

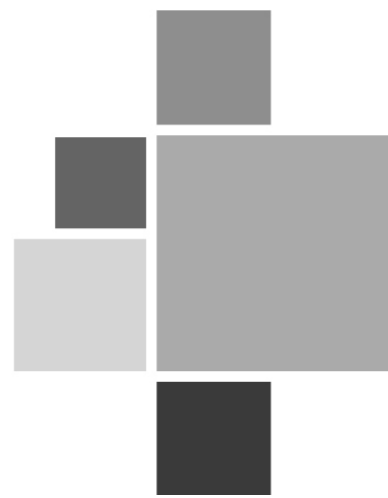
기저 공변량의 보정에 대하여 고려할 점

Points to consider on adjustment for baseline covariates

2009. 6.



식품의약품안전청
Korea Food & Drug Administration
기획조정관 통상통계담당관



인 사 말

최근 한미 FTA 의약품 협상 및 약가제도 조정 등 제약사업의 환경변화와 함께 국내외 의약산업 규모가 점차 커지고 있고, 다국가 임상시험 및 국내임상시험 실시가 증가하고 있습니다. 이에 따라 국가적으로도 신약 개발에 지원을 적극 확대하여 국내 임상시험의 질적 향상 뿐 아니라 임상자료의 활용을 더욱 높여야 할 것입니다.

식품의약품안전청에서는 유럽 의약품국 (European Medicines Agency; EMEA)의 임상통계관련 가이드라인을 번역하여 국내 임상시험의 질적 향상을 높이는데 도움이 되고자 본 해설서를 작성하게 되었습니다. 그 첫 번째로 임상시험 설계 및 분석에서 기저 공변량의 처리 방법에 대한 유럽 의약품국의 “Points to consider on adjustment for baseline covariates” 가이드라인 해설서를 작성하였습니다.

본 해설서가 과학적이고 객관적인 임상시험의 수행에 도움이 되어, 국내 임상시험 및 신약개발에 있어 국제 경쟁력을 확보하는데 기여할 수 있기를 희망합니다.

2009년 6월

식품의약품안전청 기획조정관

왕 진 호

기저 공변량의 보정에 대하여 고려할 점

(Points to consider on adjustment for baseline covariates)

2003.5.22

CPMP/EWP/2863/99

목 차

1. 설계 시 고려사항	2
1.1 총화	2
1.2 다기관 임상시험	2
1.3 동적 할당	3
2. 일차분석에 공변량을 포함하거나 제외하는 기준	3
2.1 일차 유효성 평가변수와의 연관성	3
2.2 총화	4
2.3 다기관 임상시험	4
2.4 동적 할당	4
2.5 기저값 불균형의 사후 보정	4
2.6 처리 할당에 의해 영향을 받는 공변량	5
2.7 기저값으로부터의 변화량 분석	5
3. 일차분석의 기술	5
3.1 일반적인 고려사항	5
3.2 분석에서 공변량의 개수	6
3.3 공변량과 일차 유효성 평가변수간의 관계	6
3.4 처리와 공변량간의 교호작용	7
4. 결과보고	7
4.1 일반적인 고려사항	7
4.2 기저값 비교	8
4.3 처리와 공변량간의 교호작용	8
4.4 모형 가정의 타당성	8
4.5 민감도 분석	9
5. 결론 및 권고사항	9

임상시험에서의 통계학적 원리를 다루는 ICH E9 지침서에서는 공변량의 보정문제를 간략하게 언급하고 있다. 이 지침서에는 ‘일차 유효성 평가변수에 중요한 영향을 미칠 것으로 예상되는 공변량을 찾는 것’과 ‘분석의 정확성을 높이고 군간 불균형을 해결하기 위해 공변량을 어떻게 처리할지를 구체화 하는 것’에 대한 권고사항이 포함되어 있다. 무작위배정 이후에 측정된 공변량들은 처리에 의해 영향을 받을 수 있으므로 보정 시 주의해야 한다고 언급하고 있다.

기저 공변량 (baseline covariates)이란 피험자가 시험약을 복용하기 전 (일반적으로 무작위배정 이전)에 관찰되거나 측정된 일차 유효성 평가변수에 영향을 미칠 것으로 예상되는 질적/양적 변수를 말한다.

기저 공변량은 다양한 형태로 존재하며 그것들은 임상연구의 상황에 따라 성격을 달리한다. 나이나 몸무게와 같은 인구학적 변수, 질병의 기간이나 중증도와 같은 질병의 특성, 일반적으로 받아들여지는 병리생리학적(pathophysiological) 이론을 근거로 하는 예후요인 또는 기관이나 연구자와 같은 요인이 기저 공변량이 될 수 있다. 일차 유효성 평가변수의 기저값 또한 일반적으로 기저 공변량으로 이용된다.

기저 공변량은 임상시험 내 두 단계에서 고려될 수 있다. 즉, 무작위배정 과정에서 설명되고/설명되거나 분석 시 보정될 수 있다.

기저 공변량을 보정하기 위한 다양한 분석방법이 존재하며, 그 방법은 공변량과 결과변수의 형태에 의해 선택된다. 보통 사용되는 방법은 분산 분석이나 공분산 분석 (일차 결과변수가 연속형인 경우), 로지스틱 회귀분석 (결과변수가 이산형이거나 범주형인 경우), Cox-회귀 분석 (생존자료 인 경우) 등이다.

본 문서에서는 공변량을 설명하는 데 있어서 다양한 모형이나 방법을 기술적이고 이론적인 관점에서 논의하지 않는다. 공변량의 특정 선택에 대한 임상적 적절성을 논의하는 것도 아니다. 본 내용의 대부분은 확증적 (보통 제 3상 임상시험) 무작위배정 임상시험의 일차 분석에 대해 다루고 있고, 무작위배정을 시행하지 않은 관찰적 연구는 다루고 있지 않다.

본 문서의 목적은 임상시험계획서나 통계분석 계획에서 기술될 일차분석에 언제, 왜 기저 공변량을 포함시켜야 하는지를 명확히 하고, 연구 결과를 어떻게 제시하고 해석해야 하는지를 명확히 하는 것이다. 흔히 하게 되는 질문은 ‘임상시험계획서 내에 보정된 분석과 보정되지 않은 분석 중 어느 것이 일차분석으로 명시되어야 하는가?’이다. 이것이 본 지침서의 중요한 논점이다.

1. 설계 시 고려사항

1.1 총화

무작위배정을 통해 처리군 간에 공변량 수준이 균형을 이룰 것으로 예상되지만, 실제로는 불균형이 일어나는 경우가 적지 않다. 만약 그러한 불균형이 시험군에 유리한 형태로 나타난다면 특별히 주의를 해야 한다. 총화 무작위배정은 특정 공변량 (질적 혹은 범주화된 양적 공변량)의 수준이 처리군 간에 불균형적으로 나타날 가능성을 줄이기 위해 종종 사용된다.

그 외 총화 설계를 하는 구체적인 이유는 다음과 같다.

