

## 혈청 대사체와 뇌졸중 발생위험의 용량반응 분석

지연호<sup>1</sup>, 정금지<sup>1</sup>, 임연화<sup>2,3</sup>, 이예승<sup>4</sup>, 박영자<sup>4</sup>, 지선하<sup>1</sup>

<sup>1</sup>연세대학교 보건대학원 국민건강증진연구소, <sup>2</sup>서울대학교 의과대학 환경보건센터, <sup>3</sup>서울대학교 의학연구원 환경의학연구소, <sup>4</sup>고려대학교 약학대학

### Dose-response Relationship between Serum Metabolomics and the Risk of Stroke

Yon Ho Jee<sup>1</sup>, Keum Ji Jung<sup>1</sup>, Youn-Hee Lim<sup>2,3</sup>, Yeseung Lee<sup>4</sup>, Youngja Park<sup>4</sup>, Sun Ha Jee<sup>1</sup>

<sup>1</sup>Department of Epidemiology and Health Promotion, Institute for Health Promotion, Yonsei University, Seoul; <sup>2</sup>Environmental Health Center, Seoul National University College of Medicine, Seoul; <sup>3</sup>Institute of Environmental Medicine, Seoul National University Medicine Research Center, Seoul;

<sup>4</sup>College of Pharmacy, Korea University, Sejong, Korea

**Objectives:** Except the known risk factors for stroke, few studies have identified novel metabolic markers that could effectively detect stroke at an early stage. In this study, we explored the dose-response relationship between serum metabolites and the incidence of stroke. **Methods:** We studied 213 adults in the Korean Cancer Prevention Study-II (KCPS-II) biobank and estimated dose-response relationship between serum metabolites and stroke (42 cases and 171 controls). Three serum metabolites (Acetylcholine, HexadecylAcetylGlycerol, and 1-acetyl-2-formyl-sn-glycero-3-phosphocholine) were used in this study. The analysis included (1) exploratory nonlinear analysis, (2) estimation of flexion points and slopes at below and above the points. In the model to estimate risk of incidence of stroke, we controlled for conventional risk factors such as age, sex, systolic blood pressure, type 2 diabetes, triglyceride, and smoking status. **Results:** The relationship between incidence of stroke and log-transformed 1-acetyl-2-formyl-sn-glycero-3-phosphocholine was non-linear with flexion point around intensity score of 8.8, whereas other metabolites, log-transformed Acetylcholine and HexadecylAcetylGlycerol, showed negative linear patterns. **Conclusions:** The study suggests that metabolic markers are associated with incidence of stroke, particularly, at or above the flexion point. The study result may contribute to developing a novel system for precise stroke prediction.

**Key words:** Cohort study, Dose-response, Metabolites, Stroke

## 서론

세계적으로 심뇌혈관질환은 사망원인 1위를 차지하고 있고 2012년에 1,750만 명이 심뇌혈관질환으로 사망하였으며 전체 사망의 31%를 차지하였다[1]. 또한 심뇌혈관질환 사망 중에서 절반 이상을 차지하고 있는 것이 뇌졸중 사망이다. 우리나라의 경우 2012년 순환기계통의 사망자 58,864명 중에서 뇌졸중으로 인한 사망이 25,725명이었다[2]. 이는 2012년 전체사망자 264,797명의 약 10%에 해당하는 수이다. 즉, 우리나라에서 10명 사망 중 1명은 뇌졸중으로 사망할 정도로 매우 높은

수준이다.

뇌졸중 위험요인으로 잘 알려진 것은 흡연, 식이습관, 비만, 운동부족, 당뇨병 등이 있다[3,4]. 그동안 이들 요인들은 뇌졸중의 예방에 많은 기여를 하여 왔던 것은 사실이지만, 여전히 뇌졸중의 조기발견의 필요한 새로운 인자, 즉 유전인자나 대사체(metabolite)에 대한 연구는 계속 필요로 하고 있다[5,6]. 더구나 매년 여전히 많은 수의 뇌졸중이 발생하고 사망하는 것을 보면, 기존의 위험요인만으로는 충분한 예방과 관리가 어렵다는 것이다. 또한 이러한 위험요인들에 있어서 선형관련성만 연구한 논문은 많지만 비선형 혹은 변곡점을 찾는 연구는 드물었다.

**Corresponding author:** Sun Ha Jee

50-1 Yonsei-ro, Saedaemun-gu, Seoul 03722, Korea  
Tel: +82-2-2228-1523, E-mail: jsunha@yuhs.ac

Received: July 21, 2016 Revised: August 18, 2016 Accepted: August 24, 2016

This study was funded by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (H114C2686, H113C0715).

