

Determining a BMDL of Blood Lead Based on ADHD Scores Using a Semi-Parametric Regression

Ah Hyoun Kim¹ · Mina Ha² · Byung Soo Kim³

¹Department of Applied Statistics, Yonsei University; ² College of Medicine, Dankook University

³Department of Applied Statistics, Yonsei University

(Received April 10, 2012; Revised April 26, 2012; Accepted May 25, 2012)

Abstract

This paper derives a benchmark dose(BMD) and its 95% lower confidence limit(BMDL) using a semi-parametric regression model for small lead based changes in attention-deficit hyperactivity disorder(ADHD) scores in the first wave of the Children's Health and Environment Research(CHEER) survey data, which have been regularly collected in South Korea since 2005. Ha *et al.* (2009) showed that the appearance of ADHD symptoms had a borderline trend of increasing with the blood lead concentration. Butdz-Jørgensen (EFSA, 2010a) derived the BMDL of lead corresponding to a benchmark region of 1 full intelligent quotient (IQ) score using the raw data in Lanphear *et al.* (2005, EHP). European Food Safety Authority (EFSA, 2010b) determined the BMDL of 1.2 $\mu\text{g}/\text{dl}$ as a reference point for the characterization of lead when assessing the risk of the intellectual deficit measured by IQ scores. Kim *et al.* (2011) indicated that an even lower BMDL could be obtained based on the ADHD score; however, the BMDLs depended heavily upon the model assumptions. We show in this paper that a semi-parametric approach resolves the model dependence of BMDLs.

Keywords: Attention deficit hyperactivity disorder, benchmark dose lower limit, reference point, semi-parametric regression.

1. 서론

납은 자연 상태에서 나타나는 독성 물질로 주로 식품이나 음용수, 공기, 토양 등을 통해 인체에 노출된다. 그 중 식품이 가장 주요한 납의 노출원인데, 노출된 납은 인체에 오랜 시간 머물러 높은 축적성을 보이는 것으로 알려져 있다. 납이 인체에 흡수되는 경로는 먼저 위장관, 호흡기, 식도 등을 통해 흡수된 후, 혈액 중 적혈구와 결합되어 간이나 신장 등 신체의 여러 곳을 통해 최종적으로는 뼈에 축적된다. 따라서 납이 인체에 미치는 효과는 주로 혈중 납 농도를 기초로 파악되며, 뼈나 치아의 납 농도는 과거 납의 폭로 정보를 나타낸다 (EFSA, 2010b).

납이 인체에 미치는 여러 해효과에 대해서는 다양한 연구가 이루어져 왔으며, 특히 아동을 대상으로 한 연구에서는 납이 아동의 지적 능력 저하, 인지기능 저하, 행동장애, 학교성적저하 등을 야기하는 것으로 보고된 바 있다 (Lanphear 등, 2005). 아동에 대해 주의를 요하는 혈중 납 농도 수준은 2005년 이

³Corresponding author: Department of Applied Statistics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea. E-mail: bskim@yonsei.ac.kr

전의 연구에서는 $10\mu\text{g}/\text{dl}$ 로 제시하였으며 (US CDC, 2005), 이를 1주일 간 섭취 가능량(Provisional Tolerable Weekly Intake; PTWI)으로 환산하면 $25\mu\text{g}/\text{kg b.w}/\text{day}$ 에 해당한다 (JFCFA, 2000). 그런데 이에 앞서 국제보건기구 (WHO, 1995)와 미국 질병통제센터 (US CDC, 1991)의 보고에 따르면, 납 노출에 대한 해효과는 역치가 없을 것이며 $10\mu\text{g}/\text{dl}$ 보다 적은 혈중 납 농도도 인체에 해효과를 미칠 만큼의 독성을 가질 것이라고 추측하였으나, 이를 지지하는 구체적인 자료는 Canfield 등 (2003)에서 처음 보고되었다.

이어서 과거 기준치였던 $10\mu\text{g}/\text{dl}$ 가 적절한 수준이 아니며, 납이 인체에 미치는 해효과는 역치가 존재한다는 증거가 없다는 것이 여러 연구를 통해서 밝혀졌다. 특히 아동의 지적 능력을 저하시키는 것에 관한 것으로 Lanphear 등 (2005)은 혈중 납 농도가 $10\mu\text{g}/\text{dl}$ 보다 적은 경우에도 아동의 지적 결핍을 야기할 수 있음을 밝혔다. 이들은 전 세계 7개 지역에서 아동을 대상으로 수년에 걸쳐 수집된 코호트(cohort) 자료를 통합하여 메타분석을 수행하였으며, 그 결과 $7.5\mu\text{g}/\text{dl}$ 이하의 혈중 납 농도도 아동의 지적 수준을 저해할 수 있음을 보였다. 한편 Butdz-Jørgensen (EFSA, 2010a)은 Lanphear 등 (2005)의 연구에서 사용한 7개 경시적 자료에 기초하여 아동의 IQ 점수를 1점 저하시키는 혈중 납 농도의 95% 신뢰하계를 $1.2\mu\text{g}/\text{dl}$ 로 도출하였다.

한국의 경우 지난 40년간 빠르게 진행된 산업화 과정에서 다양한 유해 물질이 배출되어 환경을 오염시켰으며, 이로 인해 아동이 신경발달장애를 겪을 수 있다는 가설이 주목을 받고 있다 (Ha 등, 2009). 실제로 주의력결핍 과잉행동장애(attention deficit hyperactivity disorder; ADHD)로 치료를 받은 5세부터 14세 사이의 아동의 수는 2003년 16,607명에서 2011년 51,657명으로 증가하였으며, 질환 아동의 수가 가장 많은 나이 그룹이 5-9세에서 10-14세로 바뀌었다고 보고되고 있다 (국민건강보험정책연구원, 2011). 환경부에서는 2005년부터 한국 아동의 건강에 환경이 어떠한 영향을 미치는가를 파악하기 위한 장기간의 코호트 조사로 CHEER(Children's Health and Environment Research) 연구를 시작하였다.

본 연구에서는 납과 아동의 주의력 결핍 과잉행동장애(ADHD)와의 상관성을 파악하고자 한다. 이와 관련된 선행 연구로 Ha 등 (2009)은 CHEER 자료를 이용하여 아동의 혈중 납 농도가 ADHD 증상을 유의적으로 심화시킬 수 있음을 보였다. 또한 Kim 등 (2011)은 Butdz-Jørgensen (EFSA, 2010a)과 유사한 방법을 이용하여 CHEER 자료를 분석한 결과, 아동의 ADHD 지표를 약간 상승하게 만드는 혈중 납 농도의 95% 신뢰하계가 $1.2\mu\text{g}/\text{dl}$ 보다 적어짐을 보였다. 본 연구에서도 동일한 CHEER 자료를 이용하여 아동의 ADHD 지표를 약간 상승하게 만드는 혈중 납 농도의 95% 신뢰하계를 산출하였는데, Kim 등 (2011)의 연구와 주요한 차이점은 선형 회귀모형이 아닌 준모수 회귀모형(semi parametric regression model)을 적용하였다는 점이다. Kim 등 (2011)의 연구에서는 자료에 변수변환을 취하거나 모형을 확장 또는 축소하였을 때 추정치가 불안정하다는 점이 한계였으나, 본 연구에서는 준모수 회귀모형을 이용함으로써 추정치가 모형에 의존적이었던 문제를 해결해 보고자 한 것이다.