- 하나 이상의 특정 예후 공변량에 대하여 처리군 간 균형을 맞추으로써 임상시험 결과의 신뢰성을 강화시킬 수 있다.
- 총화를 하면 피험자 수가 적거나 보통일 경우에도 처리효과를 더욱 효율적으로 추정할 수 있다. 무작위배정 단계에서 총화를, 분석 시에 공변량 보정을 수행하는 것은 공변량의 영향을 고려한 보완적 방법이 될 수 있다.
- 처리효과가 연령이나 인종 등과 같이 사전에 정의된 중요한 하위집단 간에 상당히 다양할 것으로 예상되는 경우, 총화는 처리효과를 해석하는 것과 처리효과를 하위집단 간에 일관성 있게 해석하는 것에 도움을 줄 수 있다. 또한 선형적으로 관심이 집중되는 특정 하위집단 분석의 신뢰성을 강화시킬 수도 있다. 그러한 교호작용이 기대된다면 임상시험은 특정 하위집단 내에서 검정력이 커져야 한다.
- 총화는 때때로 관리상의 용이성 때문에 사용될 수도 있다.

임상시험에 영향력 있는 공변량이 많은 경우에는 총화로 감당하기가 어려워진다. 특히 피험자 수가 적은 임상시험에서 많은 수의 공변량으로 총화를 하는 것은 각 층 내의 피험자 수가 적기 때문에 그 활용성이 없어지게 된다. 피험자 수가 많은 임상시험에서도 이론적으로는 많은 요인에 의해 층을 나누는 것이 가능하겠지만, 요인의 수는 임상적으로 아주 중요하거나 강력한 예후를 가진 공변량으로 제한하여야 한다.

1.2 다기관 임상시험

대부분의 다기관 임상시험의 기관 (혹은 연구자)은 알려지거나 알려지지 않은 예후 인자와 교락 (confound)되어 있을 것으로 예상되기 때문에, 또는 현실적인 이유로 인하여 기관을 총화변수로 사용한다. 만약 다기관 임상시험에서 기관이 총화변수로 사용되지 않는다면,

왜 다른 공변량이 기관보다 더 중요하게 고려되었는지에 대하여 임상시험계획서에 설명되고 정당화되어야 한다.

각 기관 내 피험자의 수가 매우 적다고 기대되는 경우에는 기관에 의한 층화 무작위배정을 하지 않을 수도 있다. 이러한 경우 무작위배정이 국가나 지역과 같은 것으로 층화될 수 있는지를 고려해야 한다. 병용요법 (co-medication), 대증요법 (palliative care) 또는 층화를 적절하게 하는 다른 요인들이 비슷한지가 층화 선택의 기준이 될 수 있다. 그러한 층화를 선택한 사유와 타당성은 임상시험계획서에 기술되어야 한다.

1.3 동적 할당

앞에서 언급한 것과 같이 피험자 수가 적은 임상시험에서 여러 예후 인자로 층화를 하는 것이 항상 가능한 것은 아니다. 이러한 경우에는 최소화 방법과 같은 동적 할당 기법이 여러 요인의 균형을 동시에 맞추기 위해 사용되기도 한다. 고정된 구조를 피한다 하더라도 동적 할당은 논쟁의 여지가 많이 남아있다. 그래서 이 방법의 사용은 권고되지 않는다. 만약 동적 할당을 사용하고자 한다면 임상적, 통계적 측면에서 그 이유가 정당화되어야 한다.

2. 일차분석에 공변량을 포함하거나 제외하는 기준

2.1 일차 유효성 평가변수와의 연관성

임상시험의 자료 분석에 공변량을 포함시키는 주요 이유는 일차 유효성 평가변수와 공변량 간에 상당히 많은 연관성이 있기 때문이다. 그러한 공변량의 보정은 일반적으로 분석의 효율성을 향상시켜, 처리효과에 대한 보다 강하고 정확한 증거 (예를 들어 작은 p-값과 좁은 신뢰구간)를 제공한다. 그러나 p-값이 작다는 것만으로는 임상적으로 유의한 효과가 있다고 확신할 수 없으며, 처리효과의 크기와 그 처리효과가 공변량 각 수준에서 일관성 있게 나타나는지를 반드시 중요하게 고려해야 한다.

일차 유효성 평가변수와의 연관성이 알려져 있거나 기대되는 공변량에 대해서는 선행 근거 (선행 임상시험이나 현재 시행중인 다른 임상시험에서 수집한 데이터를 이용할 수 있음)를 바탕으로, 그리고/또는 임상적 차원에서 정당화되어야 한다. 일차분석에 공변량을 포함하는 이유는 눈가림을 해제하기 전에 임상시험계획서나 통계분석 계획에 명확히 언급되어야 한다.

2.2 총화

일차분석은 총화에 의한 무작위배정의 제약을 반영해야 한다. 따라서 총화변수는 그것이 예후변수의 역할을 하는가에 관계없이 (즉 예후변수의 역할을 하지 못하더라도) 보통 공변량으로 포함되어야 한다. 일차분석 시 총화와 보정 간에 불일치하는 공변량이 있었다면, 이 사실을 설명하고 정당화해야 한다.

2.3 다기관 임상시험

다기관 임상시험이 기관에 의해 총화된 경우, 기관이 예후인자의 역할을 하는지에 상관없이 일차분석에서 기관이 공변량으로 보정되어야 한다. 이와 유사하게 국가나 지역과 같이 다른 것이 총화요인으로 사용되었다면 일차분석에 포함되어야 한다. 각 기관에 배정되는 피험자 수가 매우 적은 경우, 기관을 총화변수로 하여 무작위배정을 하는 것은 적절하지 않을 수도 있고 오히려 보정되지 않은 분석이 적절할 수도 있다.

적은 피험자 수를 가지는 많은 기관을 보정하는 것이 가능할 수도 있지만, 이는 최선의 해결책이 아니며 분석적 문제도 발생한다. 피험자 수가 적은 기관을 무시하고 분석할 경우 처리효과의 추정치를 신뢰하기 어려울 수 있으며, p -값이 매우 크거나 혹은 매우 작게 될 수 있다. 더구나 다른 기관의 피험자 수와 비교하기 위해 피험자 수가 적은 기관들을 통합(pooling)하는 것은 과학적인 근거가 거의 없다. 또한 무작위배정 계획에는 기관이 포함되었으나 분석 시 기관을 포함하지 않았다면 그 이유와 민감도 분석, 모의실험, 기타 다른 방법을 통하여 기관의 효과가 결론에 큰 영향을 미치지 않는다는 것을 설명하고 정당화하여야 한다.

2.4 동적 할당

동적 할당은 거의 추천되지 않는다. 만약 이를 사용한다면 할당 계획에 사용된 모든 요인이 분석에서 공변량으로 포함되어야 한다. 그러나 이러한 요구사항에도 불구하고 분석이 무작위배정 계획을 적절히 반영하는가에 대한 사항은 논쟁의 여지가 있다. 허가 신청자는 일차분석에서 얻은 결론을 지지하기 위해 수행하는 민감도 분석을 설명해야 할 것이다. 적절한 지지 분석 혹은 민감도 분석이 없으면 허가를 받지 못할 가능성이 높다.