No potential conflict of interest relevant to this article was reported.

**How to cite this article:**

Jee YH, Jung KJ, Lim YH, Lee Y, Park Y, Jee SH. Dose-response relationship between serum metabolomics and the risk of stroke. J Health Info Stat 2016;41(3):318-323. Doi: <http://dx.doi.org/10.21032/jhis.2016.41.3.318>

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permit unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2016 Journal of Health Informatics and Statistics

따라서 이 연구는 뇌졸중이 없는 일반인구를 장기간 추적하는 코호트 연구 내에서 환자-대조군 연구설계를 통하여 밝혀진 뇌졸중 발생자와 대조군의 혈청 대사체와 뇌졸중과의 용량반응 분석을 하였다. 즉, 대사체 분석을 통해 확인된 바이오 마커에 대한 역치(flexion point)를 찾아냄으로써 임상적 유용성을 높이고자 하였다.

## 연구 방법

### 연구대상

이 연구는 한국인 암 예방연구-II (Korean Cancer Prevention Study-II, KCPS-II) 바이오뱅크 코호트에서 확인된 뇌졸중 환자를 대상으로 하였다. KCPS 바이오뱅크는 2004-2013년 서울 경기 지역 11개 종합검진센터를 방문한 내원객 중에서 연구목적에 동의하고 동의서와 혈액을 공여한 대상자 156,836명으로 구축된 전향적 코호트이다. 실제로 연구대상자 등록은 2004-2005년에 5.8%, 2006-2008년 90.9%, 그리고 2009-2013년 3.3%가 이루어졌다. 연령을 20에서 84세로 제한할 경우 연구 대상자는 156,701명이었다. 이 연구설계는 기본적으로 Case-cohort 연구로서 156,701명 중에서 무작위 추출한 서브코호트 5,000명과 뇌졸중 발생자 995명을 포함하여 5,995명을 한국인 칩(Korean chip) 검사 대상으로 정하였다. 5,995명 중에서 연령의 혼란효과를 최소화하기 위해 58세에서 69세로 제한하였고, 흡연력 정보가 있는 대상자 410명을 뇌졸중 대사체 분석 대상으로 선정하였다. 이 중에서 용량반응 분석을 하고자 하는 뇌졸중 대사체 정보 세 가지가 모두 측정된 자 213명을 최종 분석 대상으로 하였다. 이 연구는 연세대학교 임상시험윤리위원회 심의위원회(institutional review board)에 심의를 거쳐 승인을 받았다(4-2014-1008). 연구자는 본 연구의 목적과 내용에 대해 피실험자에게 상세히 설명을 하고 연구 참여 동의서를 받은 후 연구를 시작하였다.

### 연구에 사용된 변수

뇌졸중의 기존의 위험요인으로 연령, 수축기 그리고 이완기 혈압, 고혈압, 공복 혈당, 당뇨병, 그리고 흡연력을 포함하였다. 이 연구에서 고혈압은 수축기 혈압 140 mmHg 이상 혹은 이완기 혈압 90 mmHg 이상 혹은 고혈압 약 복용을 한 경우로 정의하였다. 당뇨병은 공복혈당 126 mg/dL 이상 혹은 당뇨병 약 복용을 한 경우로 정의하였다. 흡연력은 비흡연, 과거흡연, 현재흡연으로 분류하였다.

### 대사체 분석

대사체 분석은 정상인(348명) thrombotic stroke 환자(62명)의 혈청을 Ultra Performance Liquid Chromatography system (Agilent 1260 In-

finity Quaternary, Agilent, Santa Clara, CA, USA)와 결합된 Agilent LC-MS/MS Q-TOF 6550 (Agilent, Santa Clara, CA, USA) mass spectrometer로 분석 후 mass spectrum을 apLCMS로 변환시켜 각각 m/z value와 intensity를 얻었다. 직교부분 최소자승 판별 분석(OPLS-DA)으로 두 그룹의 차이를 식별하여 그룹으로 분리하였으며 각 그룹의 95% 범주 내에 속하며 상위 5% 내에 들어간 403개의 대사체를 선정하였다. p-value ( $\leq 0.05$ ) filtering 후에 fold change는 case와 control의 intensity가 두 배 이상 차이가 나는 m/z를 filtering하여 45개의 대사체를 얻었다. 다음 단계로 Top 20개 대사체는 앞에서 45개의 m/z를 metlin database에 검색하여 대사체를 식별하여 선정하였다. 이 연구를 위해서는 상위 20개 대사체 중에서 뇌졸중과 관련성이 높은 세 개의 대사체에 대해서만 용량반응 분석을 실시하였다.

