

# Statistical Considerations in the Design of Biosimilar Cancer Clinical Trials

Chul Ahn<sup>1</sup> · Seung-Chun Lee<sup>2</sup>

<sup>1</sup>Department of Clinical Sciences, UT Southwestern Medical Center

<sup>2</sup>Department of Statistics, Hanshin University

(Received April 2011; accepted May 2011)

---

## Abstract

When a patent of an innovative (brand-name) small-molecule drug expires, generic copies of the innovative drug may be marketed if their therapeutic equivalence to the innovative drug has been shown. The small-molecule drugs are considered therapeutically equivalent and can be used interchangeably if two drugs are shown to be pharmaceutically equivalent with identical active substance and bioequivalent with comparable pharmacokinetics in a crossover clinical trial. However, the therapeutic equivalence paradigm cannot be applied to biosimilars since the active ingredients of biosimilars are huge molecules with complex and heterogeneous structures, and these molecules are difficult to replicate in every detail. The European Medicine Agency(EMEA) has introduced a regulatory biosimilar pathway which mandates clinical trials to show therapeutic equivalence. In this paper, we discuss statistical considerations in the design and analysis of biosimilar cancer clinical trials.

**Keywords:** Biosimilar, bioequivalence, biologics, immunogenicity, equivalence.

---

## 1. Introduction

“Biosimilar” or “Follow-on biologics” are the terms used to describe medicines that are the equivalent of generic drugs, but for biologic therapies. There is an increasing interest in biosimilars from both generic manufacturers and biopharmaceutical companies as many biological products reach the market and subsequently lose their patent protection. When the patent of an innovative (brand-name) small-molecule drug expires, generic copies of the innovative drug may be marketed if their therapeutic equivalence to the innovative drug has been shown. However, the generic approach for small-molecule drugs cannot be applied to copies of biologics due to their complexity. Complexity of many biologics is too great for them to be replicated exactly. Subtle changes in manufacturing processes, starting material and excipients may affect the efficacy and safety of biosimilars. Because it is impossible to show two biological products are identical, the term “biosimilars” was introduced in the Europe and “follow-on protein products” or “biogenerics” in the United States.

The Food and Drug Administration (2003) requires the evidence of average bioavailability through the conduct of bioavailability and bioequivalence studies for a generic product of small-molecule

---

<sup>2</sup>Corresponding author: Professor, Department of Statistics, Hanshin University, 411 Yangsan-Dong, Osan, Kyunggi-Do 447-791, Korea. E-mail: seung@hs.ac.kr

drugs. Bioequivalence studies are generally conducted by comparing the *in vivo* rate and extent of drug absorption of a test drug and an innovative drug. In a standard *in vivo* bioequivalence study design for a small-molecule drug, a conventional two-treatment, two-period, two-sequence  $2 \times 2$  randomized crossover design has been widely used. In this  $2 \times 2$  crossover design, study subjects receive a single dose of both test and innovative drugs on separate occasions through random assignment to the two possible sequences of drug administration. Current FDA bioequivalence guidance requires one pivotal bioequivalence study, as well as recommends both fed and fasted bioequivalence studies for most products (even products whose labels say there is no food effect on absorption). Generic small-molecule drugs are considered to be therapeutically equivalent to an innovative drug if pharmaceutical equivalence (identical active substances) and bioequivalence (comparable pharmacokinetics) can be demonstrated. Formal clinical efficacy and safety studies are not required for the approval of generic drugs. However, the generic approach cannot be applied to biosimilar products due to the complexity of biological products.

In this paper we focus on the design of biosimilar cancer clinical trials. We will discuss the clinical trial design considerations such as crossover design versus parallel design, bioequivalence criteria, statistical analysis methods, sample size calculation, and choice of primary and secondary endpoints for pharmacokinetic studies and clinical studies. The rest of the paper is organized as follows. Section 2 presents statistical design considerations for the biosimilar cancer clinical trials, and Section 3 presents the statistical methods for the analysis of biosimilar cancer clinical trials. Finally, we conclude with a discussion.

## 2. Design Considerations for Biosimilar Cancer Clinical Trials

### 2.1. Parallel design vs. Crossover design

Biosimilars are not generic equivalents of the innovative drugs since active ingredients of biosimilars are not identical to those of the innovative drugs. The molecules of a biological drug are much larger and more heterogeneous, and have far more complex structures than the traditional chemical drug. Proteins are generally 100–1,000 times larger than small-molecules (Schellekens, 2004).