2.5 기저값 불균형의 사후 보정

기저값이 균형 잡히지 않았다는 이유만으로 기저값을 공변량으로 포함시키는 것은 적절하지 못하다. 무작위배정 임상시험에서 명백한 불균형이 선형적으로 기대되지 않는다: 무작위배정 과정을 정확히 수행했다면 관측된 어떤 불균형이든 무작위 현상임에 틀림없다.

2.6 처리 할당에 의해 영향을 받는 공변량

처리 할당으로 인하여 영향을 받을 수 있는 공변량 (예를 들어 치료기간, 순응도, 구급약 투여와 같이 무작위배정 이후 측정된 공변량)은 확증적 임상시험의 일차분석에 일반적으로 포함되지 않아야 한다. 공변량이 직접적인 원인 혹은 다른 요인과의 연관성을 통해 처리에 의한 영향을 받을 때, 보정은 처리효과를 숨기거나 과장할 수도 있다. 그러므로 처리효과를 해석하는 데 어려움이 생긴다.

2.7 기저값으로부터의 변화량 분석

연속형 변수인 경우 일반적으로 '원 결과변수' 혹은 '기저값으로부터의 변화량' 중 하나를 일차 유효성 평가변수로 선택한다. 둘 중 어느 것을 일차변수로 하더라도 기저값은 일차분석에서 공변량으로 포함되어야 한다. 기저값을 보정하지 않은 채로 기저값으로부터의 변화량을 사용하는 것은 일반적으로 적절한 공변량 보정이라고 취급되지 않는다. 기저값이 공변량으로 모형에 포함된다면 '원 자료' 혹은 '기저값으로부터의 변화량' 중 어느 것을 사용하던지 분석에서 추정된 처리효과는 동일하다. 결론적으로 적절한 보정이 이루어진다면 일차 유효성 평가변수의 선택은 단지 해석상의 문제가 된다.

3. 일차분석의 기술

3.1 일반적인 고려사항

일차분석에 포함되는 공변량은 임상시험계획서나 통계분석 계획에 미리 정의되어야 한다. 확증적 (전형적으로 제 3상) 임상시험을 수행하는 경우, 중요한 공변량은 앞서 시행된 임상시험이나 다른 이용할 수 있는 증거로부터 미리 확인되어야 한다. 그러나 만약 계획서 작성 시점과 연구 종료시점 사이에 언급된 내용의 변화가 있다면 눈가림을 해제하기 전에 임상시험 변경계획서나 통계분석 계획에 분석방법을 수정하는 것이 적절할 수도 있다. 이전에 언급된 공변량을 제외 시키거나 새로운 공변량을 포함시키는 것에 대한 정당성을 명확히 언급해야 한다. 임상적, 통계적인 정당성 모두를 고려해야 한다. 입증된 사전 정보가 부족한 경우에는 공변량이 없거나 적은 수의 공변량만을 포함하는 간단한 모형을 사용하는 것이 바람직하다. 많은 공변량을 포함하는 분석은 잘 선택된 적은 수의 공변량만을 포함하는 분석보다 항상 신뢰도가 떨어진다.

분석에 포함된 공변량의 성질과 개수는 특히 비선형 모형에서 분석결과의 해석에 영향을 줄 수 있다. 그러한 모형에서 보정된 모수와 보정되지 않은 모수는 서로 다른 해석을 가지므로 보정된 분석을 제출할 경우에는 반드시 추정된 효과 크기의 의미를 명확하고 정확하게 설명해야 한다.

일차 유효성 평가변수와 매우 연관성이 높은 공변량을 선택하는 방법 ('변수선택방법'이라 불림)은 피해야 한다. 공변량의 임상적, 통계적 적절성은 현재 자료 이외의 다른 근거로부터 평가되고 정당화되어야 한다.

어떠한 경우에는 특정한 하나의 연구에 대하여 관련 있는 모든 민감도 분석들이 임상시험계획서에 언급되는 것은 아니다. 그러나 몇몇 민감도 분석들은 일차분석으로부터의 결론이 로버스트하다는 것을 입증하기 위해 미리 계획되어야 한다. 특히 민감도 분석은 공변량에 대한 특정한 가정을 검정하기 위해 설계되어야 한다.

3.2 분석에서 공변량의 개수

일차분석에 포함되는 공변량의 개수는 적어야 한다. 공분산분석과 같은 보정 방법을 통해 많은 수의 공변량을 보정하는 것이 이론적으로 가능하다 하더라도 간단한 모형을 미리 정의하는 것이 바람직하다. 간단한 모형에 근거한 결과는 수치적으로 보다 안정적일 것이고 그 통계적 모형을 지지하는 가정이 타당한가를 입증하기가 더 쉬우며 결과의 일반화가 용이할 수도 있다.

피험자 수가 많은 임상시험이 적은 임상시험보다 더 많은 공변량을 허용할 수 있을지 몰라도, 분석에 포함될 수 있는 공변량의 최대 개수를 정의하기 위한 공식적 규칙은 없다. 공변량들은 자주 서로 연관되어 있다. 따라서 그러한 연관성에 대한 사전 지식은 임상시험 설계 단계에서 몇몇 층화변수를 제거하는데 유용한 기초가 될 수 있다. 이러한 것을 수행할 때에는 임상적 고려사항을 참작해야 한다.

통계적 모형에 포함되는 공변량의 개수 및 총 모수의 개수는 제한적이다. 많은 수준을 가진 범주형 공변량은 효율의 손실을 발생시킬 수도 있다. 범주형 공변량의 경우 범주를 통합하거나 대안이 될 수 있는 민감도 분석을 수행하는 전략이 임상시험계획서에 미리 언급되어야 한다.

3.3 공변량과 일차 유효성 평가변수간의 관계

무작위배정 임상시험의 목표는 공변량과 일차 유효성 평가변수 간의 관계를 확인하는 것이 아니라 처리효과 차이에 대한 불편 추정치를 구하는 것이다.

선형모형이 적합하다는 가정 하에서, 처리효과의 불편추정치를 결정하는 것은 공변량과 일차 유효성 평가변수의 관계에 대한 함수의 형태 (선형 혹은 이차식)와는 관계가 없다. 그러나 공변량이나 일차 유효성 평가변수의 이상값들이 결과에 심한 영향을 줄 수 있으므로 공변량이나 일차 유효성 평가변수의 이상값에 주의를 기울여야 한다. 만약 이상값이 관측될 것으로 예측된다면 적절한 로버스트 방법을 사용하여 이상값의 영향을 최소화시킬 수 있다.

일반화 선형모형이나 비선형모형 (로지스틱 회귀분석 혹은 생존분석)을 사용하는 경우, 공변량과 일차 유효성 평가변수 간의 적절한 관계 설정은 매우 중요하며, 이러한 관계를 잘못 설정하면 처리효과의 추정치가 편의될 수 있다. 그러나 공변량과 일차 유효성 평가변수 간의 관계에

대한 사전 정보가 없는 경우 (대부분의 임상시험이 해당된다)에는 간단한 모형을 사용하는 것이 바람직하다. 예를 들어 공변량이 연속형일 때 모형은 공변량과 일차변수의 선형 관계에 근거하거나 공변량을 표본 크기에 맞는 적은 수의 수준으로 나누는 범주화에 근거한다. 그러한 경우 범주를 결정하는 법칙은 미리 정의되어야 하고 확증적 결론을 입증하기 위해 수행되는 민감도 분석은 선택된 범주화에 의해 큰 영향을 받지 않는다.

