

집단 약동학 모형을 위한 모형 진단과 적합도 검정에 대한 고찰

이은경¹

¹이화여자대학교 통계학과

(2011년 10월 7일 접수, 2011년 12월 12일 수정, 2012년 1월 31일 채택)

요약

집단 약동학 모형 추정의 결과는 환자에게 투약학 약물의 용량결정에 직접적 영향을 미치므로 추정 모형에 대한 타당도와 적합도의 검증이 중요하다. 본 논문에서는 다양한 집단 약동학 모형 적합도 검증을 위한 방법들을 비교, 분석하고 실제 임상자료를 이용하여 최적의 집단 약동학 모형을 찾고 이에 대하여 다양한 타당도, 적합도 검정을 실시하여 모형을 진단해 본다.

주요어: 집단 약동학, 혼합효과모형, 비선형모형, 모형진단.

1. 서론

약동학(pharmacokinetics)이란 약물의 흡수, 분포, 생체 내 변화 및 배설을 연구하는 분야로 약물투여 후 시간에 따른 혈중농도의 변화를 약동학 모형으로 나타낸다. 약동학 모형은 약동학적인 특성을 나타내는 약동학 모수인 분포용적(Volume of distribution; V), 청소율(Clearance), 소실속도상수(elimination rate constant; K), 흡수속도상수(absorption rate constant; K_a) 등의 함수로 이루어져 있으며 투약 후 시간에 따라 측정된 혈중농도자료를 모형에 적합시킴으로써 얻을 수 있다. 약동학 모수들은 개개인마다 다르며, 특히 나이, 성별, 인종 등의 인구학적 특성이나 병리학적 특성에 따라 다른 경향을 나타낸다. 이러한 개인차를 고려한 접근 방법을 집단 약동학이라고 부른다. 집단 약동학 모형은 개인 내 변동과 개인 간 변동을 고려한 비선형 혼합효과모형의 형태로 구현되며 이를 추정하기 위한 다양한 추정 방법들이 개발되어 있다 (이은경, 2010).

집단 약동학 모형을 추정하는 가장 큰 목적은 자료를 얻은 특정 영역, 즉 특정 환자집단 내에서 약물의 약동학적 특성을 예측하여 이를 실제 임상에 적용하는 것이다. 즉, 집단 약동학 모형 추정으로부터 얻은 약동학 모수들은 환자에게 투약학 약물의 용량 등의 결정에 직접 영향을 미치게 된다. 그러므로 모형에 대한 타당도와 적합도의 검증이 중요하다. 집단 약동학 자료분석의 결과로 얻은 최종 모형은 현재 상황에서 최선의 모형일 뿐 결코 참모형(true model)이 될 수 없으므로 관심있는 모집단을 객관적으로 잘 예측하고 참모형과의 차이가 통계적 추론에 큰 영향을 미치지 않는 모형이 최선의 모형이라고 할 수 있다. 추정한 모형이 최선의 모형인지를 판단하기 위한 여러가지 방법들이 제안되어 있다.

이 논문은 2010년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2010-0003840).

¹(120-750) 서울특별시 서대문구 이화여대길 52, 이화여자대학교 통계학과, 조교수.

E-mail: lee.eunk@ewha.ac.kr

Ette (1997, 1998)는 처음으로 집단 약동학 모형평가의 중요성을 인식하고 여러가지 기본적인 방법들을 제시하였다. 1999년 FDA는 이에 대한 가이드라인을 제시하였고 그 이후 본격적인 연구가 이루어졌다. Ette 등 (2003)과 Brendel 등 (2007)에서는 본격적으로 집단 약동/약력학 모형의 적합도 검증에 대하여 논의를 하고 있으며 특히 Brendel 등 (2007)에서는 2002년부터 2004년간 출판된 논문에서 사용한 적합도 검증 방법들을 비교하고 있다.

본 논문에서는 주로 쓰이고 있는 집단 약동학 모형 평가 방법들을 여러 관점에서 살펴보고 실제 자료를 이용하여 평가를 하여 비교, 분석하고자 한다. 2절에서는 집단 약동학 모형을 정의하고 이를 이용하여 약동학 자료분석을 하는 과정을 살펴본다. 3절에서는 기본적인 적합도 평가 방법에 대해 살펴보고 이들을 비교해 본다. 4절에서는 자료를 분할하여 적합도를 평가하는 방법을 살펴보고 5절에서는 모의시험을 이용한 적합도 평가 방법을 살펴본다. 6절에서는 실제 자료를 이용하여 최적의 약동학 모형을 찾고 선택한 최적의 모형에 대하여 여러가지 적합도 평가를 실시한 후 비교, 분석해 본다.

2. 집단 약동학 모형 적합

가장 널리 쓰이고 있는 집단 약동학 모형은 혈장농도가 커짐에 따라 오차분산도 증가하는 형태인 비례오차모형(proportional error model)로 다음과 같이 나타낼 수 있다.

$$\begin{aligned} Y_{ij} &= f(D_i, t_{ij}, \theta_i)(1 + \epsilon_{ij}) & (2.1) \\ \log \theta_i &= \log \theta + \eta_i \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \eta_i &\sim N(0, \Omega), \end{aligned}$$

여기서 Y_{ij} 는 i 번째 환자의 j 번째 측정 혈중농도를 나타내며 t_{ij} 는 i 번째 환자의 j 번째 채혈 시간, D_i 는 i 번째 환자에게 투여된 용량, 그리고 θ_i 는 i 번째 환자의 약동학 모수, θ 는 집단 약동학 모수를 나타낸다. η_i 는 약동학 모수의 개인간 변동을 나타내며 이는 평균이 0이고 분산이 Ω 인 정규분포를 따른다고 가정한다. ϵ_{ij} 는 i 번째 환자의 j 번째 혈중농도 측정에 대한 오차를 나타내며 이는 평균이 0이고 분산이 σ^2 인 정규분포를 따른다고 가정한다. 또한 f 는 임의의 약동학 식으로 대체로 일구획, 이구획, 또는 다구획 모형으로 정의되며 약물의 투여방법이나 구획모형에 따라 다른 형태의 함수모양을 나타내며 이는 약동학 모수들에 대한 비선형 함수의 형태로 나타난다.

집단 약동학 자료 분석의 첫 단계는 식 (2.1)에서 자료를 가장 잘 설명하는 약동학 함수 f 를 찾는 것이다. 이를 위하여 먼저 탐색적 자료분석과정이 필요하다. 이는 집단 약동학 자료를 모형적합 이전에 살펴보는 과정으로 가능한 모형들을 짐작해 볼 수 있게 된다. 이를 위해 가장 유용한 그림은 각 환자별로 시간에 따른 혈중농도의 변화를 살펴볼 수 있는 혈중농도곡선으로 일구획, 이구획 또는 다구획 모형 중 어떤 모형이 적합할지를 대략적으로 판단할 수 있게 된다. 탐색적 자료분석 과정을 통하여 유추한 모형을 바탕으로 다양한 형태의 약동학 함수와 여러가지 형태의 오차구조를 시도해보게 되며 이 중 자료를 가장 잘 설명하는 약동학 함수를 찾는다. 이를 기본모형(base model)이라고 부른다. 기본모형을 선택은 대체로 $-2\log\text{likelihood}$ 의 형태로 정의되는 목적함수를 이용한다. 이 목적함수값이 가장 작은 모형이 최적의 기본모형이 된다.