본 연구의 구성은 다음과 같다. 2장에서는 본 연구에 활용한 CHEER 자료 및 변수를 정의하고 벤치마크 용량에 관한 선행연구에 대해 살펴보고, 3장에서는 준모수적 모형을 구성하였다. 4장에서 모형의 적합 결과와 해석을 제시하고, 5장에서는 벤치마크 용량을 도출하기 위한 방법을 밝힌 뒤, 4장에서의 모형 적합 결과를 이용하여 산출된 벤치마크 용량의 값을 제시하고, 6장에서는 내용을 종합해 결론을 맺는다.

2. 자료 및 벤치마크 용량

2.1. CHEER 자료

CHEER 자료는 환경부에서 2005년부터 한국 아동의 건강과 환경에 관한 연구를 위해 수집한 자료로, 아동의 ADHD에 대한 심각성을 30점 척도로 지표화한 점수와 혈중 납 농도, 부모의 소득 및 교

Table 2.1. Variables in CHEER data set

변수	변수명	타입	단위
혈중 납 농도	Pb_W_B	연속형	$\mu\text{g}/\text{dl}$
ADHD 지표	ADHD	연속형	0~30
성별	sex	범주형	0 = 남자, 1 = 여자
아버지의 교육수준	dad_edu_yr	연속형	년
임신 중 흡연여부	smk_dur_preg	범주형	0 = 없음, 1 = 있음
부모의 소득	income	연속형	각 계급 별 중앙값 사용 (250 (500미만), 750 (500이상 1000미만), 1500 (1000이상 2000미만), 2500 (2000이상 3000미만), 4000 (3000이상 5000미만), 7000 (5000이상))
임신 중 음주여부	alc_dur_preg	범주형	0 = 없음, 1 = 있음
부모의 신경정신질환여부	fdr_neurop	범주형	0 = 없음, 1 = 있음

육수준 등 유용한 공변수를 포함하고 있다. 이는 한국 6개의 도시에서 1,873 명의 아동의 보호자를 대상으로 조사되었으며, 아동의 나이, 키, 몸무게의 평균(표준편차)는 각각 7(0.8)세, 121.2(9.9)cm, 27.5(6.4)kg이며, 아동의 혈중 납 농도의 중앙값(5~95 백분위수)은 1.93(0.660~3.590) $\mu\text{g}/\text{dl}$ 이다. 보호자는 아동의 ADHD 증상을 한국 아동에 맞게 수정된 Conner 척도 (So 등, 2003)을 이용하여 검사하였는데, 이는 10개의 문항으로 구성되어 있으며 각 문항별로 0, 1, 2, 3의 점수를 부여하는 검사이다. 최종 ADHD 지표는 모든 점수를 합산하여 도출되며 점수가 높을수록 ADHD 증상이 심각한 것으로 볼 수 있다. 본 연구에서 ADHD 지표를 설명하기 위한 모형에 활용된 변수에 관한 구체적인 내용은 Table 2.1에 정리되어 있다.

2.2. 벤치마크 용량

BMD(benchmark dose)란 독성 물질에 노출되지 않은 관찰치에 비해 사전에 정해진 만큼의 반응값의 증감을 야기하는 독성 물질의 양으로 정의되며, 여기서 반응값의 증가분은 BMR(benchmark region), BMD의 95% 신뢰하계를 BMDL(benchmark dose limit)이라 부른다. 앞서 소개한 Butdz-Jørgensen (EFSA, 2010a)의 연구를 예로 들어보면, 아동의 IQ 점수 1점에 대응되는 것이 BMR이고, IQ 점수 1점 하락을 야기하는 혈중 납 농도인 BMD가 1.8 $\mu\text{g}/\text{dl}$ 로 도출되었다. 또, 이에 대한 95% 신뢰하계인 BMDL이 1.2 $\mu\text{g}/\text{dl}$ 로 보고되었다. 즉, BMDL은 인체 안전 기준을 산정하는 독성 물질의 기준값이라 볼 수 있는데, BMDL 이전에 활용되었던 독성 기준값인 NOAEL(No Observed Adverse Effect Level)을 대체하기 위하여 만들어진 개념이다.

NOAEL은 대조군과 유의한 차이를 나타내지 않는 최대 실험 용량을 의미하며, 기본적으로 노출과 반응의 관계에 역치가 존재할 것으로 보고 만들어진 개념이므로 혈중 납 농도와 같이 역치 존재를 입증하기 힘든 독성 물질의 영향을 파악하기에 적합하지 않다. 뿐만 아니라 NOAEL 추정치는 실험설계에 의존적이며, 가설검정의 맥락에서 정의되었기 때문에 표본이 적어질수록 잠재적 독성 물질을 시장에 출하하려는 제조업자에게 득이 된다는 모순을 안고 있다. 이에 대한 대안으로, Crump (1984)는 이진형 자료의 BMD 산출 방법론을 제안하였으며, 연속형 자료에 대해서는 Gaylor와 Slikker (1990), Crump (1995)에 의해 고안되었는데, 이상의 방법론은 모두 동물 실험 자료를 대상으로 한 것이었다.

한편, 사람을 대상으로 하는 관찰 연구의 경우 해당 물질에 전혀 노출되지 않은 대조군을 찾기 힘들며, 반응 변수가 노출 변수 이외의 여타 중첩 변수에 의하여 영향을 받기 마련이다. Butdz-Jørgensen 등 (2001)은 기존의 실험 자료에 대한 BMD 접근법을 확장하여 관찰 연구에서 BMD 및 BMDL을 산출하는 방법을 제안하였으며, 페로 제도(Faroe Islands)의 아동에 대한 관찰 자료를 이용하여 신경심리적 장애의 척도인 BNT(Boston Naming Test) 점수에 대한 수은의 BMDL를 도출하였다. 이후에 Butdz-Jørgensen (EFSA, 2010a)은 Butdz-Jørgensen 등 (2001)에서 고안한 방법을 활용하여 아동의 IQ 점수를 1점 저하시키는 혈중 납 농도의 BMDL을 도출하였으며, 본 연구에서도 ADHD에 대한 혈중 납 농도의 BMD를 산출하는데 이와 유사한 방식으로 접근하였다.

Butdz-Jørgensen (EFSA, 2010a)의 연구에서 활용된 BMD와 BMDL의 산출 방법론은 다음과 같다. 먼저 관찰 연구에서는 노출 변수 이외의 변수에 의해 반응 변수가 영향을 받게 되므로 아래 식 (2.1)과 같이 확장된 모형을 고려해야 한다. 아래 모형에서 Y 는 반응 변수를 나타내며 d 는 노출 변수, c_1, \dots, c_k 는 그 외의 변수를 각각 의미한다.

$$Y(d, c_1, \dots, c_k) = \beta f(d) + \beta_0 + \beta_1 c_1 + \dots + \beta_k c_k + \epsilon, \quad (2.1)$$

단, $\epsilon \sim N(0, \sigma_\epsilon^2)$, f 는 단조함수이고 $f(0) = 0$.