사전 연구에서 공변량과 일차 유효성 평가변수의 관계에 대한 확립된 정보가 존재한다면 이 정보는 일차 모형에 포함되어야 한다. 또한 공변량과 일차 유효성 평가변수의 관계에 대한 함수적 형태가 임상시험계획서 또는 통계분석 계획에 미리 정의되고 정당화 되어야 한다. 독립 변수와 종속변수 사이의 관계에 대한 가정이 없는 비모수적 회귀 방법도 적용될 수 있다. 그러나 이 경우에도 단지 유의수준의 계산이 아닌 처리효과 크기의 적절한 추정치를 구하는 것이 여전히 중요하다.

3.4 처리와 공변량간의 교호작용

처리와 공변량간의 교호작용은 이미 ICH E9 지침서에 언급되어 있고 공변량 보정과 관련된 문제는 아니다. 처리효과가 기저 공변량에 따라 다를 수도 있다는 사실은 이러한 공변량의 보정을 고려하는지 여부와 관계없이 중요하다.

처리와 공변량 간 교호작용을 의심할만한 근거가 없다면, 일차 분석에는 처리효과와 공변량에 대한 주 효과항(main effect term)만을 포함해야 한다. 만약 설계단계에서 교호작용이 존재한다고 의심될 경우, 층화 무작위배정이나 하위집단분석이 미리 계획되어야 한다. 임상시험은 관련된 하위집단 내 처리효과를 발견할 수 있을 정도의 적절한 검정력을 가져야 한다.

4. 결과보고

4.1 일반적인 고려사항

임상시험계획서나 통계분석 계획에 중요한 공변량이 명확히 언급되어 있고 분석이 정확히 수행되고 해석되었다면 적절한 결론을 이끌어낼 수 있다. 그러나 공변량이나 이에 대한 보정 방법이 명백하게 제시되지 않은 경우, 동일한 효과를 가지는 여러 가지 대안적인 분석을 이용하여 적절한 결론을 도출할 수도 있다. 그러나 특정한 하나의 분석이 가장 적절하다는 것에 대한 사후 논쟁은 어려울 것이다.

4.2 기저값 비교

무작위배정법과 눈가림을 완벽하게 수행한 임상시험에서 기저값 불균형에 대한 통계적 검정은 중요한 부분이 아니다. 주요 공변량에 대한 기저값 요약은 통계 검정이 처리군 간 ‘통계적 유의성’의 차이를 나타냈는지 여부와 상관없이 임상적 관점에서 제시되고 논의되어야 한다. 피험자가 무작위로 할당되지 않아 발생하는 편이는 통계적 보정으로 해결할 수 없다. 불균형의 발생 원인을 조사하는 것으로 적절한 조치 (몇몇 피험자나 기관의 제외)가 이루어질 수 있을 것이다. 그러나 그 결과는 매우 조심스럽게 해석되어야 한다.

처리그룹 간 기저 공변량의 불균형이 단지 우연히 발생했다면, 보정되지 않은 처리효과가 관측된 불균형을 설명하지 못할 때, 보정된 처리효과는 이를 설명할 수도 있다. 이러한 불균형이 시험군이 대조군보다 더 좋은 예후를 가지는 것이라면 불균형 공변량의 보정은 매우 중요하다. 또한 어떤 유의한 처리효과도 특정 공변량에 대한 기저값의 불균형에 의한 것이 아니라는 것을 입증하는 민감도 분석이 제시되어야 한다.

기저 공변량의 불균형이 매우 심한 경우에는 어떠한 보정방법도 결과의 신뢰성을 충분히 보장할 수 없다. 그러나 상당한 기저값의 불균형 (사전에 정의된 공변량이 아니더라도)을 나타내는 변수는 일차 분석의 결론에 대한 로버스트성을 평가하기 위한 민감도 분석에서 공변량으로 사용할 수도 있다.

4.3 처리와 공변량간의 교호작용

일차분석은 임상시험계획서에 미리 언급된 공변량만을 포함해야 하며, 교호작용 항은 제외해야 한다. 그러나 ICH E9 지침서에서 권고하는 바와 같이 공변량과 처리의 교호작용은 확인되어야 한다. 흔히 교호작용에 대한 검정은 검정력이 떨어지고, 교호작용이 통계적으로 유의하지 않다는 것이 임상적으로 적절한 교호작용이 없음을 의미하지는 않는다. 반대로 말하면 교호작용에 대한 통계적 유의성에만 근거하여 교호작용의 적절성에 대한 판단을 내릴 수 없다. 통계적 유의성에 근거한 교호작용의 평가는 의미가 거의 없다.

몇몇 교호작용이 임상적 관점에서 의미가 있거나, 통계적 관점에서 유의하다고 밝혀지면 처리효과가 하위집단에 따라 달리 나타날 수도 있다는 증거가 된다. 이는 주의 깊게 검토되어야 하고 (공변량의 교호작용항이 없는) 일차분석에 근거한 결론을 신중하게 해석하고 평가해야 한다. 관측된 교호작용이 크거나 질적 교호작용이라면 임상시험의 전반적인 결론을 내리는 것이 불가능해질 수도 있다.

4.4 모형 가정의 타당성

분산분석이나 공분산분석과 같은 단순한 분석의 경우, 모형 가정의 타당성이 일반적으로 충분히

뒷받침되지 않는다. 따라서 공변량이나 일차 유효성 평가변수의 극단적인 이상값들에 주의를 기울여야 하며, 그러한 이상값이 관측될 경우 결과의 로버스트성을 평가하는 대안적인 방법을 사용하여야 한다.

일반화 선형모형이나 비선형모형의 분석에서 모형을 잘못 선택할 경우, 처리효과의 추정치 부정확할 수 있다. 그러므로 모형에 대한 가정을 신중하게 확인하여 그 결과를 마지막 연구 보고서에 제시하여야 한다. 모형에 대한 가정이 맞지 않을 경우, (임상시험계획서에 미리 정의된) 대안적 분석이 명확한 통계적, 임상적 배경에서 제안되고 정당화되어야 한다.

4.5 민감도 분석

연구 결과가, 분석에 포함된 공변량의 선택 또는 공변량과 일차 유효성 평가변수 간 가정의 선택에 따라 결론이 민감하지 않음을 입증할 수 있는 대안적인 분석방법이 항상 제시되어야 한다. 민감도 분석의 결과는 보통 탐색적으로 고려되어야 하지만 반드시 일차 분석을 뒷받침하여야 한다.