앞에서 선정된 기본모형을 바탕으로 약동학 모수의 개인간 변동을 나타내는 η_i 를 empirical bayesian 방법으로 추정한다. 이를 EBE(Empirical Bayesian Estimates)라고 부른다. EBE 값에는 집단 약동학 모수로 설명되지 못한 개인간 차이가 남아있으므로 이 개인간 차이를 나이, 성별 등의 인구학적 특성이거나 병리학적 특성을 나타내는 공변수들로 설명하는 것이 집단 약동학 분석의 두 번째 단계이다. i 번째

환자에 대해 k 개의 공변수 x_{1i}, \dots, x_{ki} 를 이용한 모형은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \log \theta_i &= \log \theta + \eta_i \\ &= \log \theta + \alpha_1 x_{1i} + \dots + \alpha_k x_{ki} + \eta_i^* \end{aligned} \tag{2.2}$$

먼저 공변수들과의 관계 파악을 위해 다양한 방법을 이용한다. EBE 값들과 공변수들의 산점도를 그려보거나 그들간의 상관계수 등을 구해보기도 한다. 다중회귀분석이나 일반화 가법모형과 같은 방법을 이용하기도 한다. 또한 AIC 기준의 stepwise procedure를 사용하기도 하고 (Liddle, 2007) Wald approximation method와 같은 방법을 사용하기도 한다 (Kowalski와 Hutmacher, 2001). 그러나 공변수의 선택 과정은 단순한 통계적 절차보다는 임상적인 지식이 더 많이 적용되어야 한다. 공변수의 선택 과정까지 마치고 나면 최종의 집단 약동학모형을 선정하게 된다.

위의 두 단계를 거쳐 선택된 최종모형은 다음 단계의 임상시험이나 신약 승인 절차에 직접적으로 영향을 주게 되므로 최종 모형에 대한 평가가 필요하게 된다. 선택된 최종 모형은 현재 상황에서 최선의 모형일 뿐 결코 참모형(true model)이 될 수 없으므로 관심있는 모집단을 객관적으로 잘 예측하고 참모형과의 차이가 통계적 추론에 큰 영향을 미치지 않는지를 잘 파악하는 것이 모형평가의 목적이 된다.

3. 기본적인 적합도 평가 방법

3.1. 그래프를 이용한 적합도 평가 방법

모형 적합 후 가장 쉽고 편리하게 적합도를 평가할 수 있는 방법은 그래프를 이용하는 방법이다. 집단 약동학 모형에서 약동학 모수들의 랜덤효과를 고려하지 않는 추정치를 집단 추정치(population prediction; PRED, $f(D_i, t_{ij}, \hat{\theta})$)라고 하며 랜덤효과까지 고려한 추정치를 개인추정치(individual prediction; IPRED, $f(D_i, t_{ij}, \hat{\theta}_i)$)라고 한다. 잔차도 관측치(observation; OBS)와 집단 추정치의 차이인 집단오차(population residual; RES)와 개인추정치와의 차이인 개인오차(individual residual; IRES)의 두 가지가 있다. 이들간의 다양한 그래프를 이용하여 모형의 적합도를 평가할 수 있다 (Ette, 1998; Karlsson과 Savic, 2007).

3.1.1. 관측치와 추정치들의 산점도 가장 먼저 그려볼 수 있는 그림은 관측치와 집단추정치(OBS vs. PRED)의 산점도와 관측치와 개인추정치(OBS vs. IPRED)의 산점도이다. 이들은 추정된 모형으로부터의 추정치가 관측치와 얼마나 일치하는지를 나타내주는 그래프로 일반적으로 $y = x$ 의 기준선과 함께 그린다. 적합도에 대한 기준은 $y = x$ 의 기준선 주변에 자료들이 퍼져있는 정도로 파악할 수 있다. 그러나 기준선 주변에 퍼져있는 정도에 대한 기준이 모호하다. 그러므로 이들 그림을 해석하는 데에는 여러 가지 문제점들이 있다. PRED는 랜덤효과를 고려하지 않은 추정치이므로 좋은 모형의 경우에도 개인간의 차이까지 반영되어 있는 OBS와의 산점도에서 기준선과는 동떨어진 패턴을 보일 수도 있다. 또한 이러한 패턴은 자료의 성질마다, 모형마다 달라지므로 기준이 되는 패턴을 찾기가 쉽지 않다. IPRED의 경우에는 랜덤효과가 포함되어 있어 PRED와의 산점도보다는 훨씬 더 기준선에 가까이 있게 된다. 그러나 상대적으로 과적합(overfit)을 밝혀내기 힘들고 또한 자료가 충분하지 않을 때에는 이 문제가 더욱 심각해 진다. 과적합이 일어나는 경우 오차항의 표준편차에 대한 추정치인 $\hat{\sigma}$ 에 비하여 IRES들의 표준편차가 작아지므로 이를 나타내는 값인 ϵ -shrinkage를 이용한다 (Savic 등, 2006; Karlsson과 Savic, 2007).

$$\epsilon_{sh} = \left(1 - \frac{SD(IRES)}{\hat{\sigma}}\right) \times 100. \tag{3.1}$$

이 값이 20~30% 이상이면 $\hat{\sigma}$ 에 비해 IRES의 표준편차가 매우 작은 것이므로 IPRED 값을 믿을 수 없게 된다.

추정치를 살펴보는 또 다른 방법은 시간(TIME)에 따른 변화를 나타내는 그림을 그리는 것이다. TIME vs. OBS의 산점도를 그리고 TIME과 PRED의 관계를 선으로 나타내어 시간에 따라 어떻게 추정이 되었는지를 파악하는 것이다. 또한 각 환자 별로 TIME vs. OBS를 선으로 나타내고 이 그림에 TIME vs. PRED와 TIME vs. IPRED의 두 선을 추가하여 각 개인별 적합도를 관찰하는 것도 필요하다. 이를 통하여 환자의 특성이 랜덤효과에 잘 반영되었는지를 확인할 수 있다 (Ette, 1998).

3.1.2. 잔차의 검토 PRED와 IPRED로부터 계산된 잔차인 RES와 IRES는 PRED, IPRED의 산점도와 마찬가지로 여러가지 문제점들이 있다. RES vs. PRED 산점도의 경우 기본패턴을 파악하기 힘들어 좋은 모형을 좋지 않게 판단할 가능성이 있고 IRES vs. IPRED 산점도는 과적합(overfit)을 가려내지 못해 좋지 않은 모형을 좋게 판단할 가능성이 높다. 그러므로 이를 보완하여 IRES를 표준화 한 잔차인 가중잔차를 이용한다. 모형추정방식에 따라 추정 분산이 달라지므로 표준화 잔차 식도 달라진다 (이은경, 2010). 약동학모수의 최대우도추정치를 구하기 위하여 개인간 차이를 나타내는 $\eta_i = \eta^*$ 에 대하여 테일러시리즈 전개에 이용한다. 이 전개의 1차식을 사용하고 $\eta^* = 0$ 을 가정하는 First Order(FO) 방법을 이용하여 모형을 추정한 경우의 표준화 잔차를 가중잔차(weighted residual; WRES)라고 하고, 1차식은 사용되지 $\eta^* = \hat{\eta}_i$ 로 가정하는 First Order Conditional Estimation(FOCE) 방법을 이용한 경우는 조건부 가중 잔차(conditional weighted residual; CWRES)라고 부른다.