A crossover design has been widely used for a generic copies of small-molecule drugs due to a short half-life. Crossover trials have the advantage of potentially reducing variability since each subject acts as his or her own control. Required sample size will be much less in a crossover design than a comparable parallel design because the within-subject variability is usually smaller than the between-subject variability, and within-subject responses to treatment are usually positively correlated. A crossover trial was used to demonstrate bioequivalence of biosimilar filgrastim and Amen filgrastim (Waller *et al.*, 2010a, 2010b; Lubenau *et al.*, 2009).

A potential problem of a crossover design is the chance of carryover effect. Carryover effects can cause treatment by period interactions, which means that the treatment effect is not constant over time. Thus, a washout period is required so that the effect of the earlier treatment is not influencing the efficacy and safety for the next treatment. A sufficient length of washout period is needed to eliminate the possible carryover effects in a crossover design. The elimination half-life of a drug is the time it takes for a drug to lose half of its pharmacologic activity. The rate at which drugs are eliminated from plasma is commonly expressed as the half-life of the drug, which is the time required for the concentration of the drug in the plasma to decrease to 50% of its initial value. The plasma concentration decrease to 25%, 12.5%, 6.25% and just over 3% at 2, 3, 4 and 5 half-lives,

respectively. Washout period should be at least five times the average mean half-life and should be extended longer if significant within-subject variation is observed in half-life (Patterson and Jones, 2006). The knowledge of half-lives of drugs is essential for the design of the clinical trial. A parallel design has not been widely used for bioequivalence studies of small-molecule drugs due to the inability of identifying and removing the intrasubject variability from the comparison between a test product and an innovative product.

Some biological drugs have a short half-life. For example, Waller *et al.* (2010a, 2010b) showed that the half-life of biosimilar filgrastim and Amgen filgrastim are 7.0 hours and 6.9 hours, respectively from a randomized Phase I trial. However, a crossover design is not appropriate for biological drugs with a long half-life, for example, monoclonal antibodies and pegylated proteins, or for proteins for which formation of anti-drug antibodies is likely. For example, the half-lives of biological drugs of Trastuzumab, Bevacizumab and Rituximab are about 2–12 days, 11–50 days and 40 hours–16 days, respectively. If cancer biological drugs are known to have a long elimination half-life and potential interference of immunogenicity, a parallel design should be considered as an alternative to a crossover design for biosimilar cancer clinical trials.

## 2.2. Sample size estimate for bioequivalence studies

Bioequivalence studies are required to be conducted using cancer patients for biosimilar drugs. Both pharmacokinetic and pharmacodynamic endpoints need to be generally investigated in biosimilar clinical trials. A Phase I randomized trial is undertaken to demonstrate equivalence of pharmacokinetic/pharmacodynamic characteristics of a test drug and an innovative drug.

Pharmacokinetics investigates what the body does to the drug, and assesses how much of the active constituents of the drug reaches its site of action. Since this assessment cannot be easily accomplished, the concentration of the drug that reaches the circulating bloodstream is taken as a surrogate. This concentration of the drug in the blood is referred to as its bioavailability. The test and innovative drugs are called bioequivalent if the two drugs have the same bioavailability. A Phase I trial should select pharmacodynamic markers that demonstrate the therapeutic efficacy of the drug, and should compare the pharmacodynamic effect of the test and innovative drugs in a population where the possible differences can best be observed. Pharmacokinetic studies can be combined with pharmacodynamic endpoints, where available.