선형모형에서, 보정된 처리효과의 추정치는 보정되지 않은 처리효과의 추정치와 비교되어야 한다. 두 추정치가 반드시 동일하지는 않더라도 서로 비슷할 것으로 예상된다. 일반적으로 보정된 분석이 더 효율적이라고 기대되므로 보정되지 않은 분석의 덜 유의한 결론이 반드시 문제가 되는 것은 아니다. 반대로 보정된 분석과 보정되지 않은 분석의 결론이 상반될 경우에는 이에 대해 신중하게 논의하고 해석해야 한다. 일차 분석과 민감도 분석의 결론이 임상적 관점과 통계적 유의성이라는 관점에서 크게 다를 경우, 임상시험의 결과가 결론에 도달하지 못할 수도 있다.

일반화 선형모형 혹은 비선형 모형에서 보정된 처리효과와 보정되지 않은 처리효과가 동일하게 해석되지 않을 수도 있고 때때로 서로 다른 결과가 도출될 수도 있다. 따라서 적절한 공변량의 선택과 일차 모형을 사전에 구체화하는 것이 매우 중요하다.

5. 결론 및 권고사항

단순 무작위배정을 사용한 임상시험의 분석이 타당할 경우에는 일반적으로 공변량을 보정할 필요가 없다. 그러나 중요한 공변량을 분석에 포함한다면 처리효과를 보다 효율적으로 추정할 수도 있으므로 이를 배제하게 되면 실제 처리효과를 발견하지 못하는 오류를 범할 수도 있다. 보정되지 않은 분석은 기술적으로 간단하고 그 해석이 쉬운 반면, 보정된 분석은 다소 복잡한 편이다. 그러므로 보정된 분석의 사용이 반드시 정당화되어 독자 혹은 심사자가 보정된 분석 내용을 명확히 이해할 수 있도록 제시하여야 한다. 보다 복잡한 테크닉을 사용할 때에도 이와 마찬가지로 명확함을 유지해야 할 것이다.

◇ 권고사항 :

1. 총화는 공변량에 대한 처리집단 간 균형을 보증하는데 사용될 수도 있고, 관리적인 이유에서도 사용될 수 있다. 총화에 사용된 변수는 일반적으로 일차모형에 공변량으로 포함되어야 한다.
2. 일차 유효성 평가변수와 밀접하게 혹은 어느 정도 관련이 있는 것으로 알려진 변수와, 임상적으로 연관이 있는 변수는 또한 일차 분석에서 공변량으로 고려되어야 한다. 이러한 변수들은 연구계획서나 통계분석 계획에 미리 언급되어야 한다.
3. 사후에 관측된 불균형 기저값의 변수를 일차분석의 공변량으로 포함하는 것은 타당하지 않다.
4. 무작위배정 이후 측정되어 처리효과에 의해 잠재적 영향을 받은 변수는 일차분석에서 공변량으로 포함해서는 안 된다.
5. 기저값이 연속형인 경우에는 대개 공변량으로 포함되어야 한다. 이는 일차 유효성 평가변수가 ‘원 결과’ 혹은 ‘기저값으로부터의 변화량’ 중 어느 하나로 정의되었을 때 적용된다.
6. 적은 수의 공변량만이 일차분석에 포함되어야 한다. 자료가 많은 임상시험이 더 많은 공변량들을 포함할 수 있다 하더라도 각 공변량을 포함하기 위한 정당성은 제시되어야 한다.
7. 사전정보가 없는 경우에는 연속형 공변량과 일차 유효성 평가변수 사이의 연관성을 간단한 함수 형태 (일반적으로 선형, 또는 연속형 척도의 이분화)로 가정하여야 한다.
8. 결과를 평가할 때, 모형 가정에 대한 타당성을 검증해야 한다. 특히 일반화 선형 모형이나 비선형 모형에서는 잘못된 가정으로 인해 처리효과의 부정확한 추정치가 도출될 수 있으므로 이것이 중요하다. 모형 가정에 대한 타당성은 결과를 평가할 때 확인되어야 한다. 선형모형이라 할지라도 이상값의 영향에 주의를 기울여야 한다.
9. 보정된 분석이 제시될 때마다, 모형 가정에 대한 타당성을 평가할 수 있도록 공변량 (적절하다면 범주화된)에 의해 생성되는 하위집단에서 처리효과의 결과가 제시되어야 한다.
10. 민감도 분석은 사전에 계획되어야 하고 일차 분석 결과의 로버스트성을 조사하기 위해 제시되어야 한다. 모순된 결과는 논의되고 설명되어야 한다. 논리적으로 설명될 수 없는 중요한 차이점 (예를 들어 보정된 분석 결과와 보정되지 않은 분석 결과의 상반된 결론)은 임상시험의 해석에 심각한 영향을 줄 수 있다.

11. 일차 유효성 평가변수 모형은 처리와 공변량의 교호작용을 포함하지 않아야 한다. 만약 사전에 교호작용이 기대될 경우, 임상시험은 특정 하위집단에서 처리효과의 추정치를 구하는 것이 가능하도록 설계되어야 한다.
12. 일차분석에 포함되지 않은 공변량의 이해를 돕고 신약개발중인 제약회사들을 돕기 위해 탐색적 분석이 수행될 수도 있다.
13. 임상시험계획서나 통계분석 계획에 명확히 언급된 대로 수행되고 해석된 일차 분석은 임상시험에서 얻은 결론을 뒷받침하여야 한다. 다양한 대안적 분석방법이 존재할 수 있기 때문에 미리 언급된 분석에 기초한 결과가 가장 신뢰성을 줄 것이다.

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS
(CPMP)

**POINTS TO CONSIDER ON ADJUSTMENT
FOR BASELINE COVARIATES**

London, 19 September 2002
CPMP/EWP/908/99

DISCUSSION IN THE EFFICACY WORKING PARTY	November 1999 / October 2001
TRANSMISSION TO CPMP	December 2001
RELEASE FOR CONSULTATION	December 2001
DEADLINE FOR COMMENTS	March 2002
DISCUSSION IN THE EFFICACY WORKING PARTY	September 2002 – April 2003
TRANSMISSION TO CPMP	May 2003
ADOPTION BY CPMP	May 2003

TABLE OF CONTENTS

INTRODUCTION.....	1
1. DESIGN CONSIDERATIONS.....	2
1.1. Stratification.....	2
1.2. Multicentre trials.....	3
1.3. Dynamic Allocation.....	3
2. CRITERIA FOR INCLUDING OR EXCLUDING A COVARIATE IN THE.....	3
2.1. PRIMARY ANALYSIS Association with the Primary Outcome.....	3
2.2. Stratification.....	4
2.3. Multicentre trials.....	4
2.4. Dynamic allocation.....	4
2.5. Baseline imbalance observed post hoc.....	5
2.6. Covariates affected by the treatment allocation.....	5
2.7. ‘Change from baseline’ analyses.....	5
3. SPECIFICATION OF THE PRIMARY ANALYSIS.....	6
3.1. General considerations.....	6
3.2. Number of covariates in the analysis.....	7
3.3. Relationship between covariates and the primary outcome.....	7
3.4. Treatment by covariate interaction.....	8
4. REPORT OF THE RESULTS.....	8
4.1. General considerations.....	8
4.2. Baseline comparisons.....	9
4.3. Treatment by covariate interaction.....	9
4.4. Validity of the model assumptions.....	10
4.5. Sensitivity analyses.....	10
5. OVERALL CONCLUSIONS AND RECOMMENDATIONS.....	11

INTRODUCTION

The note for guidance on statistical principles for clinical trials (ICH E9) briefly addresses the problem of adjustment for covariates. It advises experimenters ‘to identify the covariates expected to have an important influence on the primary outcome’ and to specify ‘how to account for them in the analysis in order to improve precision and to compensate for any lack of balance between groups’. It also cautions against adjusting for ‘covariates measured after randomisation because they may be affected by the treatments’.