3.2. 모수 추정치의 검토

모형적합으로 얻은 모수의 추정치는 환자에게 투여할 약물의 용량, 시간 간격 등을 결정하는데에 결정적인 역할을 하는 값이므로 이 추정치가 얼마나 정확한 값인지 정확도를 파악하는 것이 중요하다. 이를 위하여 추정치의 표준오차를 이용한다. 모형적합으로 얻은 약동학 모수들의 추정치는 모두 우도함수를 목적함수로 하여 이를 최대화 하는 값을 구한 것으로 최대우도추정치(Maximum likelihood estimator; MLE)가 되므로 Fisher information을 이용하여 표준오차를 구할 수 있다. 또한 최대우도추정치가 근사적으로 정규분포를 따른다는 성질을 이용하여 각 모수들의 신뢰구간도 구할 수 있다 (Sheiner, 1986). 추정치의 정확도를 파악할 수 있는 표준오차를 구하는 또 다른 방법은 Bootstrap 방법을 이용하는 것이다. 이는 resampling을 이용한 방법으로 각 모수들의 신뢰구간도 구할 수 있다. 모수들의 신뢰구간을 구하는 또 다른 방법은 Log-likelihood profiling(LLP) 방법이다 (Lindbom, 2006).

3.2.1. Bootstrap Bootstrap은 추정 모수의 정밀도를 파악하기 위한 통계적 방법으로 원 자료로부터 복원을 허용하여 추출한 bootstrap sample들을 이용하여 같은 모형을 적합시켜 모수의 추정치를 구하게 된다. 이러한 bootstrap sample을 반복하여 생성하고 이들로부터 구한 모수의 추정치들을 모아서 추정 모수의 표준오차를 구하고 추정모수의 분포도 파악할 수 있다. 이를 이용하여 신뢰구간을 구할 수 있다 (Yafune과 Ishiguro, 1999; Efron과 Tibshirani, 1986; Parke 등, 1999).

3.2.2. Log-likelihood profiling(LLP) LLP 방법은 모형 중 모수의 민감도 분석을 위해 쓰이는 방법으로 모수의 안정성, 정확도 등을 확인하고 모수공간의 형태를 확인할 수 있다. 최대우도추정치가 정규분포를 따른다는 가정 하에서 계산되는 신뢰구간이나 bootstrap 방법에 의해 구하는 신뢰구간은 모두 구간이 대칭적이어야 한다는 가정 하에서 분포에서 꼬리의 확률을 같게 하여 상한과 하한을 결정한다.

다. 이에 반하여 Log-likelihood profiling(LLP) 방법은 집단 약동학 모수의 신뢰구간을 구하는 또 다른 방법으로 다른 두 신뢰구간과는 달리 신뢰구간이 대칭적이어야 한다는 가정을 하지 않고 구하는 방법이다 (Holford와 Peace, 1992). 먼저 신뢰구간을 구하고자 하는 특정 모수를 추정치 근처의 값(추정치의 $\pm 5\%$, 10% , 20% , 30% 등에 해당하는 값)으로 고정시키고 각 고정값에 대하여 최종 모형을 다시 적합시켜 나머지 모수들을 추정하고 이 때의 최대우도함수의 값과 모든 모수들을 추정할 때의 최대우도함수 값과의 차이를 구한다. 이 차이는 근사적으로 자유도 1인 카이제곱분포를 따른다고 가정할 수 있으므로 이 성질을 이용하여 특정 모수의 신뢰구간을 구하게 된다. 먼저 모수의 고정값과 이때의 목적함수값의 차이를 그림으로 그리고 목적함수값의 차이가 3.84가 되는 지점에 선을 그린다. 이 차이가 3.84보다 작아지는 구간의 해당되는 모수의 고정값이 선택된 모수의 95% 신뢰구간이 된다.

4. 자료를 분할하는 방법들

4.1. Cross-validation

일반적인 통계모형 타당도 검정에서 사용되는 cross-validation은 자료를 n 개의 그룹으로 나누어 놓고 한번에 한 개의 그룹을 제외한 나머지 자료를 가지고 모형을 적합시킨 후 적합에서 제외된 한 개 그룹의 자료를 이용하여 예측오차를 구하여 모형적합이 잘 되었는지를 판단하는 방법으로 10-fold validation이 대표적으로 쓰이는 방법이다. 이 방법은 집단약동학 모형 타당도 검정에 그대로 적용하기 힘든 점들이 있다. 대부분 자료가 충분하지 않으며 환자 단위로 모형을 적합해야 하므로 각 환자들을 크기가 같은 n 개의 그룹으로 나누어야 하는데 이는 쉽지 않다. 특히 1상 임상시험자료인 경우는 피험자수가 20명을 넘는 경우가 많지 않으므로 더욱 더 힘들다. 이러한 집단약동학 자료의 특성을 반영한 cross-validation 방법을 개발하여 이용하고 있다. 전체 환자를 두 그룹으로 나누어 한 그룹의 자료는 모형을 적합하는데 사용하고 다른 한 그룹은 모형을 예측하는 데에 사용한다. 대체로 모형을 적합하는 쪽 그룹에는 환자의 80% 정도가 할당이 되도록 하고 다른 그룹에는 20% 정도의 환자가 할당이 되도록 한다. 이렇게 두 그룹으로 나누는 것을 랜덤하게 수차례 반복하여 모형 적합과 예측을 반복하고 이 결과들을 종합하여 선택된 모형이 자료를 잘 설명하고 있는지 판단을 하게 된다. 집단약동학에서의 cross-validation은 기존의 cross-validation과 resampling 기법을 혼합한 방법이라고 할 수 있다.

4.2. Case deletion diagnostics

Case deletion diagnostics(CDD)는 모수 추정에 가장 영향을 주는 관측값이 어떤 것인지를 파악하기 위해 이용하는 방법으로 한번에 하나의 관측값을 제외하고 추정을 하여 자료 전체를 사용했을 때의 추정치와의 차이를 비교하는 방법이다. 집단 약동학에서는 한 명의 피험자 단위로 모형을 추정해야 하므로 모형 추정에 영향을 주는 피험자를 찾는 방법으로 사용된다. 각 피험자의 영향력을 평가하기 위해서는 Cook's distance나 Covratio와 같은 값들을 사용한다.

4.3. Jackknife

Jackknife 방법은 CDD와 비슷하게 한번에 하나의 관측값을 제외하고 모형을 추정하는 방법이다. 그러나 CDD와는 달리 관측값의 영향력보다는 모수 추정시 발생할 수 있는 편의(bias)를 구하기 위해 사용된다. 집단 약동학 자료분석에서는 한 번에 한 명의 피험자를 제외하고 모형을 추정하여 구한 추정치와 원 자료를 모두 이용했을 때의 추정치의 차이를 이용하여 구하고자 하는 모수에 대한 편의를 구한다 (Sadray 등, 1999).

