀진 요인을 추가 분석함으로써 개인의 질병

발생위험도에 대한 더 정밀한 분석이 가능할 것으로 기대된다. 또한 모델의 평가에 사용된 104 샘플은 당뇨병 일지 여부를 통한 샘플링 과정으로 인해 높은 연령군의 샘플이 형성되었으나, 향후 다양한 연령에서의 유의성 여부에 관한 분석을 접근해야 한다. 본 연구가 추적 연구를 통한 분석이 아닌 단면 연구로 시간의 변화의 영향을 분석할 수 없다는 한계가 있으나, 통계 데이터의 양을 늘려 분석에 왜곡을 최소화하고자 하였다.

본 연구는 개인의 식품선호도에 따른 당뇨병 발생 위험도를 평가하는 것으로, 기존의 선호도 분석방법과 보건역학 분석방법을 융합한 모델을 제시하고, 그 타당성 분석을 수행하였다. 이러한 대규모 데이터를 이용한 개인의 성향을 분석하고, 질병 발생 위험성 여부를 분석하는 연구는 기존에 시도되지 않았던 새로운 시도였으며, 스마트폰이나 웹을 이용한 개인건강관리시스템 접목함으로써 개별적 건강관리 방법을 제시할 수 있을 것으로 기대된다.

## References

- [1] Korea National Statistical Office, "Annual report on the cause of death statistics, 2010." Statistics Korea, 2012.
- [2] H. S. Chung, J. A. Seo, S. G. Kim, N. H. Kim, D. M. Kim, C. H. Chung, and D. S. Choi, "Original Articles : Relationship Between Metabolic Syndrome and Risk of Chronic Complications in Koreans with Type 2 Diabetes," *Korean Diabetes J*, vol. 33, no. 5, pp. 392-400, 2009.  
DOI: <http://dx.doi.org/10.4093/kdj.2009.33.5.392>
- [3] Y. G. Park, E. M. Lee, C. S. Kim, J. H. Eom, J. A. Byun, N. K. Sun., "Survey on the Content and Intake Pattern of Sugar from Elementary and Middle School Food services in Daejeon and Chung-cheong Province," *JOURNAL OF THE KOREAN SOCIETY OF FOOD SCIENCE AND NUTRITION*, vol. 39, no. 10, pp. 1545-1554, OCT, 2010.  
DOI: <http://dx.doi.org/10.3746/jkfn.2010.39.10.1545>
- [4] W. H. Jeong, B. K. Jin, and E. H. Hwang, "Prevalence of Metabolic Syndrome and Assessment of Food · Nutrient Intakes among Adult Visitors of a Public Health Center in Korea," *JOURNAL OF THE KOREAN SOCIETY OF FOOD SCIENCE AND NUTRITION*, vol. 41, no. 2, pp. 205-212, 2012.  
DOI: <http://dx.doi.org/10.3746/jkfn.2012.41.2.205>
- [5] Y. J. Bae, "Evaluation of nutrient and food intake status, and dietary quality in Korean female adults according to obesity : Based on 2007-2009 Korean National Health and Nutrition Examination Survey," *The Korean Journal of Nutrition*, vol. 45, no. 2, pp. 140-149, 2012.  
DOI: <http://dx.doi.org/10.4163/kjn.2012.45.2.140>
- [6] K. W. Paek, K. H. Chun, and S. J. Lee, "A Factor of Fasting Blood Glucose and Dietary Patterns in Korean Adults Using Data From the 2007, 2008 and 2009 Korea National Health and Nutrition Examination Survey," *J Prev Med Public Health*, vol. 44, no. 2, pp. 93-100, 2011.  
DOI: <http://dx.doi.org/10.3961/jpmph.2011.44.2.93>
- [7] P. K. Newby, and K. L. Tucker, "Empirically Derived Eating Patterns Using Factor or Cluster Analysis: A Review," *Nutrition Reviews*, vol. 62, no. 5, pp. 177-203, 2004.  
DOI: <http://dx.doi.org/10.1111/j.1753-4887.2004.tb00040.x>
- [8] M. B. Schulze, K. Hoffmann, A. Kroke, and H. Boeing, "An approach to construct simplified measures of dietary patterns from exploratory factor analysis," *British Journal of Nutrition*, vol. 89, no. 3, pp. 409-418, 2003.  
DOI: <http://dx.doi.org/10.1079/BJN2002778>
- [9] Y. Ahn, Y. J. Park, S. J. Park, H. Min, H. K. Kwak, K.S. Oh, and C. Park, "Dietary Patterns and Prevalence Odds Ratio in Middle-aged Adults of Rural and Mid-size City in Korean Genome Epidemiology Study," *The Korean Journal of Nutrition*, vol. 40, no. 3, pp. 259-269, 2007.
- [10] R. M. van Dam, E. B. Rimm, W. C. Willett, M. J. Stampfer, and F. B. Hu, "Dietary Patterns and Risk for Type 2 Diabetes Mellitus in U.S. Men," *Annals of Internal Medicine*, vol. 136, no. 3, pp. 201-209, 2002.  
DOI: <http://dx.doi.org/10.7326/0003-4819-136-3-200202050-00008>
- [11] M. Yon, M. Lee, S. Oh, S. Park, C. Kwak, "Assessment of Food Consumption, Dietary Diversity and Dietary Pattern during the Summer in Middle Aged Adults and Older Adults Living in Gugoksoondam Longevity Area, Korea," *Korean Journal of Community Nutrition*, vol. 15, no. 4, pp. 536-549, 2010.
- [12] B. S. Burke, "The dietary history as a tool in research," *J Am Diet Assoc*, vol. 23, no. 12, pp. 1041-1046, 1947.
- [13] W. Willett, "Nutritional epidemiology," Oxford University Press, 2012.
- [14] Korea Centers For Disease Control & Prevention, "Guideline: The Fifth Korea National Health and