A baseline covariate is usually defined as a qualitative factor or a quantitative variable measured or observed before a subject starts taking study medication (usually before randomisation) and expected to influence the primary variable to be analysed.

There are many types of baseline covariates and their nature depends upon the context of the study. They may be demographic variables such as age or weight, disease characteristics such as duration or severity, true prognostic factors for which there is a commonly accepted pathophysiological rationale, or factors such as centre or investigator. Quite commonly baseline values of the primary outcome are also available.

A baseline covariate can be considered at two stages in a clinical trial: it can be accounted for within the randomisation process (typically by using stratified randomisation) and/or it can be adjusted for in the analysis.

There are many different techniques for adjusting for baseline covariates, the choice of which often depends on the nature of the covariate and outcome variable. Methods commonly used are analysis of variance or analysis of covariance (when the primary outcome is quantitative), logistic regression (when the outcome is binary or categorical), and Cox-regression (for time-to-event data).

This document does not discuss the technical and theoretical aspects of the various models nor methods for accounting for covariates. Nor does it discuss the clinical relevance of particular choices of covariates. Its content is mostly concerned with the primary analysis of confirmatory (typically phase III) randomised trials. Non-randomised trials, such as observational studies, are outside its scope.

The document aims to clarify when and why baseline covariates should be included in the primary analysis that will be specified in the protocol or in the statistical analysis plan, and how the results in the study report should be presented and interpreted. A question that is often encountered is whether the adjusted or unadjusted analysis should

be declared as primary in the protocol. This guidance document addresses that critical issue.

1. DESIGN CONSIDERATIONS

1.1. Stratification

Randomisation is expected to balance treatment groups among the covariate levels but, in practice, it is not unusual to observe imbalances post hoc. Such imbalances are of particular concern if they favour the experimental group. Stratified randomisation is often used to reduce the likelihood of such imbalances between treatment groups within the levels of specified covariates (generally qualitative covariates or categorised quantitative covariates).

Additional reasons why stratified designs are used include:

- ☐ Balance of treatment groups with respect to one or more specific prognostic covariates can enhance the credibility of the results of the trial.
- ☐ Stratification might improve the efficiency of the estimate of the treatment effect, especially for small or even moderate sized trials. Stratification at the stage of randomisation and adjustment for covariates in the analysis may be seen as complementary methods of accounting for covariates.
- ☐ If the effect of treatment is expected to vary substantially across important pre-specified subgroups (for example, age groups or race), then stratifying for these subgroups can help in interpreting the treatment effect and its consistency across these subgroups. This can also enhance the credibility of some subgroup analyses that are a priori of high interest. If such an interaction is expected then the trial should be powered to the treatment effect within specific subgroups.
- ☐ Stratification may sometimes be used for reasons of administrative convenience.

Stratification can become overwhelming if there are many influential covariates in the trial. This is particularly true for small trials where stratification on more than a few covariates is often not feasible due to small sample sizes within strata. Even in large trials, although theoretically possible to stratify by many factors, the number of factors should be restricted to the most clinically important and/or strongly prognostic covariates.

1.2. Multicentre trials

Most multicentre trials are stratified by centre (or investigator) either for practical reasons or because centre (or investigator) is expected to be confounded with other known or unknown prognostic factors. When multicentre trials are not stratified by centre, then the reasons to consider why other covariates are more important than centre should be explained and justified in the protocol.

When the number of patients within each centre is expected to be very small, it may not be practical to stratify the randomisation by centre. In that case it should be considered whether randomisation could be stratified by, for example, country or region. Such a choice might be driven by similarities in co-medication, palliative care or other factors that might make stratification advisable. The reasons and justification for the choice should be described in the protocol.

1.3. Dynamic Allocation

As stated above, stratification for more than a few prognostic factors is not always possible, especially for small trials. In this situation, techniques of dynamic allocation such as minimisation are sometimes used to achieve balance across several factors simultaneously.

Even if deterministic schemes are avoided, such methods remain highly controversial. Thus, applicants are strongly advised to avoid such methods. If they are used, the reasons should be justified on solid clinical and statistical grounds. (See also section II.4.)

2. CRITERIA FOR INCLUDING OR EXCLUDING A COVARIATE IN THE

2.1. PRIMARY ANALYSIS Association with the Primary Outcome

The main reason to include a covariate in the analysis of a trial is the existence of strong or moderate association between the covariate and the primary outcome measure. Adjustment for such covariates generally improves the efficiency of the analysis and hence produces stronger and more precise evidence (smaller P-values and narrower confidence intervals) of an effect. However, it should be emphasised that simply producing smaller P-values may not be sufficient to produce convincing evidence of a clinically useful effect: the size of the treatment effect and its consistency across

levels of covariates will always be important considerations.

Known or expected associations with the primary outcome variable should be justified on the basis of previous evidence (possibly data from previous or other current trials) and/or on clinical grounds. The reasons to include a covariate in the primary analysis should be explicitly stated in the protocol or in the statistical analysis plan before breaking the blind.

2.2. Stratification

The primary analysis should reflect the restriction on the randomisation implied by the stratification. For this reason, stratification variables – regardless of their prognostic value – should usually be included as covariates in the primary analysis. Any mismatch of covariates between stratification and adjustment in the primary analysis must be explained and justified.

2.3. Multicentre trials

When multicentre trials are stratified by centre, then centre should be adjusted for in the primary analysis regardless of its prognostic value. Similarly, if an alternative feature such as region or country is used as a stratification factor, then this should be adjusted for in the primary analysis. If the number of patients likely to be included at each centre (or region, etc) is very small then stratifying the randomisation in this way may not be appropriate and an unadjusted analysis may be justified.

Adjusting for many small centres might be possible but raises analytical problems for which there is no best solution. Analyses ignoring centres used in the randomisation or adjusting for a large number of small centres might lead to unreliable estimates of the treatment effect and P-values that may be either too large or too small. Furthermore, pooling small centres to form one centre of size comparable to that of other centres has little or no scientific justification. If an applicant chooses not to include centre in the analysis when it was included in the randomisation scheme, they should explain why and demonstrate through well explained and justified sensitivity analyses, simulations, or other methods that the trial conclusions are not substantially affected because of this.

2.4. Dynamic allocation

Dynamic allocation is strongly discouraged. However, if it is used, then it is imperative that all factors used in the allocation scheme be included as covariates in the analysis.

Even with this requirement, it remains controversial whether the analysis adequately reflects the randomisation scheme. Applicants will be required to describe the sensitivity analyses they intend to perform to support the conclusions from the primary analysis. Without adequate and appropriate supporting/sensitivity analyses, an application is unlikely to be successful.