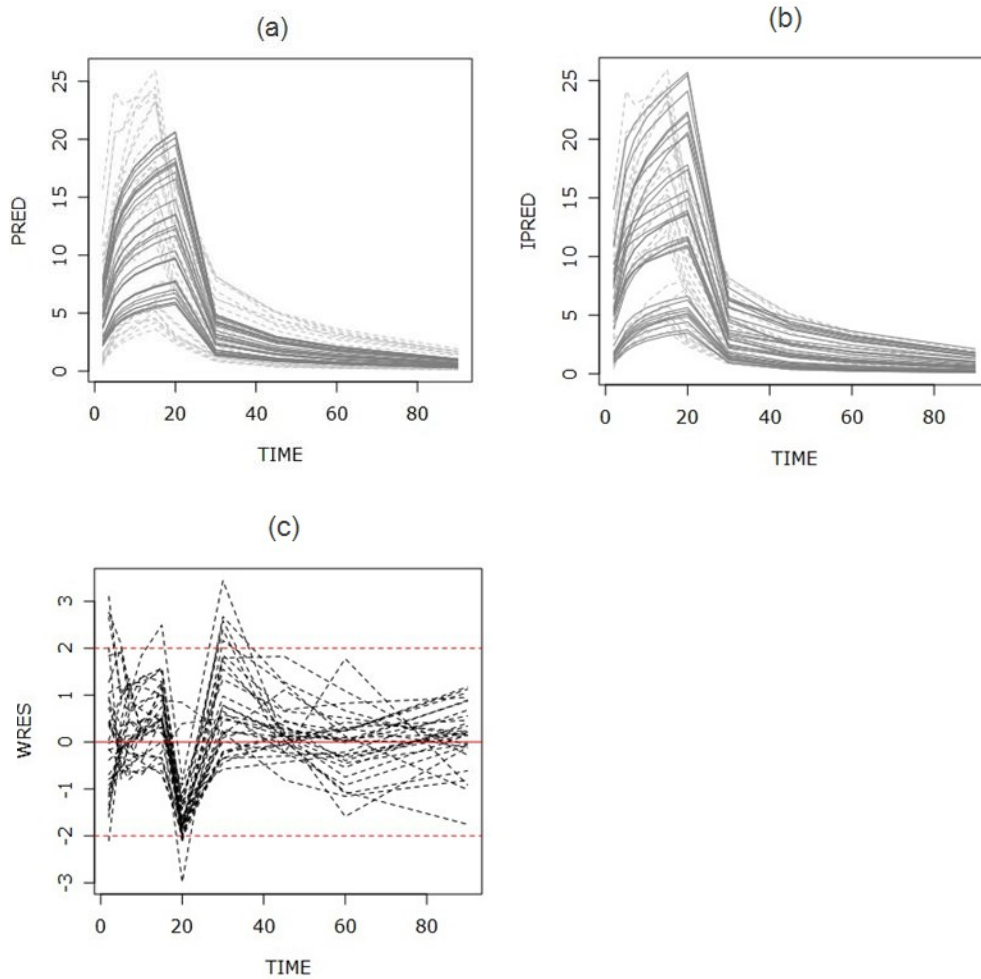


그림 6.1. 기본 모형의 적합도 평가를 위한 그림들

5. Simulation을 이용한 방법: Visual Predictive Checks

Visual predictive checks(VPC)는 최근 각광받고 있는 방법으로 실제 관측값과 최종모형으로부터 simulation을 통하여 얻은 자료간의 차이를 시각적으로 평가하는 방법으로 추정모형이 원 자료의 타당성을 재현할 수 있는지를 검토하는 방법이다 (Yano 등, 2001; Holford, 2005, 2008; Post 등, 2008). 가장 일반적인 VPC 그림은 TIME vs. OBS의 산점도 위에 1000번의 모의시험으로부터 얻은 자료의 각 시간 별 5%, 50%, 95% 분위값을 연결한 선을 그려서 실제 관측값이 5%와 95% 분위값 바깥쪽에 각각 5%씩 존재한다면 최종모형이 자료를 잘 설명하고 있다고 할 수 있다.

6. 예제: Microemulsion propofol

Propofol은 수면마취제로 쓰이는 것으로 Microemulsion propofol은 새로 개발된 formula이다 (Lee 등, 2008). 30마리의 쥐를 10마리씩 세 그룹으로 나누어 각각 0.5, 1.0, 1.5mg/kg/min으로 20분간 지속

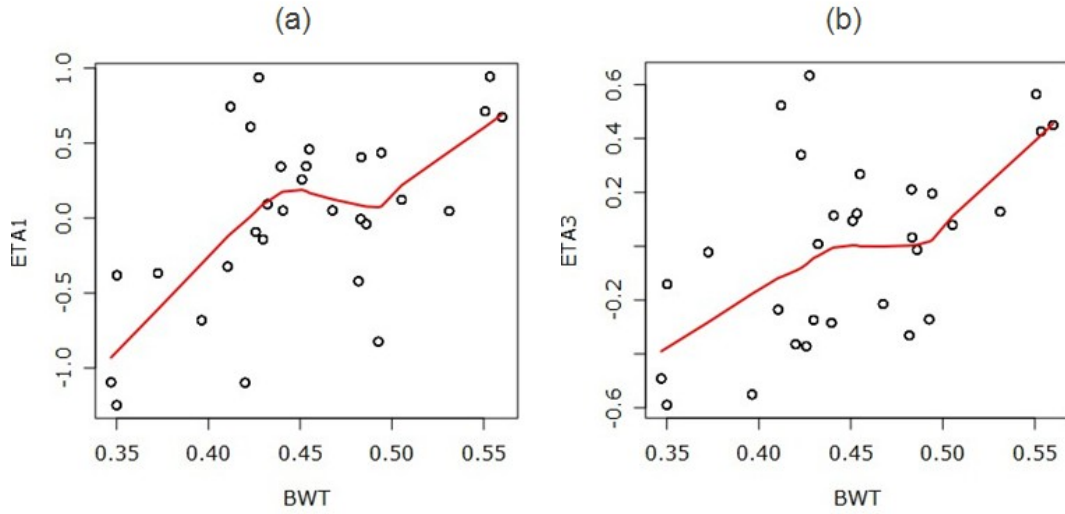


그림 6.2. 기본모형의 ETA 추정치와 공변수 BWT와의 산점도

정주하고 지속정주 시작부터 90분간 혈중약물농도를 측정된 자료이다. 기본모형으로는 2구획 지속정주 모형을 선택하였다. 모형에 이용된 기본모형의 약동학 모수는 V_1 , V_2 , CL_1 , 그리고 CL_2 이다. 약동학 모수 추정을 위해 NONMEM의 FO 방법을 이용하였다. 기본 모형에서 사용한 모수의 분포식은 다음과 같다.

$$\begin{aligned}
 V_1 &= \theta_1 \exp(\eta_1), \quad \eta_1 \sim N(0, \omega_1), \\
 V_2 &= \theta_2, \\
 CL_1 &= \theta_3 \exp(\eta_3), \quad \eta_3 \sim N(0, \omega_3), \\
 CL_2 &= \theta_4.
 \end{aligned}$$

그림 6.1은 기본 모형에 대한 적합도 검정을 나타내는 그림으로 (a)는 관측값과 PRED의 시간에 따른 변화를 나타내는 profile 그림이고 (b)는 IPRED의 시간에 따른 변화를 나타내는 profile 그림이다. 점선으로 나타난 부분은 원자료의 시간에 따른 변화를 나타내는 profile을 나타낸다. PRED의 경우 혈중농도가 최고가 되는 지점과 급격히 떨어지는 지점에서 많은 차이를 보인다. IPRED는 최고 지점과 급격히 떨어지는 지점 모두 PRED 보다 원 자료와 일치하는 것을 알 수 있다. (c)는 시간에 따른 WRES의 변화를 나타내는 그림으로 처음 농도가 올라가는 지점, 최고가 되는 지점, 그리고 급격히 떨어진 지점에서 기준선인 -2와 2를 넘어가는 것을 볼 수 있다.