- Nutrition Examination Survey (KNHANES V-2)“, 2012.
- [15] H. Choi, H. Bae, J. Seo, and K. Yoon, "Music Recommendation System Using Extended Collaborative Filtering Based On Emotion & Context Information Fusion.," in Proceedings of Korea Society of Broadcast Engineers, pp. 82-84. 2011.
- [16] H. J. Kwon, D. K. Chung, and K. S. Hong, "A Multimedia Recommender System Using User Playback Time," Journal of Korean Society for Internet Information, vol. 10, no. 1, pp. 111-121, 2009.
- [17] E. H. Jeong, and B. K. Lee, "A Design of HPPS(Hybrid Preference Prediction System) for Customer-Tailored Service," JOURNAL OF KOREA MULTIMEDIA SOCIETY, vol. 14, no. 11, pp. 1467-1477, NOV, 2011.
- [18] G. Adomavicius, and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," Knowledge and Data Engineering, IEEE Transactions on, vol. 17, no. 6, pp. 734-749, 2005.
- [19] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp. 175-186, 1994.
- [20] K. Y. Chung, "Human Sensibility Ergonomics Makeup Recommendation System using Context Sensor Information," JOURNAL OF THE KOREA CONTENTS ASSOCIATION, vol. 10, no. 7, pp. 23-30, JUL, 2010.
- [21] H. Jung, H. Yoo , K. Y. Chung, and Y. H. Lee, "Performance Analysis of Intelligence Pain Nursing Intervention U-health System," JOURNAL OF THE KOREA CONTENTS ASSOCIATION, vol. 13, no. 4, pp. 1-7, APR. 2013.
- [22] S. H. Jeon, N. H. Kim, "Prevalence of Diabetes and Impaired fasting glucose according to Food frequency Similarity in Korea," Journal of The Korea Academia-Industrial cooperation Society, vol.14, no.2, pp. 751-758, 2013.

**전 소혜(So-Hye Jeon)**

[준회원]



- 2008년 2월 : 연세대학교 의용전자공학과 (공학사)
- 2008년 9월 ~ 현재 : 연세대학교 생체공학협동과정 석박사통합과정

<관심분야>

의료정보시스템, PHR, 유헬스, 데이터마이닝

**김 남 현(Nam-Hyun Kim)**

[정회원]



- 1982년 2월 : 연세대학교 대학원 전기공학과 (공학석사)
- 1986년 2월 : 연세대학교 대학원 전기공학과 (공학박사)
- 2008년 9월 ~ 2012년 8월 : 연세의료원 의료정보실장
- 2011년 3월 ~ 현재 : 연세의대 의학공학교실 주임교수

<관심분야>

의료정보시스템 개발 및 표준화, 유헬스, PHR 등