2.5. Baseline imbalance observed post hoc

Baseline imbalance in itself should not be considered an appropriate reason to include a baseline measure as a covariate. A pronounced baseline imbalance is not expected a priori in a randomised trial: if the randomisation process has worked correctly, any observed imbalance must always be a random phenomenon.

2.6. Covariates affected by the treatment allocation

A covariate that may be affected by the treatment allocation (for example, a covariate measured after randomisation such as duration of treatment, level of compliance or use of rescue medication) should not normally be included in the primary analysis of a confirmatory trial. When a covariate is affected by the treatment either through direct causation or through association with another factor, the adjustment may hide or exaggerate the treatment effect. It therefore makes the treatment effect difficult to interpret.

2.7. 'Change from baseline' analyses

When the analysis is based on a continuous outcome there is commonly the choice of whether to use the raw outcome variable or the change from baseline as the primary endpoint.

Whichever of these endpoints is chosen, the baseline value should be included as a covariate in the primary analysis. The use of change from baseline without adjusting for baseline does not generally constitute an appropriate covariate adjustment. Note that when the baseline is included as a covariate in the model, the estimated treatment effects are identical for both 'change from baseline' and the 'raw outcome' analysis. Consequently if the appropriate adjustment is done, then the choice of endpoint becomes solely an issue of interpretability.

3. SPECIFICATION OF THE PRIMARY ANALYSIS

3.1. General considerations

Covariates to be included in the primary analysis must be pre-specified in the protocol or in the statistical analysis plan. When a confirmatory (typically phase III) trial starts, the important covariates should have already been identified through previous trials and other available evidence. However, if the state of knowledge changes between the writing of the protocol and the completion of the study it may be appropriate to re-consider and update the description of the analysis in a protocol amendment or in the statistical analysis plan prior to unblinding. The justification (at this time) for including new covariates (or excluding others that were previously identified) should be stated unambiguously. Both clinical and statistical justifications should be considered. When there is a lack of such established prior knowledge, it is safer to use a simple model with no, or only a few, covariates. In all cases, analyses including many covariates will always be less convincing than analyses with fewer, well-chosen, covariates.

The nature and the number of covariates included in the analysis may affect the interpretation of the analysis, especially in non-linear models. In such models the adjusted parameters and unadjusted parameters have different interpretations: it is essential that in any presentation of adjusted analyses, the applicant clearly and precisely explains the meaning of the estimated effect size.

Methods that select covariates by choosing those that are most strongly associated with the primary outcome (often called 'variable selection methods') should be avoided. The clinical and statistical relevance of a covariate should be assessed and justified from a source other than the current dataset.

In some cases, not all of the relevant sensitivity analyses for a particular study can be anticipated in the protocol. However some sensitivity analyses should be pre-planned to establish whether the conclusions drawn from the primary analysis are robust. In particular, sensitivity analyses should be designed to test specific assumptions about covariates.

3.2. Number of covariates in the analysis

No more than a few covariates should be included in the primary analysis. Even though methods of adjustment, such as analysis of covariance, can theoretically adjust for a large number of covariates it is safer to pre-specify a simple model. Results based on such a model are more likely to be numerically stable, the assumptions underpinning the statistical model are easier to validate and generalisability of the results may be improved.

There is no formal rule for specifying the maximum number of covariates that can be included in any analysis, although larger trials might tolerate more covariates than smaller trials. Potential covariates are often strongly correlated and so knowledge of the correlation can be a useful basis for eliminating some stratification variables at the planning stage. Clinical considerations should be taken into account when doing this.

Limitations should be placed on the number of covariates included in the statistical model and on the total number of parameters. Categorical covariates with many levels may lead to a loss of efficiency. For such covariates, strategies to combine categories or to carry out alternative sensitivity analyses should be pre-specified in the protocol.

3.3. Relationship between covariates and the primary outcome

The aim of a randomised clinical trial is not to determine the true relationship between covariates and the primary outcome variable but to provide an unbiased estimate of the true difference between the treatments.

Under the assumption that a standard linear model holds, the determination of an unbiased estimate of the treatment effect does not depend upon using the correct functional form (such as linear or quadratic) to relate the covariates to the primary outcome. However, attention should be paid to outlying values of either the covariates or the outcome variable as these may have undue influence on the results. If the possibility of outlying values is foreseen, then their influence can be minimised by using suitable robust methods.

If the analysis is a generalised linear or non-linear model (such as logistic regression or survival analysis), the issue of an appropriate relationship between the covariates and the outcome is more crucial and mis-specification of this can lead to biased estimates of the treatment effect. However, in the absence of any well-established prior knowledge about the relationship between the covariates and the outcome (which is

often the case in most clinical trials) the model should use a simple form. For example, when the covariate is continuous, then the model might be based on a linear relationship between the covariate and outcome, or on a categorisation of the covariate into a few levels, the number of levels depending upon the sample size. In such a case, the rules for determining how the categories will be described should be pre-specified and sensitivity analyses conducted to ensure substantive conclusions are not highly dependent on the categories selected.

If there is well-established prior information from previous studies about how the covariates are related to the outcome, then the primary model should incorporate this information. The functional form that relates the covariates to the outcome should be pre-specified and justified in the protocol or statistical analysis plan. Nonparametric regression methods may be applied which avoid assumptions about the relationship between the dependent and independent variables. However, in these cases, it is important that appropriate estimates of the size of the treatment effect are still attainable, not just the calculation of significance levels.

3.4. Treatment by covariate interaction

This has already been addressed in ICH E9 and is not an issue specifically related to adjustment for covariates. The fact that the treatment effect may be different depending on the baseline value of a covariate is a matter for concern whether adjustment for this covariate is considered or not.

If there is no reason to suspect an interaction between treatment and a covariate then the primary analysis should only include the main effects for treatment and covariate. Conversely, if a substantial treatment by covariate interaction is suspected at the design stage, then stratified randomisation and/or subgroup analyses should be pre-planned accordingly.

The trial should have adequate power to detect treatment effects within relevant subgroups.

4. REPORT OF THE RESULTS

4.1. General considerations

If the key covariates were specified clearly in the protocol or in the statistical analysis plan and the analysis was correctly performed and interpreted, then appropriate

conclusions can be safely drawn. However, if the covariates and the method of adjustment for them were not specified unambiguously, then a number of alternative analyses may be equally valid. It will be difficult for the applicant to argue post hoc that a particular analysis is the most relevant.

4.2. Baseline comparisons

Statistical testing for baseline imbalance has no role in a trial where the handling of the randomisation and blinding has been fully satisfactory. Baseline summaries with respect to the main covariates should be presented and discussed from a clinical point of view, irrespective of whether a statistical test indicated a ‘statistically significant’ difference between treatment groups.

If the process of allocating patients to treatments has, in fact, not been random then any resulting bias cannot be corrected by any statistical adjustment. The appropriate actions (possibly excluding some patients or centres) will follow from investigations into the cause of the imbalance. The results should be interpreted very cautiously in such cases.