기본 모형에서 η_1 과 η_3 의 추정치를 구하여 이들과 공변수인 쥐의 몸무게(weight)간의 산점도를 그려 관계를 살펴보았다. 그림 6.2에서 η_1 과 η_3 모두 weight와 선형관계를 보이고 있음을 알 수 있다. 이를 토대로 몇가지 모형을 시도해 보았다. Model 1은 V_1 에 해당되는 η_1 에 weight와의 선형관계를 넣은 것이고 Model 2는 CL_1 에 해당되는 η_3 에 weight와의 선형관계를 넣은 것이다. Model 3은 이 두가지를 모두 넣은 모형이다 (표 6.1). 표 6.2은 각 모형을 NONMEM의 FO 방법을 이용하여 추정했을 때의 목적함수값과 모수의 추정치, 그리고 추정치의 표준오차를 정리한 표이다.

Model 1과 Model 2는 모두 8개의 모수를 추정하고 있으며 목적함수의 값도 비슷하다. Model 3은 9개의 모수를 추정하여 Model 1, Model 2 보다 하나의 모수를 더 추정하고 있으나 목적함수의 값은 10정

표 6.1. 시도된 여러가지 모형들

Model 1	$V1 = \theta_1 \exp(\theta_5 * BWT + \eta_1)$
Model 2	$CL1 = \theta_3 \exp(\theta_6 * BWT + \eta_3)$
Model 3	$V1 = \theta_1 \exp(\theta_5 * BWT + \eta_1)$ $CL1 = \theta_3 \exp(\theta_6 * BWT + \eta_3)$

표 6.2. 모형에 따른 목적함수 값과 약동학 모수의 추정치, 그리고 표준오차

Parameter	Base model	Model 1	Model 2	Model 3
OBJ value	318.692	307.369	308.541	297.03
TH1(V1)	0.1430 (0.018400)	0.0082 (0.006360)	0.1430 (0.018600)	0.0081 (0.007810)
TH2(V2)	0.4220 (0.020500)	0.4220 (0.020000)	0.4230 (0.020500)	0.4220 (0.020600)
TH3(CL1)	0.0249 (0.001560)	0.0249 (0.001490)	0.0061 (0.002460)	0.0065 (0.002750)
TH4(CL2)	0.0176 (0.000665)	0.0176 (0.000684)	0.0176 (0.000727)	0.0176 (0.000682)
TH5		6.3500 (1.610000)		6.3700 (2.030000)
TH6			3.1200 (0.852000)	2.9800 (0.895000)
OMEGA1	0.3890 (0.624000)	0.2580 (0.508000)	0.3890 (0.623000)	0.2540 (0.504000)
OMEGA3	0.1140 (0.338000)	0.1140 (0.337000)	0.0805 (0.028400)	0.0798 (0.283000)
SIGMA	0.0363 (0.191000)	0.0363 (0.190000)	0.0363 (0.190000)	0.0363 (0.191000)

도 차이가 난다. 그러므로 본 논문에서는 Model 3을 최종 모형으로 선택하기로 하고 Model 3에 대한 적합도 검정을 실시해 보고자 한다.

그림 6.3은 Model 3에 대한 적합도 검정을 나타내는 그림이다. (a)는 PRED의 시간에 따른 변화를 나타내는 profile 그림으로 기본모형에서의 그림과는 달리 용량별로 확연히 다른 반응을 나타냄을 확인할 수 있다. 반면 (b) IPRED와의 그림에서는 기본모형에서와 크게 다르지 않음을 볼 수 있다. (c)의 시간에 따른 WRES 변화 그림에서는 기본모형에서 기준선인 -2와 2를 넘어갔던 지점에서 대부분의 값들이 기준선 안에 있음을 볼 수 있다.

표 6.3은 여러가지 방법으로 구한 약동학 모수들의 추정치를 비교한 것이다. 원 자료를 이용한 방법과 Bootstrap, 그리고 cross validation은 비슷한 결과를 나타내고 있다. 반면 Jackknife의 경우 BWT를 공변수로 이용한 TH1, TH3, 그리고 TH5, TH6의 값은 다른 방법으로 부터 얻은 추정치와 상당한 차이를 보이고 있다. 이는 몇몇의 피험자 자료가 모수결정에 큰 영향을 미치고 있음을 나타내는 것으로 CDD로 이를 확인할 수 있다.

그림 6.4는 각 피험자에 대한 Covratio와 Cook's distance를 산점도로 나타낸 것이다. 12번 피험자의 경우 covratio는 크지 않으나 Cook's distance 값이 다른 피험자들에 비해 상당히 크다는 것을 알 수 있다. 또한 18번 피험자의 경우 Cook's distance는 크지 않은 반면 Covratio가 상당히 크다. 그림 6.5는 이 두 피험자의 원 자료의 profile(원형 점이 있는 검은선)과 PRED(넓은 점선), 그리고 IPRED(좁은 점선) profile을 보여주고 있다. 12번 피험자 자료의 경우 IPRED, PRED 모두 원 자료의 최고농도에 못미치는 농도를 예측하고 있으며 18번 피험자의 경우 최고농도의 양은 예측하고 있으나 최고농도의 도달 시간을 예측하지 못하고 있다.

표 6.4는 Fisher information을 이용한 근사적 방법으로 구한 모수의 95% 신뢰구간, Bootstrap 방법으로 구한 신뢰구간, Jackknife 방법으로 구한 신뢰구간, 그리고 log-likelihood profiling 방법으로 구한 신뢰구간을 나타내고 있다. TH1의 경우 근사적 방법과 Jackknife 모두 0을 포함하는 구간을 나타내고 있으며 Bootstrap 방법은 그렇지 않다. TH2의 경우 근사적 방법과 Bootstrap 방법, 그리고