즉, 위 모형에서 β 의 추정치를 도출하면, $\hat{\beta}f(d)$ 이 BMR이 되며, 이러한 BMR에 대응되는 d 의 값이 BMD가 된다. BMD의 95% 신뢰하계가 BMDL이므로, 그 산출 식은 아래 식 (2.2)와 같다.

$$\hat{\beta}_u f(d) = \text{BMR}, \quad \text{단, } \hat{\beta}_u = \hat{\beta} + 1.645 \text{s\hat{e}}(\hat{\beta}). \quad (2.2)$$

한편, BMDL을 산출하기 위하여 위 식 (2.2)를 이용하는 대신 모수적 붓스트랩(bootstrap) 방법을 이용할 수도 있다. 기존 모형에서의 반응값 추정치에 랜덤하게 발생한 오차 등을 더하여 반응변수를 재생성한 뒤, 그 반응변수를 이용해서 모형을 새로 적합하면 BMD도 새로 산출할 수 있는데, 이 과정을 반복하여 BMD의 분포를 얻은 뒤 하위 5%에 해당하는 값으로 BMDL을 구하는 방식이다. 본 연구에서 활용하는 준모수적 회귀모형에는 붓스트랩 방식을 이용하는 것이 더 적절하므로, 본 연구에서도 이상의 방식을 이용하여 BMDL을 도출하였다.

3. 준모수적 모형

본 연구는 기본적으로 앞서 정의한 설명변수 가운데 자료의 종류가 명목형의 경우는 모수적 인자로, 혈중 납의 농도를 포함한 모든 연속형 설명변수는 비모수적 인자일 수도 있다는 가능성을 염두에 두고 모형을 처음 설계한 뒤, 이들 변수를 비모수적 인자로 가정하는 것이 적절한가에 대한 검토 후 최종 모형을 구성하였다. 앞서 밝힌 바와 같이 ADHD 지표를 설명하는데 의미있는 7가지 설명변수는 Pb_W_B, sex, dad_educ_yr, smk_dur_preg, income, alc_dur_preg, fdr_neurop이며, 이들 변수로 ADHD를 설명하기 위한 모형을 아래 모형 1과 같이 구성하였다. 아래 모형에서 연속형 자료인 Pb_W_B, dad_educ_yr, income은 우선 비모수적 인자로 가정하며, 나머지는 가변수로 구성하였다. 또한 아래 식에서 두 번째 등호는 주어진 모형의 비모수적 인자를 벌점화 스플라인(penalized spline)으로 두고 정의된 식으로, 각 해당 비모수적 변수 별 계수인 u_k^p, \dots, u_k^I 의 값이 커지지 않도록 제약을 둔 상태로 각 계수의 최소제곱 추정량(least square estimators)을 도출해야 한다.

모형 1

$$\text{ADHD} = \beta_0 + f(\text{Pb_W_B}) + \beta_s \cdot \text{sex} + g(\text{dad_educ_yr}) + \beta_m \cdot \text{smk_dur_preg} + h(\text{income})$$

Table 3.1. Check of the nonlinear relationship of continuous explanatory variables against the response

Pb_W_B	dad_edu_yr	income	AIC
비선형	선형	선형	9553.40
선형	비선형	선형	9553.54
선형	선형	비선형	9546.06
비선형	선형	비선형	9545.89
선형	비선형	비선형	9546.06
비선형	비선형	비선형	9545.89

$$\begin{aligned}
 & + \beta_a \cdot alc_dur_preg + \beta_f \cdot fdr_neurop + \epsilon, \\
 = & \beta_0 + \beta_p \cdot Pb_W_B + \sum_{k=1}^{K_p} u_k^p (Pb_W_B - \kappa_k^p)_+ + \beta_s \cdot sex + \beta_d \cdot dad_edu_yr \\
 & + \sum_{k=1}^{K_d} u_k^d (dad_edu_yr - \kappa_k^d)_+ + \beta_m \cdot smk_dur_preg + \beta_I \cdot income + \sum_{k=1}^{K_I} u_k^I (income - \kappa_k^I)_+ \\
 & + \beta_a \cdot alc_dur_preg + \beta_f \cdot fdr_neurop + \epsilon, \tag{3.1}
 \end{aligned}$$

단, $\epsilon \sim N(0, \sigma_\epsilon^2)$ 이고, 모형에서 모수적 변수의 계수는 각각 $\beta_s, \beta_m, \beta_a, \beta_f$, 비모수적 변수의 계수는 u_k^p, \dots, u_k^I 이고, C_p, C_d, C_I 가 사전에 주어진 상수(평활 모수와 일대일 대응)라고 할 때 $\sum_{k=1}^{K_p} (u_k^p)^2 < C_p, \sum_{k=1}^{K_d} (u_k^d)^2 < C_d, \sum_{k=1}^{K_I} (u_k^I)^2 < C_I$ 를 만족한다. 그리고 f, g, h 는 각 비모수적 변수 별 평활 함수를 나타내며, $\kappa_1^p, \dots, \kappa_{K_p}^p, \kappa_1^d, \dots, \kappa_{K_d}^d, \kappa_1^I, \dots, \kappa_{K_I}^I$ 는 각 해당 변수 별 매듭(knots)을 의미하고, 절단 함수인 u_+ 는 $u > 0$ 인 경우는 u , 그렇지 않은 경우 0인 함수로 정의된다.

위 모형에서 비선형 함수인 f, g, h 의 평활 정도를 결정해야 하는데, 이는 평활 모수(smoothing parameter)와 대응 관계에 있는 f, g, h 각각의 자유도 df_p, df_d, df_I 에 반영되며 본 연구에서는 이를 REML 방법을 이용하여 추정하였다 (Ruppert 등, 2003). f, g, h 의 자유도 df_p, df_d, df_I 의 REML 추정치는 위 모형 1을 혼합효과모형(Mixed Effect Model)로 재구성한 뒤, 그 모형에서의 모수인 랜덤 효과와 오차항의 분산에 대한 최선선형불편예측치(BLUP)를 이용하여 구할 수 있으며, 본 연구에서 이러한 BLUP의 추정은 R의 lme 함수를 이용하였다. CHEER 자료에 대해 이상의 논의된 바에 따라 모형을 적용한 결과, Pb_W_B의 함수 f 는 자유도를 $df_p = 1.473$, dad_edu_yr의 함수 g 는 자유도를 $df_d = 1.000$, income의 함수 h 는 자유도를 $df_I = 2.614$ 로 두는 것이 적절한 것으로 도출되었다.