When there is some imbalance between the treatment groups in a baseline covariate that is solely due to chance then adjusted treatment effects may account for this observed imbalance when unadjusted analyses may not. If the imbalance is such that the experimental group has a better prognosis than the control group, then adjusting for the imbalance is particularly important. Sensitivity analyses should be provided to demonstrate that any observed positive treatment effect is not solely explained by imbalances at baseline in any of the covariates.

In the unlikely case of a very strong baseline imbalance, no adjustment may be sufficiently convincing to restore the reliability of the results. However, a strong baseline imbalance in a variable (not necessarily a pre-specified covariate) may also be a reason for including that variable as a covariate in a sensitivity analysis to allow assessment of the robustness of the conclusions drawn from the primary analysis.

4.3. Treatment by covariate interaction

The primary analysis should include only the covariates pre-specified in the protocol and no interaction terms. However, treatment by covariate interactions should be explored, as recommended in the ICH E9 guideline. Tests for interactions often lack statistical power and the absence of statistical evidence of an interaction is not evidence that

there is no clinically relevant interaction. Conversely, an interaction cannot be considered as relevant on the sole basis of a significant test for interaction. Assessment of interaction terms based on statistical significance tests is therefore of little value.

If some interactions turn out to be large from a clinical point of view or significant from a statistical point of view, this provides evidence that the effect of treatment may vary across subgroups. These findings should be examined carefully; conclusions based on the primary analysis (with no interaction) should be interpreted cautiously and commented on. If the observed interaction is particularly large in size or qualitative in nature, then interpretation of the overall results of the trial may become impossible.

4.4. Validity of the model assumptions

In the case of simple analysis of variance or covariance, model assumptions generally hold under quite weak conditions. Attention should be paid to outlying extreme values of either the covariates or the primary variable and if such outlying values are observed, then alternative methods should be used to assess the robustness of the conclusions.

If the analysis is a generalised linear or non-linear model, then mis-specification of the model could lead to incorrect estimates of the treatment effect. Thus, assumptions must be checked carefully and the findings presented in the final study report. If the model assumptions do not hold, alternative analyses (ideally pre-specified in the protocol) should be proposed and justified on clear statistical and clinical grounds.

4.5. Sensitivity analyses

Alternative analyses should always be presented to confirm that the conclusions of the study are not sensitive to the choice of covariates included or the choice of the relationship between covariates and outcome that has been assumed. Findings based on these sensitivity analyses should normally be considered exploratory but necessary to support the primary analysis.

For ordinary linear models, adjusted estimates of the treatment effect should be compared to unadjusted estimates. The estimates of the size of the treatment effect would be expected to be similar although not necessarily identical. Since there is generally an expected gain in efficiency with the adjusted analysis, a less significant result for an unadjusted analysis is not necessarily cause for concern. Conversely, if there are strong discrepancies between the conclusions drawn from adjusted and unadjusted analyses, these should be discussed and interpreted whenever possible. If

the conclusions from the primary analysis and the sensitivity analyses are very different in terms of clinical and statistical significance, then the results of the trial could become inconclusive.

For generalised linear models or non-linear models, adjusted and unadjusted treatment effects may not have the same interpretation and, sometimes, different results may be obtained from adjusted and unadjusted analyses. Thus, the choice of the appropriate covariates and the pre-specification of the primary model are critically important.

5. OVERALL CONCLUSIONS AND RECOMMENDATIONS

Valid analyses of trials that have used simple randomisation generally do not require adjustment for covariates. However, the inclusion of influential covariates may result in more efficient estimates of treatment effects and so their exclusion may result in the failure to detect real treatment effects, if they exist.

Whilst unadjusted analyses are technically simple and easy to interpret, adjusted analyses can be rather complex. Therefore the use of adjusted analyses should be justified and they should be presented in such a manner that the reader or the reviewer can understand them unambiguously. The use of more complex techniques should not be an excuse for the absence of clarity.

The following recommendations are made:

- ☐ Stratification may be used to ensure balance of treatments across covariates; it may also be used for administrative reasons. The factors that are the basis of stratification should normally be included as covariates in the primary model.
- ☐ Variables known a priori to be strongly, or at least moderately, associated with the primary outcome and/or variables for which there is a strong clinical rationale for such an association should also be considered as covariates in the primary analysis. The variables selected on this basis should be pre-specified in the protocol or the statistical analysis plan.
- ☐ Baseline imbalance observed post hoc should not be considered an appropriate reason for including a variable as a covariate in the primary analysis.

- ☐ Variables measured after randomisation and so potentially affected by the treatment should not normally be included as covariates in the primary analysis.
- ☐ If a baseline value of a continuous outcome measure is available, then this should usually be included as a covariate. This applies whether the primary outcome variable is defined as the 'raw outcome' or as the 'change from baseline'.
- ☐ Only a few covariates should be included in a primary analysis. Although larger data sets may support more covariates than smaller ones, justification for including each of the covariates should be provided.
- ☐ In the absence of prior knowledge, a simple functional form (usually either linearity or dichotomising a continuous scale) should be assumed for the relationship between a continuous covariate and the outcome variable.
- ☐ The validity of model assumptions must be checked when assessing the results. This is particularly important for generalised linear or non-linear models where mis-specification could lead to incorrect estimates of the treatment effect. Even under ordinary linear models, some attention should be paid to the possible influence of extreme outlying values.
- ☐ Whenever adjusted analyses are presented, results of the treatment effect in subgroups formed by the covariates (appropriately categorised, if relevant) should be presented to enable an assessment of the validity of the model assumptions.
- ☐ Sensitivity analyses should be pre-planned and presented to investigate the robustness of the primary results. Discrepancies should be discussed and explained. In the presence of important differences that cannot be logically explained – for example, between the results of adjusted and unadjusted analyses – the interpretation of the trial could be seriously affected.
- ☐ The primary model should not include treatment by covariate interactions. If substantial interactions are expected a priori, the trial should be designed to allow separate estimates of the treatment effects in specific subgroups.
- ☐ Exploratory analyses may be carried out to improve the understanding of covariates not included in the primary analysis, and to help the sponsor with the ongoing development of the drug.

- ☐ A primary analysis, unambiguously pre-specified in the protocol or statistical analysis plan, correctly carried out and interpreted, should support the conclusions which are drawn from the trial. Since there may be a number of alternative valid analyses, results based on pre-specified analyses will carry most credibility.

〈기저 공변량의 보정에 대하여 고려할 점〉
자문위원 명단

강승호 (연세대학교)

김선우 (삼성의료원)

남정모 (연세대학교)

박용규 (가톨릭대학교)

기저 공변량의 보정에 대하여 고려할 점

발 행 일 : 2009년 6월

발 행 기 관 : 식품의약품안전청 기획조정관 통상통계담당관

발 행 인 : 왕진호

편 집 위 원 장 : 남봉현

편 집 위 원 : 김은희 장정훈 김현정 김문신 이석배

연 락 처 : 식품의약품안전청 기획조정관 통상통계담당관

전 화 번 호 : 02) 380-1661, 1662

팩 스 번 호 : 02) 356-2893