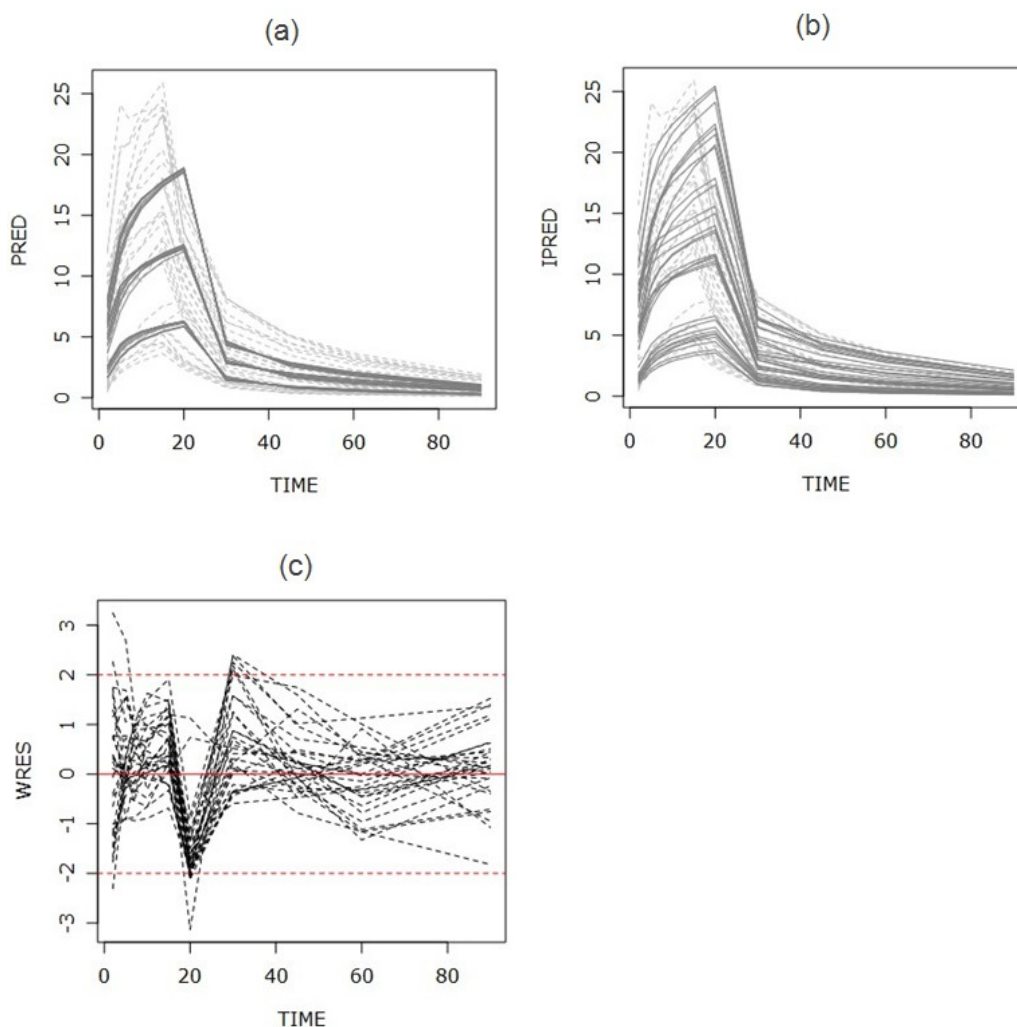


그림 6.3. 최종 모형의 적합도 평가를 위한 그림들

표 6.3. Bootstrap, Jackknife, CV를 이용한 모수의 추정치 비교

Parameter	Original est.	Bootstrap	Jackknife	Cross Validation
TH1(V1)	0.0081	0.0089	-0.0021	0.0088
TH2(V2)	0.4220	0.4199	0.4094	0.4224
TH3(CL1)	0.0065	0.0062	0.0141	0.0062
TH4(CL2)	0.0176	0.0176	0.0185	0.0176
TH5	6.3700	6.5837	8.1293	6.3050
TH6	2.9800	3.1478	0.1090	3.1070

LLP 방법은 비슷한 구간을 나타내고 있으며 Jackknife 방법은 다른 구간을 나타내고 있다. TH3에서는 Bootstrap이 가장 좁은 구간을 나타내고 있으며 근사적인 방법은 Bootstrap 보다 조금 넓은 구간을 나

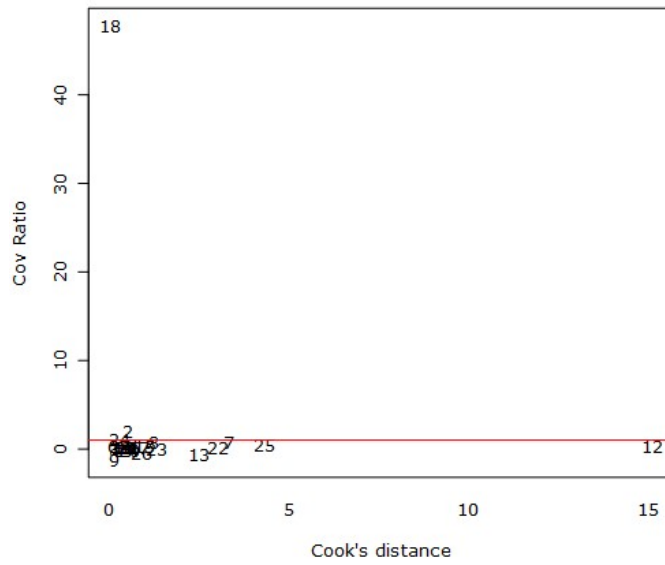


그림 6.4. CDD 결과: Covratio와 Cook's distance의 산점도

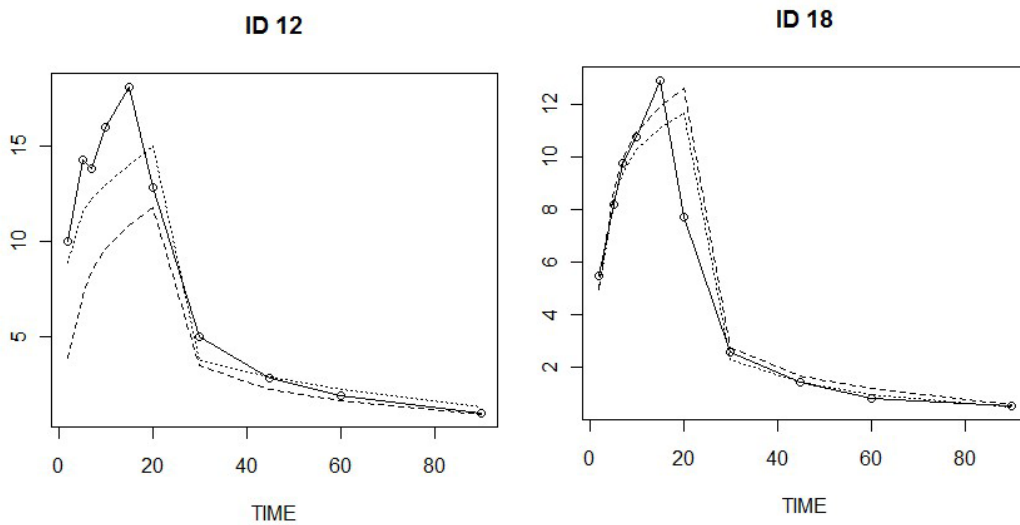


그림 6.5. 12번, 18번 피험자의 profile 그림. 원자료는 원형 점이 있는 검은선으로, PRED는 넓은 점선으로, 그리고 IPRED는 좁은 점선으로 나타남.