The EMEA bioequivalence guideline (2010) for small-molecule drugs says that “Area under the Curve (AUC) is the most reliable reflection of the extent of absorption”. The FDA guidelines for small-molecule drugs also say that the statistical analysis should be based on the non-compartmental pharmacokinetic parameters  $AUC_{0-t}$  (Area Under the Curve from time 0 to the last measurable time point),  $AUC_{0-\infty}$  (Area Under the Curve from time 0 to  $\infty$ ) and  $C_{max}$  (maximum concentration) derived from the drug concentration-time curve. AUC is also the most reliable reflection of the extent of absorption for biosimilar drugs. Thus, AUC is widely used as a primary endpoint for bioequivalence studies. Secondary endpoints may include maximum concentration ( $C_{max}$ ), time of maximum concentration ( $T_{max}$ ), trough concentration at steady state ( $C_{trough,SS}$ ), average concentration ( $C_{av}$ ), trough concentration before any dose ( $C_{trough}$ ), minimum concentration ( $C_{min}$ ), peak to trough fluctuation (PTF) ratio, clearance at steady state ( $CL_{SS}$ ), terminal elimination rate constant ( $\lambda_z$ ), mean residence time at steady state ( $MRT_{SS}$ ), time of maximum concentration ( $t_{max}$ ) and elimination half-life ( $T_{1/2}$ ). Safety is also assessed as a secondary endpoint. Safety endpoints include immunogenicity, observed adverse events (AEs) based on the National Cancer Institute (NCI)

common toxicity criteria, clinical laboratory tests, physical examinations, and vital assessments.

There are three forms of bioequivalence: average bioequivalence (ABE), population bioequivalence (PBE) and individual bioequivalence (IBE). The FDA guidance recommends the use of an ABE criterion to compare bioavailability measures for replicate and nonreplicate bioequivalence studies of both immediate- and modified-release products. The EMEA guideline states that “To date, most bioequivalence studies are designed to evaluate average bioequivalence. Experience with population and individual bioequivalence studies is limited. Therefore, no specific recommendation is given on this matter”. In this paper, we will restrict our attention to ABE.

The bioequivalence of two formulations is used as a surrogate criterion for therapeutic equivalence, *i.e.*, equivalence of efficacy and safety of the two formulations. Hence, in order to have a reasonably high likelihood of therapeutic equivalence one should not consider an unacceptably wide range of the ratio of mean pharmacokinetic parameters as alternative hypotheses in designing a study for bioequivalence.

The EMEA indicates that a bioequivalence between two drug products can be claimed if the 90% confidence interval (CI) of the ratio of means of the primary PK parameters, such as AUC and  $C_{max}$ , is entirely within the bioequivalence limits of (80%, 125%). That is, bioequivalence is concluded if the average bioavailability of the test drug is within (80%, 125%) that of the innovative drug for the log-transformed pharmacokinetic responses such as AUC. Note that the distributions of many pharmacokinetic characteristics like AUC and  $C_{max}$  tend to be skewed and the underlying variances increase with the expected means. They follow a lognormal distribution rather than a normal distribution. Regulatory agencies such as FDA and EMEA suggest that a log-transformation be performed before data analysis for assessment of bioequivalence. Based on existing EMEA guidance documents on some biosimilars, AUC should be considered as the primary parameter. It is well known that the inclusion of 90% confidence interval in the acceptance range (80%, 125%) is equivalent to the rejection of the two one-sided null hypotheses at the 5% level (Hauschke *et al.*, 2007).

However, as a secondary endpoint such as  $C_{max}$  or  $C_{trough}$ , in certain cases, the EMEA and WHO allow a wider interval of (75%, 133%) for the ratio of average bioavailability to address any safety and efficacy concerns for patients. If a wider interval is used, it must be pre-specified in the protocol.

Let the primary endpoint be the area under the curve at steady state ( $AUC_{SS}$ ) at month 6. Assume that the expected ratio of means of  $AUC_{SS}$  between the test and innovative drugs is 1.0, and the coefficient of variation is 30%. Assume that data will be analyzed in the log-scale using  $t$ -tests for differences in means, and that each  $t$ -test is made at the 5% level. When the sample size in each group is 31, a two-group parallel design will have 80% power to reject both the null hypothesis that the ratio of the  $AUC_{SS}$  means of a test drug to an innovative drug is below 80% and the null hypothesis that the ratio of the  $AUC_{SS}$  means of a test drug to an innovative drug is above 125%, that is, that the test and innovative drugs are not equivalent, in favor of the alternative hypothesis that the means of the two groups are equivalent. The sample size estimate can be computed using a commercially available sample size software such as nQuery Advisor 7.0 (Elashoff, 2007). Suppose that no more than 40% of the study subjects will dropout from the study, and at least 85% of study subjects will reach a steady state by month 6. Then, the number of patients needed for the bioequivalence study is 61 patients in each group with a total of 122 patients.