### 뇌졸중 진단

이 연구팀은 연세대학교와 국민건강보험공단과 공동연구를 체결하였고, 대상자 개인의 상병 자료 사용을 동의한 개인정보활용 동의서를 제출한 경우에 한해서 연구자료로 활용하였다. 이 연구는 연구시작 이후 2010년까지 새롭게 진단된 뇌졸중을 포함하였다. 즉 2010년까지 발생된 뇌졸중 환자에 대해서 개인정보활용 동의서를 제출하고 뇌졸중 진단을 확인하였다. 2011년 이후는 개인정보보호 강화로 인해 뇌졸중 진단 확인을 할 수 없는 상태이다. 뇌졸중은 국민건강보험공단의 청구 자료에 있는 상병코드 International Classification of Disease (ICD)-10 code (I60-I69)에 해당하는 경우 뇌졸중으로 진단하였고, 그 중에서 첫 번째 입원에 해당하는 경우를 본 연구에서 뇌졸중 발생으로 정의하였다. 국민건강보험공단의 뇌졸중 상병코드에 대한 정확도 연구는 2000년 Park et al. [7]에 의해서 수행되었으며, 당시 뇌졸중 상병코드의 83%가 정확한 것으로 보고된바 있다.

### 통계분석방법

1단계, 비선형성(non-linearity) 분석: 뇌졸중의 기존 위험요인으로 연령, 성, 당뇨병, 수축기 혈압, 중성지방, 흡연력을 포함한 Generalized additive model (GAM) [8,9]의 로지스틱 회귀모형을 통해 비선형성(non-linearity) 분석을 하였다. 이 때 뇌졸중 발생과의 선형관계 여부를 확인하기 위하여 연령, 수축기 혈압, 중성지방에 대해 평활 함수를 주었다(Model 1). 모형에서 기존 위험요인을 보정한 후 세 가지 대사체, 즉 Acetylcholine (AC), HexadecylAcetylGlycerol (HAG), 1-acetyl-2-formyl-sn-glycero-3-phosphocholine (PC)와 뇌졸중 간의 연관성을 분석하였다(Model 5). 세 가지 대사체는 정규화를 위해 로그함수로 치환하였다. 로지스틱 회귀모형은 다음과 같았다.

$$\text{logit } E(Y) = \beta_0 + \beta_1 \times \text{metabolite} + S(x_1) + \dots + S(x_p)$$

이때 종속변수 Y는 뇌졸중 발생 여부이고, metabolite는 AC, HAG, PC 중 1개씩 모델에 포함했으며,  $S(x_i)$  ( $i=1,2,\dots,p$ )는 공변량들(연령, 수축기 혈압, 중성지방)의 평활함수이며, penalized spline이 평활함수로 사용되었다[10]. 나머지 공변량들(성, 당뇨병, 흡연력)에 대해서도 추가로 보정을 해주었다. 준모수적 모형인 GAM을 통해 Akaike's Information Criterion (AIC) [11]을 최소로 하는 평활모수(estimated degrees of freedom, EDF)를 추정하였다. 위의 모형에서 대사체 수준에 따른 뇌졸중 발생 관계의 비선형성 여부를 확인하기 위해 대사체 변수에 평활함수를 사용하여 분석하였다. 이 연구에서 사용된 모수적 공변량과 평활함수를 사용한 비모수적 공변량은 다음과 같았다.

$$\text{logit } E(Y) = \beta_0 + \beta_1 \times \text{SEX} + \beta_2 \times \text{TDM} + \beta_3 \times \text{EXSMOK} + \beta_4 \times \text{CUSMOK} + S(\text{AGE}) + S(\text{SBP}) + S(\text{TG}) + S(\log(\text{AC})) + S(\log(\text{HAG})) + S(\log(\text{PC}))$$

이때 종속변수 Y는 뇌졸중 발생 여부이고, Type2 diabetes mellitus (TDM)는 제2형 당뇨병, Ex-smoker (EXSMOK)는 이전 흡연자, Current-smoker (CUSMOK) 현재 흡연자, AGE는 연령, Systolic blood pressure (SBP)는 수축기 혈압을 뜻하며, 세 개의 metabolite AC, HAG, PC와 뇌졸중 발생과의 비선형 관계는 평활함수 ( $S(\log(\text{AC}))$ ,  $S(\log(\text{HAG}))$ ,  $S(\log(\text{PC}))$ )를 사용한 그래프를 통해 확인할 수 있다. 이 단계의 통계분석은 R 프로그램(v3.2.1) (R development Core Team, <https://cran.r-project.org/>)의 mgcv package의 gam 함수를 통해 이루어졌다.