타내고 있다. Jackknife 방법은 추정모수의 차이로 구간도 매우 다르게 나타내고 있다. TH4의 경우도 TH2와 비슷하다. TH5의 경우 Bootstrap이 가장 좁은 구간을, 그 다음이 LLP, 그리고 근사적 방법이 가장 넓은 구간을 나타내고 있다. TH6에서는 Bootstrap이 가장 좁은 구간을, 그리고 근사적인 방법이 그 다음을 나타내고 있으며 LLP는 조금 다른 구간을 나타내고 있다. 또한 Jackknife의 경우는 0을 포함하는 구간을 나타내고 있다. Bootstrap을 이용한 신뢰구간은 Bootstrap sample로 부터 구해진 실제

표 6.4. 근사적 방법, Bootstrap, Jackknife 그리고 LLP를 이용한 약동학 모수 95% 신뢰구간 비교

Parameter	근사적 방법	Bootstrap	Jackknife	LLP
TH1(V1)	(-0.0072, 0.0234)	(0.0021, 0.0264)	(-0.0176, 0.0134)	
TH2(V2)	(0.3816, 0.4624)	(0.3858, 0.4555)	(0.3688, 0.4501)	(0.389, 0.460)
TH3(CL1)	(0.0011, 0.0119)	(0.0033, 0.0112)	(0.0091, 0.0191)	
TH4(CL2)	(0.0163, 0.0189)	(0.0162, 0.0188)	(0.0170, 0.0199)	(0.016, 0.019)
TH5	(2.3912, 10.3488)	(3.7080, 9.3920)	(4.6762, 11.5824)	(3.031, 10.095)
TH6	(1.2258, 4.7342)	(1.8780, 4.3500)	(-1.5718, 1.7927)	(1.361, 5.153)

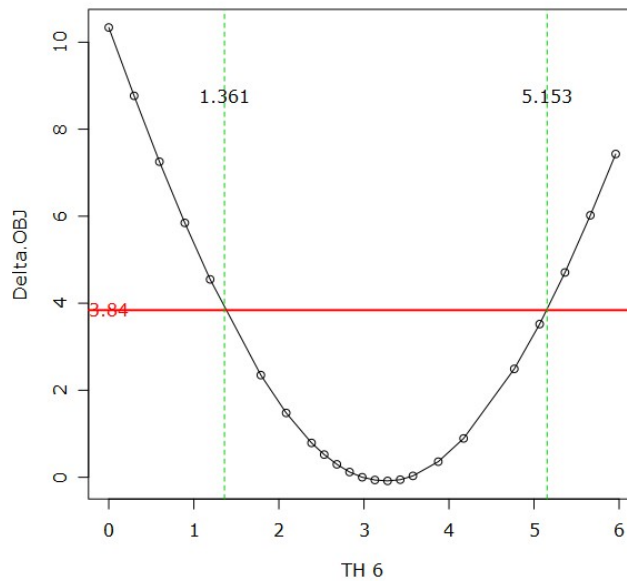


그림 6.6. TH6의 log-likelihood profiling 방법을 이용한 신뢰구간

추정치를 이용하고, 또한 LLP를 이용한 신뢰구간은 추정치 근방의 값으로 고정시킨 후 모형을 추정하여 얻은 목적함수 값을 사용하는 방법이므로 이 두 방법 모두 모수의 신뢰구간이 0을 포함하지 않게 된다. 즉, 상한, 하한 모두 0보다 크거나 같은 값을 갖게 된다. 그러나 근사적 방법이나 Jackknife 방법은 $\pm 1.96 * SE$ 의 방법을 사용하므로 하한이 0보다 작아질 수 있다. 그림 6.6은 TH6의 LLP 신뢰구간을 구하는 방법을 그림으로 나타낸 것이다. TH1과 TH3의 경우 모형이 TH1과 TH3 근방의 값에서 안정적으로 추정되지 않아 LLP 방법으로 신뢰구간을 구할 수 없었다.

마지막으로 그림 6.7은 visual predictive checks(VPC)의 결과를 그림으로 나타낸 것이다. (a)는 기본 모형에 대한 VPC 결과이고, (b)는 최종모형에 대한 VPC 결과이다. 회색영역은 모의시험으로부터 얻은 자료의 90% 영역을 의미한다. (a)에서 기본모형으로부터 얻은 모의시험 자료의 대부분은 원 자료보다 낮은 농도를 나타내고 있음을 알 수 있다. 반면 (b)로부터 최종 모형으로부터 얻은 모의시험자료의 90% 영역이 원자료의 대부분을 포함하고 있음을 알 수 있다. 5% 분위값 바깥쪽에는 거의 관측값이 없으나 95% 분위값 바깥쪽에는 5%정도의 관측값이 존재함을 알 수 있다. 이는 우리가 선택한 최종 모형이 나쁘지 않음을 나타낸다.

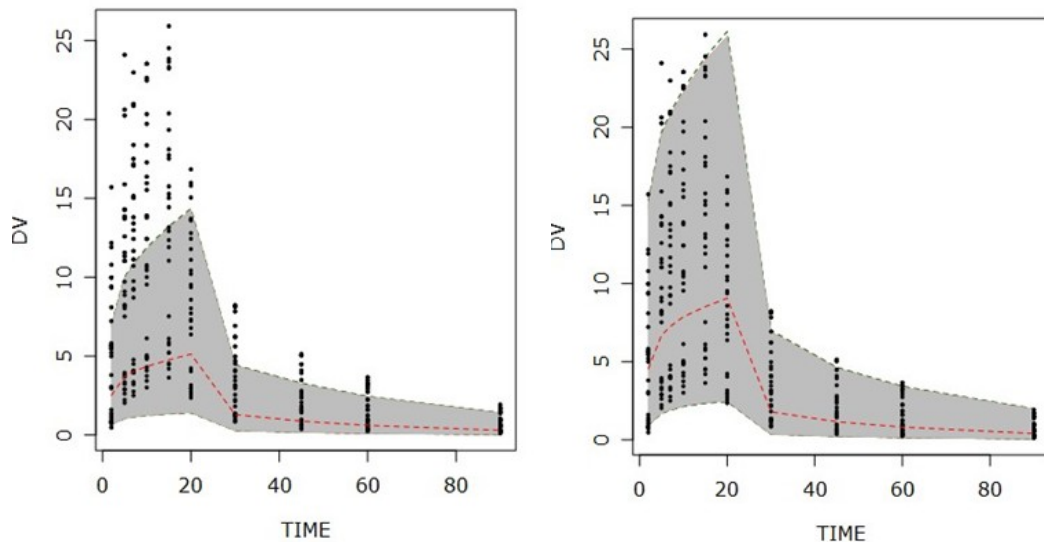


그림 6.7. 기본모형과 최종모형의 Visual predictive checks 결과

7. 토론

본 논문에서는 다양한 집단약동학 모형의 진단과 타당도 검증을 위한 방법을 살펴보고 실제 자료를 이용하여 약동학 모형을 추정하고 추정 모형에 대한 적합도 검정을 해보았다. 약동학 모형 추정을 위한 소프트웨어들이 다양한 추정방식으로 개발되어 있기는 하나 아직까지는 NONMEM이 가장 범용적으로 쓰이고 있는 소프트웨어이다. 그러나 NONMEM은 모든 결과물을 text 방식으로 제공하고 있어 추정 모형의 적합도 검정을 위해서는 또 한번 자료를 가공해야할 뿐 아니라 다른 소프트웨어를 이용해야한다.