한편, 모형 1에서는 모든 연속형 변수를 비모수적 인자로 가정하였는데, 이러한 가정이 적절한가에 대한 검토가 필요하다. 즉 Pb_W_B, dad_edu_yr, income가 반응 변수인 Pb_W_B와 비선형의 관계를 가지는가에 대한 것이다. 우선 탐색적 분석으로 아래 Figure 4.1의 산점도를 살펴보면 ADHD를 dad_edu_yr에 대한 비선형함수로 보기 어렵다는 것을 알 수 있다. 또한 Table 3.1에서 확인되듯이 dad_edu_yr는 선형, Pb_W_B, income는 비선형 관계로 가정하는 것이 AIC 기준에서 적합도가 가장 좋아 이를 최종 모형으로 채택하기로 하였다.

이상의 내용을 종합하면 반응 변수인 ADHD를 7개의 설명변수 (Pb_W_B, sex, dad_edu_yr, smk_dur_preg, income, alc_dur_preg, fdr_neurop)로 설명하는 모형을 구성하되, Pb_W_B는 자유도가 1.473인 비모수적 인자로, income은 자유도가 2.614인 비모수적 인자로 두고 나머지 5개의 변수인 sex, dad_edu_yr, smk_dur_preg, alc_dur_preg, fdr_neurop는 모수적 인자로 두는 준모수 회귀모형이 적정한 것으로 판단된다. 최종 모형인 모형 2는 다음과 같이 구성된다.

모형 2

$$\begin{aligned}
\text{ADHD} &= \beta_0 + f(\text{Pb_W_B}) + \beta_s \cdot \text{sex} + \beta_d \cdot \text{dad_edu_yr} + \beta_m \cdot \text{smk_dur_preg} \\
&\quad + h(\text{income}) + \beta_a \cdot \text{alc_dur_preg} + \beta_f \cdot \text{fdr_neurop} + \epsilon, \\
&= \beta_0 + \beta_p \cdot \text{Pb_W_B} + \sum_{k=1}^{K_p} u_k^p \left(\text{Pb_W_B} - \kappa_k^p \right)_+ + \beta_s \cdot \text{sex} + \beta_d \cdot \text{dad_edu_yr} \\
&\quad + \beta_m \cdot \text{smk_dur_preg} + \beta_I \cdot \text{income} + \sum_{k=1}^{K_I} u_k^I \left(\text{income} - \kappa_k^I \right)_+ + \beta_a \cdot \text{alc_dur_preg} \\
&\quad + \beta_f \cdot \text{fdr_neurop} + \epsilon, \tag{3.2}
\end{aligned}$$

단, $\epsilon \sim N(0, \sigma_\epsilon^2)$ 이고, 기타 표기에 대한 정의는 식 (3.1)과 유사하다.

4. 모형 적합 결과

R의 gam 함수를 이용하여 CHEER 자료에 대하여 위 모형 2을 적합하였다. 결과 중 일부를 살펴보기 위하여 아래 Figure 4.1에서 아동의 혈중 납 농도, 아버지의 교육수준, 소득 수준, 부모의 신경정신질환 여부와 반응변수인 ADHD 간의 산점도를 보여주고 있다. 산점도에서 추정된 ADHD의 값을 선으로 표현하여 적합의 정도를 비교할 수 있도록 하였는데, 추정치에 해당하는 값은 수평축의 기준변수를 제외한 나머지 설명변수의 평균값을 추정된 회귀식에 대입하여 계산된 것이다. 추정된 결과를 살펴보면 우선 Pb_W_B와 ADHD 간에 용량의존적 증가가 보이고 있음을 주목할 필요가 있다. 이는 혈중 납의 농도가 증가하면 ADHD가 심해진다는 주장에 대한 근거가 될 수 있는데, 한편, dad_edu_yr나 income의 경우 ADHD와 음의 관계를 보이고 있는데, 이는 아버지의 교육 수준이 증가하고 소득수준이 증가할수록 ADHD 증상을 덜 보이는 것으로 해석할 수 있으며, 비모수 함수로 추정된 소득수준 변수의 경우 일정 수준까지 증가하는 동안은 ADHD 지표 추정치가 비교적 급격하게 떨어지다 일정 수준을 넘어서면 수평이 유지되는 것을 보이고 있다. 또한 fdr_neurop과 ADHD도 양의 상관관계를 가지는데 부모가 신경정신적 질환을 겪은 아동의 ADHD 점수가 높아짐을 알 수 있다.

5. BMDL의 산출 방법론 및 산출 결과

5.1. 준모수 회귀함수에서 BMD를 찾는 특수성

본 연구에서 BMD는 사전에 정의되어 있는 일정한 수준(BMR) 만큼 ADHD를 증가시킬 수 있는 혈중 납의 농도(Pb_W_B)를 의미한다. 앞서 식 (2.2)에서 살펴본 BMD 산출식은 선형 회귀모형을 이용하는 경우에만 적용될 수 있으며, 본 연구에서와 같은 준모수 회귀함수를 이용하는 경우에 BMD를 도출하는 것은 좀 더 특수한 문제이다. 본 연구에서는 정해진 BMR에 Pb_W_B가 0일 때의 ADHD 추정치 평균을 더한 뒤, 이 값을 만족하도록 하는 Pb_W_B를 구함으로써 BMD를 산출하였다. 구체적으로, 사전에 정해져 있는 BMR에 대하여 아래 식 (5.1)을 만족하는 Pb_W_B의 값이 BMD가 된다.

$$\hat{f}(0) + \text{BMR} \approx \hat{f}(\text{Pb_W_B}). \tag{5.1}$$

앞서 모형 2로 추정된 준모수 회귀모형에 Pb_W_B는 0, 그 외 나머지 설명변수는 해당 변수의 평균값을 대입하여 $\hat{f}(0)$ 를 도출해 보았으며, 그 결과 $\hat{f}(0) = 6.3249$ 로 추정되었다. 즉, 정해진 BMR에 대하여 추정치가 $6.3249 + \text{BMR}$ 로 나오도록 하는 Pb_W_B의 값을 찾아 이를 BMD로 둘 수 있다.