**2단계, 변곡점 전후 기울기 분석:** 일반선형 회귀모형(generalized linear model, GLM)을 통해서 변곡점을 찾고 그 전후의 기울기 방향과 크기를 확인하였다. 1단계 비선형성 분석에서 의미 있었던 변수만을 포함하였다. 특히 1단계에서 유의한 비선형성을 보인 보정 변수들은 natural cubic spline (ns)을 통해 각각의 매듭(knot) 수를 지정하여 그 방향성과 크기를 분석하였다.

역치 효과(threshold effect)를 확인하기 위해 GLM의 로지스틱 회귀 모형을 사용하였다. 보정 변수의 비선형성은 ns를 이용하였다.

$$\text{logit } E(Y) = \beta_0 + \beta_1 \times \text{metabolite} + \beta_2 \times (\text{metabolite} - \xi) + ns(x_1) + \dots + ns(x_p),$$

$$(x - \xi)_+ = \max\{x - \xi, 0\}$$

여기서, Y는 뇌졸중 발생 여부이고, metabolite는 세 개의 로그 변환한 대사체  $\log(\text{AC})$ ,  $\log(\text{HAG})$ ,  $\log(\text{PC})$  중 하나를 노출변수로 지정하고,  $ns(x_i)$ 는 노출변수를 제외한 나머지 두 로그 변환 대사체들과 연령변수에 취했으며,  $\xi$ 는 역치값이다. 이는 다시 말해 역치값 ( $\xi$ ) 전후로 기울

기가 각각 다음과 같음을 의미한다.

$$\begin{cases} \text{logit } E(Y) = \beta_0 + \beta_1 \times \text{metabolite} + ns(x_1) + \dots + ns(x_p), & \text{metabolite} \leq \xi \\ \text{logit } E(Y) = \beta_0 - \beta_2 \xi + (\beta_1 + \beta_2) \times \text{metabolite} + ns(x_1) + \dots + ns(x_p), & \text{metabolite} > \xi \end{cases}$$

즉, 이 두 직선들은 역치값 ( $\xi$ )에서 만나게 된다. 지정된 범위 내에서 탐색단위마다 격자탐색을 하여 편차를 최소로 하는 모형의 역치점 (optimum-threshold)을 찾아준다. 이 단계의 통계분석은 R 프로그램의 HEAT package의 *threspt* 함수[12]를 통해 이루어졌다. 최적 역치 탐색 분석은 EDF가 2 이상인 대사체에 대해서만 하였다.

## 연구 결과

### 연구대상의 일반적 특성

이 연구에서 뇌졸중 발생자는 42명이고 대조군은 171명이었다. 뇌졸중 발생자 중에서 허혈성 뇌졸중 환자는 18명, 출혈성 뇌졸중은 7명이었다. 연구대상자의 평균 연령은 62세(표준 편차 3.2세)이었고(남자 61.9세, 여자 62.2세), 연령범위는 58세에서 69세이었다. 수축기 혈압은 뇌졸중에서 정상대조군 보다 평균 2.8 mmg 높았으나 이완기 혈압은 두 군이 비슷하였다(Table 1).

**Table 1.** General characteristics of study participants\*

Variables	Stroke (n=42)	Healthy control (n=171)
	Mean (SD), %	Mean (SD), %
Age (y)	63.4 (3.3)	61.7 (3.1)
Systolic BP (mmHg)	128.6 (15.0)	125.8 (16.2)
Diastolic BP (mmHg)	76.6 (8.9)	76.4 (11.8)
Total cholesterol (mg/dL)	198.2 (36.8)	203.7 (37.4)
HDL cholesterol (mg/dL)	54.1 (13.3)	52.4 (12.5)
Triglyceride (mg/dL)	148.0 (88.4)	139.0 (76.3)
Fasting serum glucose (mg/dL)	99.4 (26.5)	98.2 (20.0)
Sex (female)	50.0	36.3
Diabetes	26.2	24.0
Smoker		
Past	9.5	23.4
Current	19.1	12.3
Metabolite		
log(AC)	9.2 (0.6)	9.5 (0.5)
log(HAG)	8.8 (1.1)	9.8 (1.1)
log(PC)	9.7 (0.9)	8.9 (0.7)

SD, standard deviance; BP, blood pressure; HDL, high-density lipoprotein; AC, acetylcholine; HAG, HexadecylAcetylGlycerol; PC, 1-acetyl-2-formyl-sn-glycero-3-phosphocholine.