A proportion of steady-state achievement can be estimated through a literature review or simulation using the results of pharmacokinetic studies of an innovative drug. For example, Bruno *et al.* (2005)

investigated population pharmacokinetics of trastuzumab in patients with HER2+ metastatic breast cancer. They showed that a two-compartment linear pharmacokinetic model best described the data. Based on the paper by Bruno *et al.* (2005), steady-state has been considered as achieved if the half-life was  $< 44.25$  hours, which is the theoretical threshold to reach at least 90% of the actual steady-state level by cycle 8 (147 days). Simulation study shows that 69% of the subjects would be at steady-state at cycle 8 using the parameter estimates of the final population pharmacokinetic model given in Table 3 of Bruno *et al.* (2005). Since only 69% of patients are expected to reach the steady-state, the required number of patients can be computed by  $n^* = n/(1 - 0.69)$ , where  $n$  is the number of patients needed assuming that all patients reach the steady-state.

### 2.3. Sample size estimate for equivalence studies

Equivalence trials (Wellek, 2003) seek to determine whether a test drug is therapeutically similar to an innovative drug with the treatment effect being between  $-\Delta$  and  $\Delta$ , where  $\Delta$  is the preset margin of the treatment effect. It is important to set up an equivalence margin that is acceptable to the regulatory authorities.

Overall survival(OS) and progression-free survival(PFS) have been widely used as the primary endpoint in cancer clinical trials for innovative drugs. However, use of standard endpoints such as OS and PFS may require thousands of patients treated for several years, which may be unrealistic for biosimilar cancer clinical trials. Use of surrogate markers generally reduces the number of patients and shortens the duration of the trial. Thus, the surrogate endpoint such as overall response rate(ORR) is a more realistic primary endpoint in biosimilar cancer clinical trials. Waller *et al.* (2010c) and Gatzemeier *et al.* (2009) used the duration of severe neutropenia as the primary endpoint, and incidence of febrile neutropenia as the secondary endpoint in a Phase III randomized equivalence trial of biosimilar filgrastim versus Amgen filgrastim in patients with breast cancer and lung cancer. Surrogate markers need to be validated and their use as a primary endpoint needs very careful consideration. The use of surrogate marker should be discussed in advance with the regulatory authorities. Other efficacy endpoint not chosen as primary endpoint can be selected for secondary endpoints. That is, OS and PFS can be selected as secondary endpoints.

Assume that the test drug is therapeutically equivalent to the innovative drug if the difference in the overall response rate(ORR) is within the interval of  $(-15\%, 15\%)$ . Assume that both the test and innovative drugs have 50% ORR. Sample size ( $n$ ) for equivalence studies can be obtained using the following sample size formula assuming no dropout.

$$n = (1 + \lambda)p(1 - p)(z_{1-\alpha/2} + z_{1-\beta/2})^2/(\lambda d^2), \quad (2.1)$$

where  $p$  is the overall response rate,  $d$  is the margin,  $\lambda$  is the proportion of patients allocated to biosimilar drug, and  $z_{1-\alpha/2}$  is the  $100(1 - \alpha/2)$  percentile of a standard normal distribution. Let  $r$  be the rate of dropout. Then, the required sample size is  $n^* = n/(1 - r)$ . From the above example,  $n = (1 + 1) * 0.5 * 0.5 * (1.96 + 1.282)^2/(1 * 0.15^2) = 234$ . If a dropout rate of 10% is assumed, then 260 ( $= 234/0.9$ ) patients are randomized to each treatment group. Note that the sample size formula in Equation (2.1) is different from that used for superiority designs. In superiority designs,  $z_{1-\beta}$  is used instead of  $z_{1-\beta/2}$ .

A sample size can be also obtained using a commercially available software such nQuery advisor sample size software. When the sample size in each group is 231, the observed two-sided 95% confidence interval for the difference of ORR will be expected to lie between  $-15\%$  and  $15\%$  with 80%

power, which is based on 5000 simulations using the Newcombe-Wilson score method to construct the confidence interval (Newcombe, 1988). If a dropout rate of 10% is assumed, then 257 patients are randomized to each treatment group.

There is a very minimal difference in sample size estimates obtained from simulation through nQuery sample size software and the Equation (2.1) due to differences in methods used for sample size estimation.