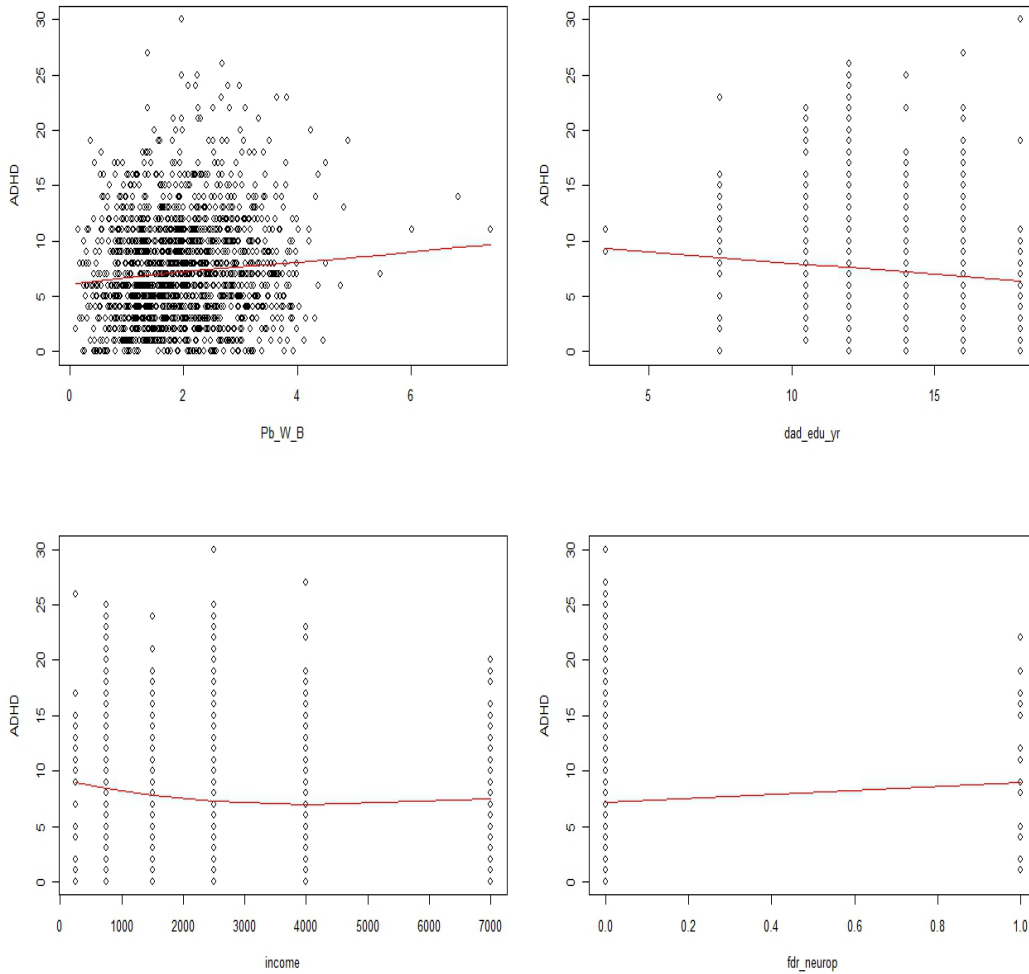


Figure 4.1. Scatter plots of ADHD versus major explanatory variable

5.2. 도함수 산점도

앞서 논의된 내용 중 ($Pb_W_B, \hat{f}(Pb_W_B)$)가 단조 증가하는 형태인지 보다 엄밀히 파악하기 위하여 $\hat{f}(Pb_W_B)$ 에 대한 1차 도함수와 그 95% 신뢰구간을 도출해 보았다. 1차 도함수에 대한 신뢰구간의 하계가 0보다 큰 경우에는 $\hat{f}(Pb_W_B)$ 이 단조 증가하는 형태가 통계적으로 유의하다고 해석할 수 있다. 기존의 모형을 그대로 이용할 경우 \hat{f} 에 대한 1차 도함수는 0이 되어 도함수 산점도를 도출할 수 없으므로, Pb_W_B 에 대한 2차 항까지 포함한 준모수 회귀 모형으로 다시 적합해 줄 필요가 있다. 이때 Pb_W_B 이외의 나머지 변수의 경우 Pb_W_B 에 대한 \hat{f} 의 도함수 산점도에는 영향을 미치지 않으므로 앞 모형과 동일하게 정의하였다. 이 경우 \hat{f} 의 도함수와 도함수에 대한 95% 신뢰구간은 앞서 모형 1에서 비모수적 인자의 자유도를 결정하던 방법과 동일하게 원 모형을 혼합 효과 모형(Mixed Effect Model)로 재구성한 뒤 모수에 대한 BLUP 추정치를 이용하면 쉽게 도출할 수 있다 (Ruppert 등, 2003, 6.4절).

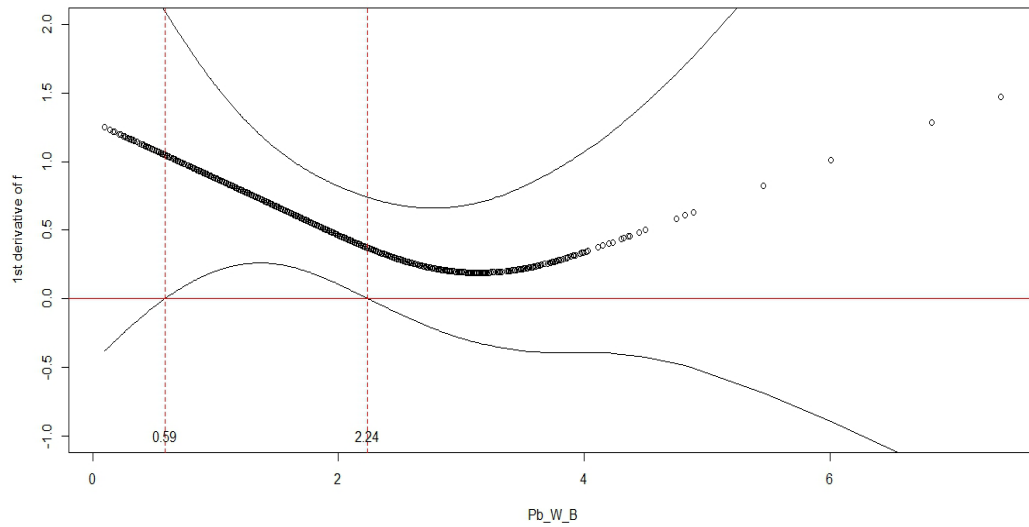


Figure 5.1. The first derivative and its variability band of ADHD against the blood lead concentration

이상의 과정을 통해 Pb_W_B 에 대한 \hat{f} 의 도함수 산점도를 도출한 결과는 아래 Figure 5.1과 같다. 도함수 추정치는 모두 양의 값을 가지고 있으므로 Pb_W_B 에 대해 f 는 단조 증가한다고 말할 수 있으나, 도함수 추정치의 신뢰구간을 살펴보면 Pb_W_B 이 $1.5\mu\text{g}/\text{dl}$ 를 기준으로 더 작아지거나 커질수록 변동 폭이 점차 커지는 것을 알 수 있다. 그래도 Pb_W_B 가 $0.59\text{--}2.24\mu\text{g}/\text{dl}$ 범위의 값에 대해서는 f 가 단조 증가하는 형태를 가지는 것이 통계적으로 유의하다고 말할 수 있다. 앞서 밝힌 바와 같이 Lanphear 등 (2005)은 혈중 납 농도가 $7.5\mu\text{g}/\text{dl}$ 보다 작은 경우에도 아동의 지적 발달을 저해할 수 있음을 밝혔는데, 이상의 결과에 의하면 이보다 훨씬 적은 수준인 $0.59\text{--}2.24\mu\text{g}/\text{dl}$ 범위의 값에 대해서 아동의 행동장애에 미치는 효과가 통계적으로 유의하다는 점에서 의미가 있다.