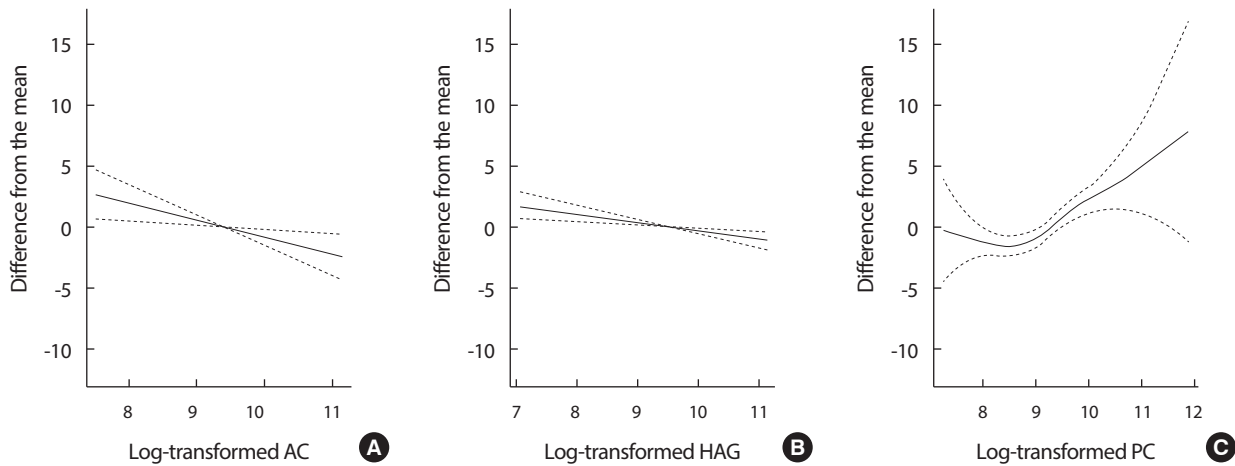
\*Data are expressed as mean (SD) unless otherwise indicated.

**Table 2.** General additive model comparison before and after adding three metabolites

Smooth terms	Model 1		Model 2		Model 3		Model 4		Model 5	
	EDF	p-value	EDF	p-value	EDF	p-value	EDF	p-value	EDF	p-value
s(log(AC))			1.00	0.001					1.00	0.007
s(log(HAG))					1.53	2.13E-05			1.00	0.001
s(log(PC))							2.28	6.11E-06	3.23	<0.000
s(AGE)	1.00	0.005	1.00	0.009	2.35	0.044	1.00	0.001	1.00	0.003
s(SBP)	1.39	0.678	1.00	0.438	5.86	0.405	1.73	0.544	1.00	0.491
s(TG)	3.70	0.356	3.79	0.368	4.42	0.301	3.88	0.139	8.29	0.056
Parametric coefficients	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept	-1.82	<.0001	-1.93	2.83E-06	-2.28	2.51E-06	-2.41	1.82E-06	-3.03	<0.000
Sex (female)	0.62	0.162	0.64	0.167	0.81	0.122	0.86	0.101	0.86	0.187
Diabetes	0.06	0.898	0.01	0.988	0.03	0.948	0.41	0.428	0.58	0.361
Past smoker	-0.63	0.322	-0.53	0.420	-0.87	0.230	-0.57	0.439	-1.12	0.222
Current smoker	0.72	0.203	0.57	0.327	0.86	0.196	0.54	0.462	0.06	0.946
Adjusted R <sup>2</sup>	0.06		0.12		0.23		0.29		0.50	

Model 1: Conventional risk factors (age, sex, systolic blood pressure, type 2 diabetes, triglyceride, and smoking status.); Model 2: Model 1+log(AC); Model 3: Model 1+log(HAG); Model 4: Model 1+log(PC); Model 5: Model 1+log(AC)+log(HAG)+log(PC); s( ): penalized smoothing term.

AC, Acetylcholine; HAG, HexadecylAcetylGlycerol; PC, PC2010U; SBP, systolic blood pressure; TG, triglyceride; EDF, estimated degrees of freedom.



**Figure 1.** Association of three metabolites with the risk of stroke incidence after controlling for age, sex, diabetes, systolic blood pressure, and triglyceride. (A) AC, Acetylcholine; (B) HAG, HexadecylAcetylGlycerol; (C) PC, 1-acetyl-2-formyl-sn-glycero-3-phosphocholine.