Gatzemeier *et al.* (2009) did not formally estimate the sample size for the Phase III randomized clinical trial by statistical sample size calculations. They showed that the probability of observing at least one case of a specific adverse event is approximately 70% (91%) with the sample size of 240 patients if a true incidence rate of adverse event is 0.5% (1%).

### 3. Statistical Methods for Biosimilar Cancer Clinical Trials

#### 3.1. Statistical methods for bioequivalence studies

The EMEA bioequivalence guideline (2010) states that “The pharmacokinetic parameters under consideration should be analyzed using ANOVA. The data should be transformed prior to analysis using a logarithmic transformation. A confidence interval for the difference between formulations on the log-transformed scale is obtained from the ANOVA model. This confidence interval is then back-transformed to obtain the desired confidence interval for the ratio on the original scale. A non-parametric analysis is not acceptable.”

For both PK and PD analyses, pharmacokinetic and pharmacodynamic parameters such as AUC and  $C_{max}$  are analyzed using analysis of variance (ANOVA) tests or Student’s  $t$ -tests. The ANOVA tests or  $t$ -tests are conducted to examine if treatment effects are significantly different between treatment groups. The ratios of the means of AUC and  $C_{max}$  are calculated and presented with the 90% confidence interval. The test and innovative drugs are determined as bioequivalent when the 90% confidence intervals of the ratio of the means of the PK parameters such as AUC and  $C_{max}$  are within 80% to 125%. For the PD parameters, ANOVA tests or Student’s  $t$ -tests are conducted after log transformation of pharmacodynamic parameters such as AUC and  $C_{max}$ . The non-compartmental analysis to derive the pharmacokinetic and pharmacodynamic parameters and ANOVA tests and Student’s  $t$ -tests can be done using software such as WinNonLin or bear, an R package. If the 90% confidence interval around the geometric mean ratio of the test and innovative drugs falls within bioequivalence limits of 0.80 to 1.25 (or  $\pm 0.2231$  on the natural log scale) for both  $\log(\text{AUC})$  and  $\log(C_{max})$ , average bioequivalence (ABE) can be claimed.

EXAMPLE 3.1. In a study of Steinijans *et al.* (1989),  $2 \times 2$  crossover design was applied to 18 subjects. The primary extent characteristic  $\text{AUC}_{0-\infty}$  was obtained as shown in Table 3.1. It is assumed that the original measurements follow lognormal distributions with parameters  $(\mu_T, \sigma^2)$  and  $(\mu_R, \sigma^2)$  for the test and reference drugs, respectively. Accordingly a logarithmic transformation would be done prior to data analysis. After the transformation, the response of the  $k$ th subject in the  $j$ th period of the  $i$ th sequence is modeled as

$$Y_{ijk} = \mu + \nu_i + s_{i(k)} + p_j + F_{ij} + \epsilon_{ijk},$$

where  $\mu$  are the overall mean,  $\nu_i$  is the fixed effect of sequence  $i$ ,  $s_{i(k)}$  is the random effect of subject  $k$  in sequence  $i$ ,  $p_j$  is the fixed effects of period  $j$ ,  $F_{ij}$  is the fixed effect of the formulation

**Table 3.1.** The  $AUC_{0-\infty}$  values in the study of Steinijs *et al.* (1989).  $2 \times 2$  crossover design was employed with Sequence 1 = TR, Sequence 2 = RT

Subject number	Sequence	Period 1	Period 2	Subject number	Sequence	Period 1	Period 2
1	1	228.04	288.79	2	2	339.03	329.76
3	1	288.21	343.37	4	2	242.64	258.19
5	2	249.94	201.56	6	1	217.97	225.77
7	1	133.13	235.89	8	2	184.32	249.64
9	1	213.78	215.14	10	1	248.98	245.48
10	1	163.93	134.89	12	2	209.30	231.98
13	2	207.40	234.19	14	1	245.92	223.39
15	2	239.84	241.25	16	2	211.24	255.60
17	1	188.05	169.70	18	2	230.36	256.55