5.3. BMD와 BMDL의 산출

아래 Figure 5.2은 BMR이 0.5로 주어졌을 때 $\hat{f}(0) + \text{BMR} \approx \hat{f}(Pb_W_B)$ 를 만족하는 Pb_W_B 를 도출한 것이다. \hat{f} 이 비선형 함수이므로 닫힌 해로 정리되지 않으므로, $\hat{f}(0) + \text{BMR}$ 와 $\hat{f}(Pb_W_B)$ 의 차이의 절대값이 가장 작을 때의 Pb_W_B 를 찾는 알고리즘을 이용하였는데 도출 결과 BMD는 $0.84\mu\text{g}/\text{dl}$ 였다.

BMDL은 앞서 구한 BMD의 95% 신뢰하계에 해당하는 값이다. 본 연구에서는 Pb_W_B 와 ADHD의 관계를 비모수로 가정하였기 때문에 BMD의 95% 신뢰하계를 구하기 위해서는 시뮬레이션을 통해 데이터를 재생성한 뒤 BMD를 도출하는 과정을 반복하여 BMD의 분포를 통하여 구할 수 있다. 그 과정은 구체적으로 다음과 같다.

먼저 모형 2를 혼합효과모형(Mixed Effect Model)로 변환하면 오차항과 랜덤효과 변수가 기대값이 0인 정규분포를 따르게 되는데, 해당 변수 별 분산의 BLUP 추정치를 이용하여 오차항과 랜덤효과 변수를 모든 자료 관찰치 수만큼 생성한다. 여기에 앞서 적합한 모형2의 계수 추정치를 반영하여 ADHD 변수를 새로 생성해준다. 이렇게 새로 생성된 ADHD 변수와 기존의 모든 설명 변수를 모형 2에 적용하여 다시 모형을 적합해 주면 새로운 BMD를 도출할 수 있다. 이 과정을 10,000번 반복하여 BMD의 분포

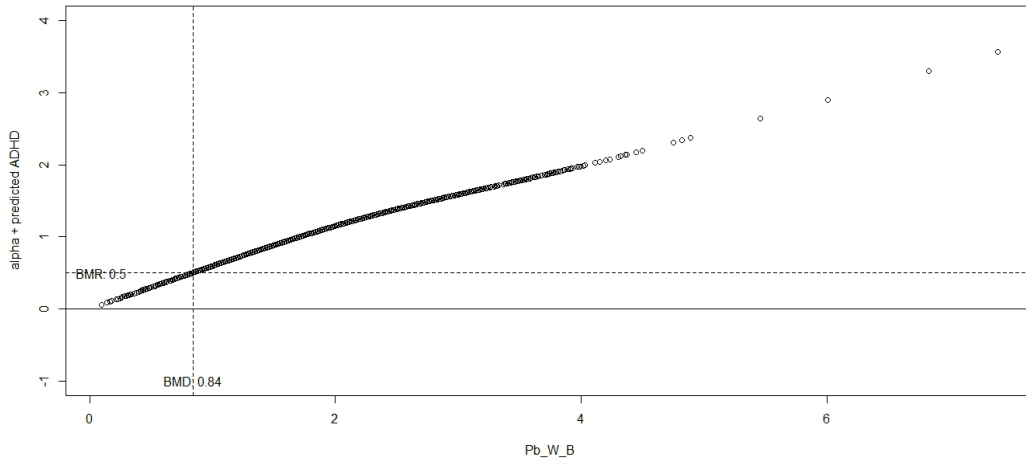


Figure 5.2. The BMD of the blood lead concentration which raises 0.5 (BMR) of ADHD

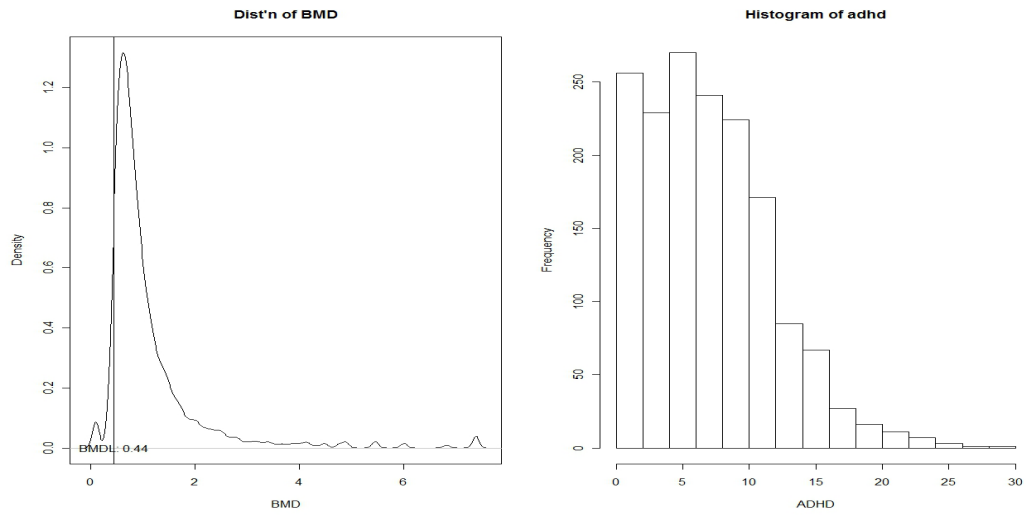


Figure 5.3. The distribution of BMD of the blood lead concentration which raises 0.5 of the ADHD, the 95% lower confidence limit(BMDL), and the histogram of ADHD

를 만들어낸 뒤, 그 분포에서 하위 5%의 퍼센타일에 해당하는 값을 찾아 그 값을 BMDL로 둔다. 이상의 방법으로 BMD의 분포를 도출한 결과는 아래 Figure 5.3(왼편)과 같으며, 분포의 95% 신뢰하계인 BMDL은 $0.44\mu\text{g}/\text{dl}$ 로 추정되었다.

5.4. 변수 변환을 통한 모형 보정

앞 절의 Figure 5.3(왼편)에서 BMD의 분포는 양의 왜도를 갖는 형태를 보이고 있다. 이는 정규가정으로부터의 이탈에서 기인하는 것인데, 아래 Figure 5.3(오른편)에서와 같이 ADHD의 분포도 이와 유사

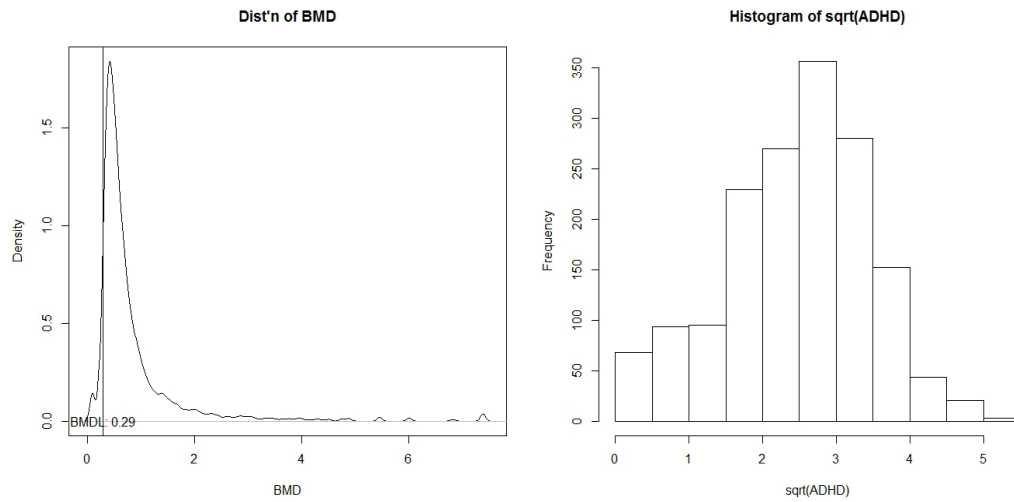


Figure 5.4. The distribution of BMD of the blood lead concentration which raises 0.1 of the square root of ADHD, the 95% lower confidence limit (BMDL), and the histogram of the square root of ADHD