### 1단계 GAM을 이용한 비선형분석

Table 2의 Model 1과 같이 기존 위험요인 중에서는 연령만이 유의하였으며 선형적 관련성을 보였다. 수축기 혈압이나 중성지방은 EDF가 각각 1.4과 3.7로 변곡점 존재의 가능성을 보였지만 통계적으로 유의하지 않았다. 세 가지 대사체를 추가한 결과 연령과 log(AC)와 log(HAG)는 EDF가 1로 변곡점을 보이지 않았고, log(PC)에서는 3.2의 EDF를 보였다(model 5). 수정된 결정계수 (adjusted-R<sup>2</sup>)는 기존 위험요인만을 넣은 모형은 0.06이었고(model 1) 대사체를 추가한 모형은 0.50으로(model 5) 설명력이 크게 증가하였다.

Figure 1은 세 대사체의 뇌졸중 발생과의 관계를 보였다. 이 그림으

로부터 log(AC)와 log(HAG)는 뇌졸중 발생과 음의 선형관계를 보이는 반면 log(PC)에서는 양의 비선형 관계를 알 수 있다(Figure 1C). 또한 log(PC)는 8에서 10구간에 변곡점이 있는 것으로 보였다.

### 2단계 GLM을 이용한 변곡점 및 전후 기울기 분석

1단계에서 세 가지 대사체의 뇌졸중 여부와의 관계에 대한 비선형성을 파악하였다. 2단계에서는 뇌졸중 발생에 있어서 대사체 log(PC)의 역치 전후 (<threshold와 ≥threshold) 기울기를 추정하였다. Table 3과 Figure 2에서와 같이 log(PC)의 구간을 8에서 12로 할 경우 Optimum threshold가 8.8로 나왔으며, intensity score 8.8 이후의 기울기는

**Table 3.** The association between log-transformed PC and stroke incidence < or ≥ optimum threshold level (8.8 in log-transformed PC)<sup>1</sup>

	Estimate	SE	p-value
Intercept	6.21	7.92	0.433
< Threshold	-0.78	0.92	0.397
≥ Threshold	2.45	0.56	< 0.000
Sex (female) <sup>2</sup>	0.99	0.46	0.031

PC, 1-acetyl-2-formyl-sn-glycero-3-phosphocholine; SE, standard error.

<sup>1</sup>Optimum threshold was estimated between 8 and 12 in log-transformed PC.

<sup>2</sup>Reference of sex was male.

2.4512로 유의하였다(Table 3).

Figure 2는 다른 공변량들을 평균치로 채우고 log(PC) 변수만을 원 자료로 사용했을 때 log(PC)에 대한 뇌졸중 발생 추정치 그림이다. 즉 다른 위험요인들이 평균치일 때 Figure 1C와 마찬가지로 log(PC)의 intensity score 8.8 이후로 뇌졸중 발생이 급증함을 보여주고 있다.

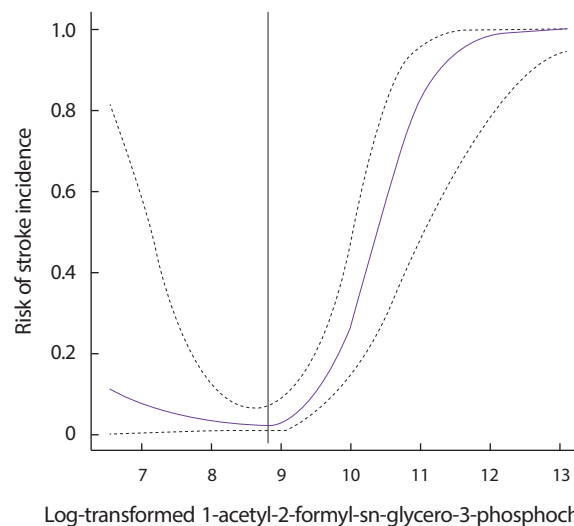
## 고찰

이 연구는 전향적 코호트 연구에서 확인된 뇌졸중 환자 42명과 코호트 내 sub-cohort에서 무작위로 뽑은 대조군 171명에서 선정된 뇌졸중 대사체에 대해서 비선형모형을 통하여 용량반응 분석을 실시하였다. 분석에 포함한 세 가지 대사체 AC, HAG, PC 중에서 PC에서 의미 있는 변곡점을 확인할 수 있었다.