**Table 3.2.** Analysis of variance table for the log-transformed AUC

Source of variation	df	Sum of squares	Mean square	F-value	p-value
Between-Subject					
Sequence	1	0.09637	0.09637	1.3870	0.2561
Subject(Sequence)	16	1.11172	0.06948	3.7071	0.0063
Within-Subject					
Period	1	0.04467	0.04467	2.3831	0.1422
Formulation	1	0.00003	0.00003	0.0017	0.9676
Residual	16	0.29989	0.01874		

administered in period  $j$  of sequence  $i$ , and  $\epsilon_{ijk}$  is the customary random error. It is assumed that  $p_1 + p_2 = 0$  and  $F_T + F_R = 0$  where  $F_T = F_{11} = F_{22}$  and  $F_R = F_{12} = F_{21}$ . The  $s_{ik}$ 's and the  $\epsilon_{ijk}$ 's are assumed to be independent normal random variables with mean 0. The variances of  $s_{ik}$  is  $\sigma_s^2$  and the variance of  $\epsilon_{ijk}$  is  $\sigma_T^2$  and  $\sigma_R^2$  for the test and reference formulations, respectively. Here we assume  $\sigma_T^2 = \sigma_R^2 = \sigma^2$ .

The ANOVA tests for various effects are shown in Table 3.2. The statistical analysis indicates that the subject effect is significant, but other effects are not. In particular, the p-value for the formulation effect is too large to reject the hypothesis  $\mu_T = \mu_R$ . However, this is an indirect method and cannot be used for bioequivalence assessment. Both regulatory agencies require to test the hypotheses

$$H_0: \mu_T - \mu_R \geq \log(1.25) \text{ or } \mu_T - \mu_R \leq \log(0.80) \text{ versus } H_a: \log(0.80) < \mu_T - \mu_R < \log(1.25)$$

at the 5% significance level to claim average bioequivalence for AUC. The regulatory agencies also recommended to use the “two one-sided tests” (TOST) proposed by Schuirmann (1987).

Note that  $(D - (\mu_T - \mu_R))/S_D$  has a  $t$ -distribution with  $m + n - 2$  degrees of freedom where  $D = 1/2 \{(\bar{Y}_{11\bullet} - \bar{Y}_{12\bullet}) + (\bar{Y}_{22\bullet} - \bar{Y}_{21\bullet})\}$ ,  $S_D^2 = 1/2 (1/m + 1/n)$  MSE, and  $m$  and  $n$  are the number of subjects in the first and the second sequence, respectively. Thus, a 90% confidence interval can be obtained as  $(-0.07777, 0.08157)$  which is included in the equivalence range of  $(-0.2231, 0.2231)$ . Therefore, we can claim the average bioequivalence of AUC.

The TOST is indeed an size  $\alpha$  test. Note however, it does not have theoretical appeal in spite of the recommendation of the regulatory agencies. For instance, Berger and Hsu (1996) showed that the TOST is biased and suffers from a lack of power. Its power is noticeably smaller than the test procedures proposed by Berger and Hsu (1996) and Brown *et al.* (1997). Those tests are unbiased as well.

### 3.2. Statistical methods for equivalence studies

A confidence interval approach is used for evaluation of efficacy by constructing a 95% confidence interval for the difference in the ORR between the test and innovative drugs. The equivalence criteria for this analysis is the closed interval  $[-\Delta, \Delta]$  in that the confidence limits of the constructed 95% confidence interval of the difference in proportions must be entirely bound by the limits of the closed interval  $[-\Delta, \Delta]$ . Data analysis is conducted using the intent-to-treat principle for the primary endpoint. The confidence interval is constructed using a simple model without adjustment for covariates and strata used for randomization. Secondary analyses are done using an analysis of covariance (ANCOVA) method with baseline measures as covariates. Kaplan-Meier survival curves are constructed for time-to-event data. Log-rank tests are used to examine if the time-to-event such as OS and PFS is significantly different between the test and innovative drug groups. Fishers exact tests, chi-square tests, and Mantel-Haenszel tests are conducted for binary or categorical variables. Safety data is presented with descriptive analysis, along with Fishers exact test for categorical variables and  $t$ -tests for continuous variables.

## 4. Discussion

Even though the statistical methods for approval of generic small molecule drugs are well established, current statistical methods for small-molecule drug products may not be appropriate for biologic products due to the complexity of biological products. The currently available EMEA CHMP guidelines for biosimilar drugs are helpful for the design and analysis of biosimilar clinical trials. However, these guidelines cannot be definitive since no two cases are identical.