Table 5.1. Comparison of BMDLs based on the original and the transformed models

	원 모형 결과	변수변환 모형 결과
평균(표준편차)	7.43(4.813)	2.539(0.994)
BMR	0.5	0.1
BMD	0.84	0.59
BMDL	0.44	0.29

한 형태를 가지고 있어 ADHD의 변수 변환을 고려해 볼 필요가 있어 보인다.

따라서 추가적으로 ADHD 변수에 제곱근을 취한 뒤, 앞서 논의된 것과 동일한 절차에 따라 모형을 추정하고 BMD와 BMDL를 도출해 보았다. 제곱근을 취했을 때의 BMD의 분포를 도출한 결과와 반응변수의 히스토그램은 아래 Figure 5.4와 같다. BMD의 분포의 오른쪽 꼬리가 더 길어진 것을 확인하였으며, 반응 변수도 보다 대칭형의 분포로 바뀌었음을 확인할 수 있다.

아래 Table 5.1은 원 모형에서의 결과와 변수변환된 모형에서의 결과를 종합적으로 비교한 것이다. BMR의 경우 자료 분석 이전에 적정한 값을 미리 설정하기 어렵지만 반응변수의 5% 또는 10% 정도의 적은 값을 두는 것이 일반적이며, 본 연구에서는 반응변수 표준편차의 약 1/10 정도에 해당하는 값으로 두었다. 즉, ADHD를 이용한 모형의 경우는 0.5, ADHD의 제곱근을 이용한 모형의 경우는 0.1로 두었다. 또한 각 BMR에 대하여 ADHD를 반응변수로 하는 모형의 BMDL은 $0.44\mu\text{g}/\text{dl}$ 이고, ADHD의 제곱근을 반응변수로 하는 모형의 BMDL은 $0.29\mu\text{g}/\text{dl}$ 로 도출되었는데 BMDL 추정치가 모형 선택이나 변수변환에 민감하게 변할 수 있다는 것을 감안했을 때 본 연구의 결과는 상당히 안정적인 것이라 말할 수 있다. 비교를 위해 동일한 자료를 이용한 선형 회귀모형을 적합하여 ADHD에 대한 BMDL을 도출한 Kim 등 (2011)의 결과를 아래 Table 5.2에 정리하였다. 이 경우는 ADHD의 제곱근을 반응변수로 둔 모형과 ADHD를 반응변수로 둔 모형간의 BMDL의 차이는 최대 5배 이상 정도 나는 것을 확인할 수 있다.

마지막으로 원 모형과 변수변환을 이용한 모형 각각에 대한 잔차 산점도와 잔차의 정규확률그림을 아래

Table 5.2. BMDLs based on the linear regression models of Kim *et al.* (2001)

	ADHD를 이용한 모형	ADHD의 제곱근을 이용한 모형	
Pb_W_B	유의함	유의함	유의함
sex	유의함	유의함	유의함
dad_edu_yr	유의함	유의함	유의함
smk_dur_preg	유의함	유의함	유의함
income	유의함	유의함	유의함
area(6 levels)	유의하지 않음	유의함	유의함
alc_dur_preg	유의함	유의하지 않음	유의하지 않음
fdr_neurop	유의함	유의하지 않음	유의하지 않음
Pb_W_B ²	유의하지 않음	제외	유의함
Pb_W_B ³	유의하지 않음	제외	유의함
BMR	0.5	0.1	0.1
BMD	1.012	0.838	0.198
BMDL	0.749	0.612	0.130

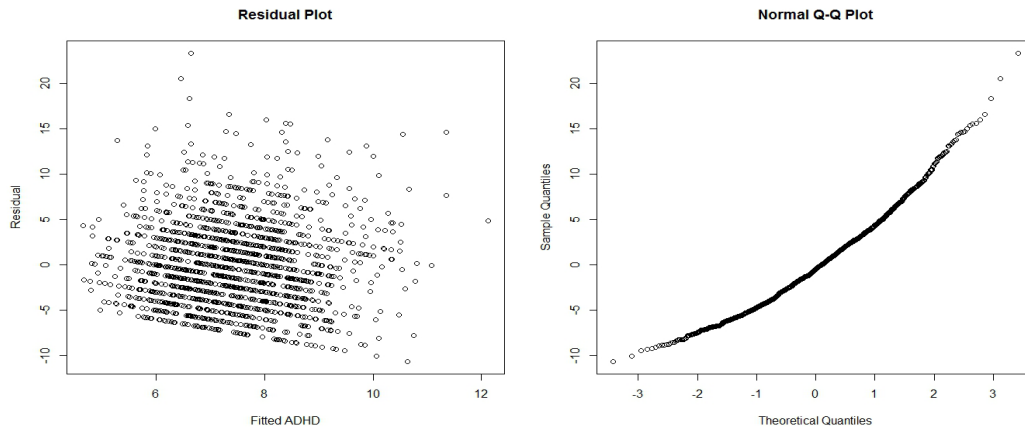


Figure 5.5. Residual plot of the original model and its normal probability plot

Figure 5.5와 Figure 5.6에 제시하였다. 두 모형의 모두 모형의 등분산성, 독립성 가정에 크게 문제가 없음을 확인할 수 있으며, 잔차의 분포도 정규분포로부터 크게 이탈하지 않았음을 확인할 수 있다.