이 연구는 대사체 PC와 뇌졸중 발생과의 관계에서 변곡점의 존재 가능성을 제시하고 있다. 연구에 사용된 다른 두 대사체인 AC와 HAG는 변곡점 없는 선형관계를 보였다. PC의 변곡점 이후 PC가 두 배 증가하면 뇌졸중 발생이 44.7% [(2<sup>2.4512</sup>-1)×100] 증가하였다. 반면 선형적인 모형에서는 PC, HAG, AC 모두 유의한 선형을 보였다.

용량반응 분석을 통한 변곡점 확인에 대한 이전 연구 사례를 보면 특정 변수와 사망과의 용량반응관계가 거의 선형에 가깝고 변곡점이 나타나지 않은 연구들도 있었다[13,14]. 따라서 잠재적 변곡점 존재 여부는 조심스럽게 해석되어야 한다[15].

대사체와 관련한 이전 연구는 acrylamide에 의해 형성된 hemoglobin cysteine adduct level과 그 epoxide 대사체 glycidamide의 관계를 선형모형과 비선형 모형에 적합성을 각각 비교하였다[16]. 또한 의사결정 나무 귀납법 (decision tree induction)에 근거한 방법을 이용해 대사체 NADPH<sub>2</sub>의 역치를 찾은 연구도 있다[17]. Floter 등은 대사체 농도의 유의미한 역치를 밝힘으로써 대사체가 특정 반응을 나타내는 안정된 상태를 찾아내는 것이 대사체 역동성이나 역학을 이해하는 데 중요한 단계라고 주장했다.



**Figure 2.** Association between log-transformed 1-acetyl-2-formyl-sn-glycero-3-phosphocholine metabolite and the risk of stroke incidence.

이 연구에서 분석한 세 가지 대사체의 변곡점이 실제 뇌졸중 진단이나 예방관리에 효율적으로 사용될 수 있는지는 추후 연구를 통해서 확인되어야 할 부분이다. 그렇지만 문헌에 근거하여 이번 연구에 포함한 세 가지 물질 AC, HAG, PC와 뇌졸중 발생과의 관련 기전을 정리하면 다음과 같았다. AC는 혈관벽에 작용하여 동맥경화를 유발함으로써 뇌졸중을 일으키고[18,19], PC는 platelet activating factor (PAF)의 한 종류로 작용하여 혈전 생성을 유발함으로써 뇌졸중을 일으키는 것으로 보고된 바 있다[18].

HAG 또한 1-o-hexadecyl-2-acetyl-sn-glycerol-3-phosphorylcholine이라는 PAF의 전구체로 사용되면서 혈전생성을 유발하여 뇌졸중을 일으키는 것으로 보고된 바 있다[19]. 그렇지만 이러한 대사체들이 실제로 인구집단에서 뇌졸중을 조기 진단하는 바이오마커로 활용이 되는 바는 없는 것으로 사료되며, 뇌졸중 진단에 활용되기까지는 여전히 많은 연구가 필요하다고 본다.

이 연구의 장점은 연구설계가 전향적 연구라는 점이다. 즉, 대사체에 사용된 혈청은 뇌졸중 발생 약 7년 이전에 이미 확보하여 -70도 이하 냉동 보관한 상태로 유지를 하였고, 이후 뇌졸중 발생 이후에 시료를 분석하였으므로 인과관계에서 중요한 범주 중에 하나인 시간적 전후 관계를 만족하고 있다는 점이다[20].

그렇지만 몇 가지 연구의 제한점이 있다. 연구에 사용된 뇌졸중은 국민건강보험공단의 청구자료에 있는 상병코드만을 사용하였다. 비록 일부 표본을 통해 정확도가 83%로 보고된 바 있지만 연구의 제한점으로 남아 있다[7]. 연구에 포함된 뇌졸중은 42명이었고, 이 중 허혈성 뇌졸중 18명, 그리고 출혈성 뇌졸중이 7명이었다. 대사체와 뇌졸중과의 관련성은 뇌졸중 세부 형태에 따라 다를 수 있지만 표본크기의 제한으로

로 전체 뇌졸중에 국한하여 분석한 것은 제한점이다. 또 다른 제한점으로는 최종 용량반응 분석에 포함된 연구대상자 수가 적을 뿐만 아니라 대표성 있는 표본을 얻은 것이 아니므로 이 연구결과를 일반인구에 적용하기에는 무리가 있을 것으로 판단된다. 특히 최종 분석에 사용된 대사체에서 실험실적으로 분석결과가 0이거나 결측인 경우는 분석에서 제외하였다. 또한 이 연구에서 사용한 대사체의 자료가 intensity score로서 특정 물질을 정량한 것이 아니고, 추정되는 intensity score에 기반하여 근사한 물질을 추정한 것이다. 이 부분은 본 논문의 범위를 넘어서 대사체의 확인과 정량에 대한 부분이지만 역시 추후 연구가 계속 필요한 부분이라고 본다. 따라서 이 연구결과는 위에 언급한 여러가지 제한점이 있으며 추후 연구를 통해서 재평가될 필요성이 있다.