이처럼 집단 약동학 자료분석은 모형추정과정은 대부분 NONMEM에 의존하고 있으며 모형추정 후의 모형검토과정을 위하여 또다른 소프트웨어를 사용해야하는 불편함이 있다. Jonsson과 Karlsson (1999)은 NONMEM을 이용하여 약동학 자료분석을 NONMEM을 이용하여 분석할 때 이를 도와주는 도구인 Xpose를 개발하였다. 텍스트 파일의 형태로 제공하는 NONMEM의 결과를 이용하여 여러가지 그림을 그려주고 간단한 모형의 적합도 검정을 할 수 있도록 되어 있다. 최근에는 Xpose에서도 새로운 적합도 검정 방법들이 제공되고 있기는 하지만 아직은 자유롭게 쓰기에는 한계가 있다. Xpose 이외에도 몇몇 소프트웨어들이 개발되어 있으나 범용적으로 쓰이고 있는 것은 없을 뿐 아니라 대부분의 소프트웨어들은 본 논문에서 제시한 다양한 적합도 검정 방법들 중 일부만을 제공하고 있다. 본 논문에서는 모형에서의 약동학 모수 추정은 NONMEM 7.2를 이용하였으며 그 이외의 적합도 검정 및 통계적 방법은 R을 이용하였다.

집단 약동학 모형은 대부분 자료의 갯수가 충분하지 않은 상황에서 복잡한 비선형 함수를 추정해야하는 문제점을 가지고 있어 모형적합이 불안정한 경우가 많다. 이로 인하여 다양한 진단방법을 모두 다 통과하지 못하더라도 임상적으로 의미가 있다고 판단이 되는 경우 최적의 모형으로 삼고 다음의 임상시험을 진행하는 경우도 발생하게 된다. 그러므로 한 두가지의 진단방법보다는 모형을 진단하기 위하여 여러가지 방법을 시도해 보는 것이 중요하다.

참고문헌

- 이은경 (2010). 집단 약동학 모형에 대한 통계학적 고찰, <응용통계연구>, **23**, 511-520.
- Brendel, K., Dartois, C., Comets, E., Lemenuel-Diot, A., Laveille, C., Tranchand, B., Girard, P., Laffont, C. M. and Mentre, F. (2007). Are population pharmacokinetic and/or pharmacodynamic models adequately evaluated? A survey of the literature from 2002 to 2004, *Clinical Pharmacokinetics*, **46**, 221-234.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statistical Science*, **1**, 54-77.
- Ette, E. I. (1997). Stability and performance of a population pharmacokinetic model, *Journal of Clinical Pharmacology*, **37**, 486-495.
- Ette, E. I. (1998). Statistical graphics in pharmacokinetics and pharmacodynamics: A tutorial, *The Annals of Pharmacotherapy*, **32**, 818-828.
- Ette, E. I., Williams, P. J., Kim, Y. H., Lane, J. R., Liu, M. J. and Capparelli, E. V. (2003). Model approximateness and population pharmacokinetic modeling, *Journal of Clinical Pharmacology*, **43**, 610-623.
- Holford, N. (2005). The visual predictive check-superiority to standard diagnostic(Rorschach) plots [abstr 738] *PAGE*, <http://www.page-meeting.org/?abstract=738>.
- Holford, N. (2008). Model Evaluation = Visual predictive checks [abstr 1434] *PAGE*, <http://www.page-meeting.org/?abstract=1434>.
- Holford, N. G. G. and Peace, K. E. (1992). Results and validation of a population pharmacodynamic models for cognitive effects in Alzheimer patients treated with tacrine, *Proceedings of the National Academy of Sciences*, **89**, 11471-11475.
- Jonsson, E. N. and Karlsson, M. O. (1999). Xpose - an S-plus based population pharmacokinetic/pharmacodynamic model building aid for NONMEM, *Computer Methods and Programs in Biomedicine*, **58**, 51-64.
- Karlsson, M. O. and Savic, R. M. (2007). Diagnosing model diagnostics, *Clinical Pharmacology and Therapeutics*, **82**, 17-20.
- Kowalski, K. G. and Hutmacher, M. M. (2001). Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test, *Journal of Pharmacokinetics and Pharmacodynamics*, **28**, 253-275.
- Lee, E., Lee, S., Park, D., Ki, K., Lee, E., Lee, D. and Noh, G. (2008). Physicochemical properties, pharmacokinetics, and pharmacodynamics of a reformulated microemulsion propofol in rats, *Anesthesiology*, **109**, 436-447.
- Liddle, A. R. (2007). Information criteria for astrophysical model selection, *Monthly Notices of the Royal Astronomical Society: Letters*, **377**, L74-L78.
- Lindbom, L. (2006). Development, application and evaluation of statistical tools in pharmacometric data analysis, *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy*, **33**, 46.
- Parke, J., Holford, N. H. G. and Charles, B. G. (1999). A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models, *Computer Methods and Programs in Biomedicine*, **59**, 19-29.
- Post, T. M., Freijer, J. I., Ploeger, B. A. and Danhof, M. (2008). Extensions to the visual predictive check to facilitate model performance evaluation, *Journal of Pharmacokinetics and Pharmacodynamics*, **35**, 185-202.
- Sadray, S., Jonsson, E. N. and Karlsson, M. O. (1999). Likelihood-based diagnostics for influential individuals in non-linear mixed effects model selection, *Pharmaceutical Research*, **16**, 1260-1265.
- Savic, R., Wilkins, J. J. and Karlsson, M. O. (2006). (Un)informativeness of empirical Bayes estimate based diagnostics [abstr T3360] *AAPS Journal*, **8(S2)**.
- Sheiner, L. B. (1986). Analysis of pharmacokinetic data using parametric models. III. Hypothesis tests and confidence intervals, *Journal of Pharmacokinetics and Biopharmaceutics*, **14**, 539-555.
- US FDA (1999). *Guidance for Industry - Population Pharmacokinetics*.
- Yafune, A. and Ishiguro, M. (1999). Bootstrap approach for constructing confidence intervals for population pharmacokinetic parameters. I: A use of bootstrap standard error, *Statistics in Medicine*, **18**, 581-599.
- Yano, Y., Beal, S. L. and Sheiner, L. B. (2001). Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check, *Journal of Pharmacokinetics and Pharmacodynamics*, **28**, 171-192.

Model Validation Methods of Population Pharmacokinetic Models

Eun-Kyung Lee¹

¹Department of Statistics, Ewha Womans University

(Received October 7, 2011; Revised December 12, 2011; Accepted January 31, 2012)

Abstract

The result of the analysis of a population pharmacokinetic model can directly influence the decision of the dose level applied to the targeted patients. Therefore the validation procedure of the final model is very important in this area. This paper reviews the validation methods of population pharmacokinetic models from a statistical viewpoint. In addition, the whole procedure of the analysis of population pharmacokinetics, from the base model to the final model (that includes various validation procedures for the final model) is tested with real clinical data.

Keywords: Population pharmacokinetics, mixed effect model, nonlinear function, model validation.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0003840).

¹Assistant Professor, Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong, Seodaemun-Gu, Seoul 120-750, Korea. E-mail: lee.eunk@ewha.ac.kr