In this paper, we briefly discuss the design and analysis issues for biosimilar cancer clinical trials. Comparative immunogenicity data is mandatory for the approval of biosimilar drugs. Monitoring of safety including immunogenicity is also essential for the development of biosimilar drugs. The statistical methods need to be developed to show therapeutic equivalence in product characteristics and efficacy between a biosimilar drug and an innovative drug. Until then, biosimilar drugs will be approved on a case-by-case basis with a clinical data requirement.

## References

- Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets, *Statistical Science*, **11**, 283–319.
- Brown, L. D., Hwang, J. T. and Munk, A. (1997). An unbiased test for the bioequivalence problem, *The Annals of Statistics*, **25**, 2345–2367.
- Bruno, R., Washington, C. B., Lu, J., Lieberman, G., Banken, L. and Klein, P. (2005). Population pharmacokinetics of trastuzumab in patients with HER2+ metastatic breast cancer, *Cancer Chemotherapy and Pharmacology*, **56**, 361–369.
- Elashoff, J. D. (2007). *nQuery Advisor. Version 7.0 User Guide*, Statistical Solutions, Los Angeles.
- EMA Committee for Medicinal Products for Human Use. (2010). *Guideline on the Investigation of Bioequivalence*, CPMP/EWP/QWP/1401/98 Rev. 1. London.
- Gatzemeier, U., Ciuleanu, T., Dediu, M., Ganea-Motan, E., Lubenau, H. and Del Giglio, A. (2009). XM02, the first biosimilar G-CSF, is safe and effective in reducing the duration of severe neutropenia and incidence of febrile neutropenia in patients with small cell or non-small cell lung cancer receiving platinum-based chemotherapy, *Journal of Thoracic Oncology*, **4**, 736–740.
- Hauschke, D., Steinijias, V. and Pigeot, I. (2007). *Bioequivalence Studies in Drug Development: Methods and Applications*, John Wiley & Sons, New York.



- Lubenau, H., Bias, P., Maly, A., Siegler, K. E. and Mehlretter, K. (2009). Pharmacokinetic and pharmacodynamic profile of new biosimilar filgrastim XM02 equivalent to marketed filgrastim neupogen: Single-blind, randomized, crossover trial, *Biodrugs*, **23**, 43–51.
- Newcombe, R. G. (1988). Interval estimation for the difference between independent proportions: Comparison of eleven methods, *Statistics in Medicine*, **17**, 873–890.
- Patterson, S. and Jones, B. (2006). *Bioequivalence and Statistics in Clinical Pharmacology*, Chapman & Hall/CRC, London.
- Schellekens, H. (2004). How similar do 'biosimilars' need to be?, *Nature Biotechnology*, **22**, 1357–1359.
- Schuirman, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability, *Journal of Pharmacokinetics and Biopharmaceutics*, **15**, 657–680.
- Steinijans, V. W., Sauter, R., Jonkman, J. H., Schulz, H., Stricker, H. and Blume, H. (1989). Bioequivalence studies: Single vs multiple dose, *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **27**, 261–266.
- Waller, C. F., Bronchud, M., Mair, S. and Challand, R. (2010a). Pharmacokinetic profiles of a biosimilar filgrastim and Amgen filgrastim: Results from a randomized, phase I trial, *Annals of Hematology*, **89**, 927–933.
- Waller, C. F., Bronchud, M., Mair, S. and Challand, R. (2010b). Comparison of the pharmacodynamic profiles of a biosimilar filgrastim and Amgen filgrastim: Results from a randomized, phase I trial, *Annals of Hematology*, **89**, 971–978.
- Waller, C. F., Semiglazov, V. F., Tjulandin, S., Bentsion, D., Chan, S. and Challand, R. (2010c). A phase III randomized equivalence study of biosimilar filgrastim versus Amgen filgrastim in patients receiving myelosuppressive chemotherapy for breast cancer, *Onkologie*, **33**, 504–511.
- Wellek, S. (2003). *Testing Statistical Hypotheses of Equivalence*, Chapman & Hall/CRC, London.
- U.S. Food and Drug Administration, Center for Drug Evaluation and Research (2003). *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations*. Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville MD 20857.