결론적으로 이 연구에서 용량반응 분석에 사용한 대사체 AC, HAG, PC 중에서 PC의 경우 의미 있는 뇌졸중 발생과 관련 있는 의미 있는 변곡점을 확인할 수 있었다. 향후 이를 활용한 뇌졸중 예측모형 개발이나 조기 진단에 활용할 수 있도록 후속 연구가 계속될 필요가 있다.

## REFERENCES

- World Health Organization. Fact sheet on cardiovascular diseases. Available at <http://www.who.int/mediacentre/factsheets/fs317/en/> [accessed on June 12, 2016].
- Statistics Korea. 2012 Annual report on the causes of death statistics. Daejeon: Statistics Korea; 2013.
- Jee SH, Suh I, Kim IS, Appel LJ. Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: the Korea Medical Insurance Corporation Study. *JAMA* 1999;282(22):2149-2155.
- Del Gobbo LC, Kalantarian S, Imamura F, Lemaitre R, Siscovick DS, Psaty BM, et al. Contribution of major lifestyle risk factors for incident heart failure in older adults: the Cardiovascular Health Study. *JACC Heart Fail* 2015;3(7):520-528.
- Ho JE, Larson MG, Ghorbani A, Cheng S, Chen MH, Keyes M, et al. Metabolomic profiles of body mass index in the framingham heart study reveal distinct cardiometabolic phenotypes. *PLoS One* 2016;11(2):e0148361.
- Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet* 2008;4(11):e1000282.
- Park JK, Kim KS, Kim CB, Lee TY, Lee KS, Lee DH, et al. The accuracy of ICD codes for cerebrovascular diseases in medical insurance claims. *Korean J Prev Med* 2000;33:76-82 (Korean).
- Hastie TJ, Tibshirani RJ. Generalized additive models. *Stat Sci* 1986;1(3):297-318.
- Hastie TJ, Tibshirani RJ. Generalized additive models. London, UK: Chapman & Hall; 1990.
- Venables WN, Ripley BD. Modern applied statistics with S-PLUS. New York: Springer-Verlag; 1999.
- Akaike H. Information theory and an extension of the maximum likelihood principle. Budapest: Akademiai Kiado; 1973.
- Lim YH, Ohn IS, Kim H. 2015 R: Health effects of air pollution and temperature (HEAT). Retrieved from <http://www.R-project.org> [accessed on June 12, 2016].
- Daniels MJ, Dominici F, Samet JM, Zeger SL. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *Am J Epidemiol* 2000;152(5):397-406.
- Ostro BD. A search for a threshold in the relationship of air pollution to mortality; a reanalysis of data on London winters. *Environ Health Perspect* 1984;58:397-399.
- Kim SY, Lee JT, Hong YC, Ahn KJ, Kim H. Determining the threshold effect of ozone on daily mortality: an analysis of ozone and mortality in Seoul, Korea, 1995-1999. *Environ Res* 2004;94(2):113-119.
- Calleman CJ, Stern LG, Bergmark E, Costa LG. Linear versus nonlinear models for hemoglobin adduct formation by acrylamide and its metabolite glycidamide: Implications for risk estimation. *Cancer Epidemiol Biomarkers Prev* 1992;1(5):361-368.
- Flöter A, Nicolas J, Schaub T, Selbig J. Threshold extraction in metabolite concentration data. *Bioinformatics* 2004;20(10):1491-1494.
- Iank ML, Snyder F, Byers LW, Brooks B, Muirhead EE. Antihypertensive activity of an alkyl ether analog of phosphatidylcholine. *Biochem Biophys Res Commun* 1979;90(4):1194-1200.
- Kramer RM, Jakubowski JA, Deykin D. Hydrolysis of 1-alkyl-2-arachidonoyl-sn-glycero-3-phosphocholine, a common precursor of platelet-activating factor and eicosanoids, by human platelet phospholipase A2. *Biochim Biophys Acta* 1988;959(3):269-279.
- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58(5):295-300.