6. 토의 및 결론

EFSA (2010a)에서 납의 위해평가를 하면서 세 가지 다른 종점에 대하여 각각 다른 BMDL을 제시하였다. IQ를 종점으로 할 때 1%의 하락을 BMR로 두고 구한 BMDL은 1.2 μ g/dl, 이완기 혈압 1% 상승을 BMR로 두고 구한 BMDL은 3.6 μ g/dl, 그리고 만성 신장병 발병율을 10% 증가시키는 것을 BMR로 할 때 구한 BMDL은 1.5 μ g/dl로 제시하였다. 본 연구에서는 ADHD를 종점으로 사용하였고 ADHD 표준 편차의 1/10을 BMR로 하여 BMDL을 새롭게 제시하고 있다. 이를 위해 본 연구에서 사용한 CHEER 자료는 한국 아동의 건강, 특히 ADHD 지표로 표현될 수 있는 정신적 행동장애에 소량의 혈중 납 농도가 어떠한 영향을 미칠 수 있는가에 대한 유용한 정보를 담고 있다. 이러한 CHEER 자료를 이용하여

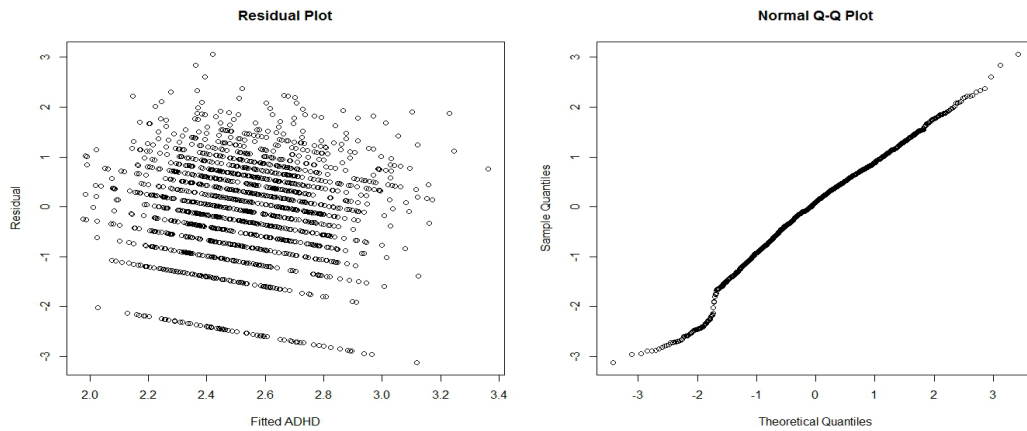


Figure 5.6. Residual plot of the transformed model and its normal probability plot

아동의 ADHD 지표를 소폭 상승시킬 수 있는 혈중 납 농도의 BMDL을 산출하였으며, 그 결과 ADHD 지표 0.5의 BMR을 상승시키는 혈중 납 농도의 BMDL은 $0.44\mu\text{g}/\text{dl}$ 로, ADHD의 제공근 지표 0.1의 BMR을 상승시키는 혈중 납 농도의 BMDL은 $0.29\mu\text{g}/\text{dl}$ 로 각각 도출되었다.

Wheeler와 Bailer (2009)이 제안한 모형평균 방법을 참고하여, 각 모형에서 도출된 BMDL의 값을 해당 모형 별 적합도를 나타내는 AIC를 이용하여 가중평균하여 도출된 단일 BMDL을 제안해 볼 수 있다. 원 모형의 AIC 값이 변수변환을 취한 모형의 AIC 값에 비해 약 2.13배 높으므로, 원 모형에서 나온 BMDL의 가중치를 $1/(1 + 2.13)$ 으로, 변환 모형에서 나온 BMDL의 가중치를 $2.13/(1 + 2.13)$ 으로 두고 가중평균하면 $0.33\mu\text{g}/\text{dl}$ 으로 도출된다. 따라서 $0.33\mu\text{g}/\text{dl}$ 을 아동의 ADHD 증상에 관한 안전기준을 산정하는 혈중 납 농도의 기준값으로 보는 것이 적절할 것으로 보인다. 이는 EFSA (2010b)에서 아동의 IQ 점수에 대해 산출한 BMDL $1.2\mu\text{g}/\text{dl}$ 보다 낮은 수준이라는 점에서 주목할 필요가 있다. 또한 본 연구에서는 준모수 회귀모형을 활용하여 이전 연구에서의 BMDL의 산출 값이 모형이나 변수 변환 등에 민감한 부분을 어느 정도 극복하였다는 점이 진전이라 말할 수 있다.

References

- National Health Insurance Corporation, Korea. (2012). Data request through personal communication.
- Budtz-Jørgensen, E., Keiding, N. and Grandjean, P. (2001). Benchmark dose calculation from epidemiological data, *Biometrics*, **57**, 698–706.
- Canfield, R. L., Henderson, C. R. Jr., Cory-Slechta, D. A., Cox, C., Jusko, T. A. and Lanphear, B. P. (2003). Intellectual impairment in children with blood lead concentrations below 10 microgram per deciliter, *The New England Journal of Medicine*, **348**, 1517–1526.
- Crump, K. (1984). A new method for determining allowable daily intakes, *Fundamental and Applied Toxicology*, **4**, 854–871.
- Crump, K. (1995). Calculation of benchmark doses from continuous data, *Risk Analysis*, **15**, 79–89.
- EFSA (2010a). An international pooled analysis for obtaining a benchmark dose for environmental lead exposure in children. (Question N°. EFSA-Q-2009-01078).
- EFSA (2010b). EFSA Panel on Contaminants in the Food Chain (CONTAM); Scientific Opinion on Lead in Food, *EFSA Journal* 2010, **8**(4), 1570 [147pp]. doi:10.2903/j.efsa.2010.1570. Available online: www.efsa.europa.eu.

- Gaylor, D. W. and Slikker, W. Jr. (1990). Risk assessment for neurotoxic effect, *Neurotoxicology*, **11**, 211–218.
- Ha, M., Kwon, H.-J., Lim, M.-H., Jee, Y.-K., Hong, Y.-C., Leem, J.-H., Sakong, J., Bae, J.-M., Hong, S.-J., Roh, Y.-M. and Jo, S.-J. (2009). Low blood levels of lead and mercury and symptoms of attention deficit hyperactivity in children: A report of the children's health and environment research (CHEER), *Neurotoxicology*, **30**, 31–36.
- JECFA (2000) Evaluation of Certain Food Additives and Contaminants, *WHO Technical Report Series*, **896**, 80–83.
- Kim, B.-S., Ha, M. and Kwon, H.-J. (2011). A benchmark dose analysis for lead exposure on the attention deficit hyperactivity disorder data of Korean children, *Biometrie und Medizinische Informatik*, **18**, 59–69.
- Lanphear, B. P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D. C., Canfield, R. L., Dietrich, K. N., Bornschein, R., Green, T., Rothenberg, S. J., Needleman, H. L., Schnaas, L., Wasserman, G., Graziano, J. and Roberts, R. (2005). Low-level environmental lead exposure and children's intellectual function: an international pooled analysis, *Environmental Health Perspectives*, **113**, 854–899.
- Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*, Cambridge University Press, New York.
- So, Y. K., Kim, Y. S., Choi, N. K., Kim, S. J., Non, J. S. and Ko, Y. J. (2003). The reliability and validity of Korean Conner's parent and teacher rating scale, *J Child Adolesc Psychiatry*, **14**, 183–194.
- US CDC (1991). Preventing Lead Poisoning in Young Children: A Statement by the Centers for Disease Control. Atlanta, GA: Centers for Disease Control and Prevention.
- US CDC (2005). MMWR. Morbidity and mortality weekly report, **54**(20), 513–516.
- Wheeler, M. W. and Bailer, J. A. (2009). Comparing model averaging with other model selection strategies for benchmark dose estimation, *Environmental and Ecological Statistics*, **16**, 37–51.
- WHO (1995). Environmental Health Criteria 165–Inorganic Lead. Geneva: International Programme on Chemical Safety, World Health Organization.