Enliven Archive
Intensifying Thoughts

Research Article | www.enlivenarchive.org | Enliven: Biosimilars and Bioavailability

Statistical Method for Margin Selection in Biosimilar Clinical Studies

Jiancheng Cheng[1*], Wenting Qiu[1], Ni Gao[1], Jun Zhu[1*], and Shein-Chung Chow[2]

[1]*Shanghai Henlius Biotech, Inc., China*

[2]*Duke University School of Medicine, Durham, North Carolina, USA*

[*]**Corresponding author**: Jiancheng Cheng, Shanghai Henlius Biotech, Inc., China. E-mail: Jiancheng_Cheng@henlius.com

Jun Zhu, Shanghai Henlius Biotech, Inc., China. E-mail: Jason_Zhu@henlius.com

Received Date: 30th October 2021
Accepted Date: 15th December 2021
Published Date: 26th December 2021

Citation: Cheng J, Qiu W, Gao N, Zhu J, Chow S-C. Statistical Method for Margin Selection in Biosimilar Clinical Studies. Enliven: Biosimilars Bioavailab. 2021; 6(1): 004.

1) Abstract

A stepwise approach to provide the totality-of-the-evidence regarding the safety and efficacy of the proposed biosimilar products is recommended for the regulatory approval of biosimilar products by the United States Food and Drug Administration (FDA). The stepwise approach involves similarity assessment on (i) critical quality attributes (CQAs) that are relevant to clinical outcomes in analytical studies, (ii) extent and rate of drug absorption in pharmacokinetic and pharmacodynamics (PK/PD) studies, and (iii) safety, tolerability and efficacy primary endpoints in clinical and immunogenicity studies. Similarity assessment depends heavily on the selection of similarity margins. For analytical and PK/PD similarity assessment, similarity margins are well-established. For clinical similarity assessment, however, no explicit and universally accepted similarity margins exist in practice. In this article, we reviewed several methods provided in literature and regulatory guidance for the determination of similarity margins. Extensive simulation studies were conducted to evaluate the relative performances of these methods. A couple of case studies were presented to illustrate the proposed methods for selection of similarity margins.

2) Keywords: Stepwise approach; Totality-of-the-evidence; Meta-analysis; 95%–95% Method

3) Introduction

For the approval of biosimilar products, the United States Food and Drug Administration (FDA) recommends a stepwise approach for providing the totality-of-the-evidence regarding the safety and efficacy of the proposed biosimilar products for regulatory approval. The stepwise approach includes (i) analytical similarity assessment based on critical quality attributes (CQAs) that are relevant to clinical outcomes, (ii) pharmacokinetic and pharmacodynamics (PK/PD) assessment in terms of the extent and rate of drug absorption, and (iii) clinical and immunogenicity similarity assessment in terms of safety and efficacy parameters. For analytical similarity assessment, FDA suggested that the similarity limit (δ) or equivalence acceptance criterion (EAC) of 1.5σR (i.e., $EAC=\delta=1.5\sigma_R$) should be used, where σ_R is the population standard deviation of the reference product [1]. The commonly considered similarity margin for *in vivo* PK/PD bio-similarity testing is the one-size-fits-all criterion of 80% to 125% [2,3]. Unlike analytical and PK/PD similarity assessment, FDA published a guidance on similarity margin selection in non-inferiority clinical trials based on a meta-analysis by retaining 50%–80% treatment effect in active-control trials [4].

Despite of 50%–80% allowance on retention of treatment effect, the recommended similarity margins of FDA are often too narrow to be of practical use. In practice, no explicit and universally accepted similarity margins exist for clinical and/or immunogenicity similarity assessment. Most recently, following the guidance published by FDA in 2016, a meta-analysis in conjunction with the 95%–95% method was proposed [5]. The first 95% refers to the confidence interval (CI) of the estimated effect of control based on historical data. And the second 95% refers to CI used to test the null hypothesis in the non-inferiority design (relevant to equivalence design, the 90% CI is used to test the symmetric assumptions). In certain circumstances, a two-step method with some adjustments on the 95%–95% method was designed to define the largest acceptable margin and the clinical margin. The largest acceptable margin was defined by the effect of active control, and the clinical margin was depended on clinical judgement of maintaining size for the largest acceptable margin. Recently, a meta-analysis study utilizing the modified 95%–95% method has received FDA's approval [5].

In this article, our attention will be placed on the performance of Dr. He's method He et al. [5] when assessing bio-similarity between a proposed biosimilar (test) product and an innovative biological (reference) product. Several statistical methods for determination of similarity margin are briefly described. Extensive clinical trial simulations were conducted to evaluate relative performances of the methods under study. And A couple of case studies including (i) Oncologic Drug Advisory Committee (ODAC) for evaluation of Avastin biosimilar held on July 12–13, 2017 at FDA in Silver Spring, Maryland, USA and (ii) a recent submission concerning a biosimilar product in treating patients with metastatic Colorectal Cancer (mCRC) were presented.

4) Statistical Methods for Similarity Margin Selection

4.1) FDA's Recommendation

Without loss of generality, in this section we will focus on non-inferiority margin selection as similarity margin selection can be applied directly under the assumption of symmetry. Let T, C and P denote the new or test treatment, the active protocol agent which has been demonstrated to be superior to a placebo and the placebo, respectively. Thus, if T falls within C−M and C+M, we consider T and C are therapeutically equivalent assuming that the right side of C is improving and the left side of C is worsening. Thus, if T falls on the left-hand side of C−M, i.e., T<C−M or C−T>M, we claim that T is inferior to C or C is superior to T. On the other hand, T is considered non-inferior to C if it falls on the right side of C−M, i.e., C−M<T or C−T<M. In this case, hypotheses for testing non-inferiority between T and C can be described as follows.

H_0 : C−T> M (or C−M>T, T is inferior to C)
H_a : C−T< M (or C−M<T, T is not inferior to C).

Thus, we would reject the null hypothesis that T is inferior to C and conclude that the difference between T and C is less than a clinically meaningful non-inferiority margin (M) and hence T (test treatment) is at least as effective as (or not worsen than) C, the active control agent. If T is not inferior to C and is superior to P, then (i) T > C−M or T−C > −M and (ii) $T\text{-}P > \delta$, where $M \geq \delta$. To provide a better understanding, the relationships among T, C, P and M are illustrated in Figure 1 below.

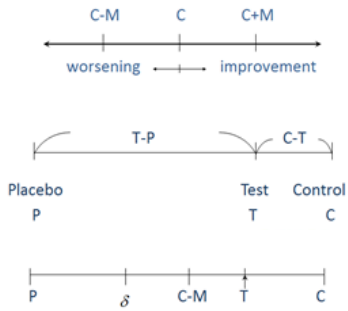

Figure 1. Relationships among T, C, P and M

Hung et al. [6] proposed the concept of retention ratio, denoted by r, which represented the effect of the test treatment (i.e., T−P) and the effect of the active control agent (i.e., C−P) as compared to a placebo control regardless the presence of the placebo in the study. That is,

$$r = \frac{T-P}{C-P}$$

where r is a fixed constant between 0 and 1. Chow and Shao [21] introduced the parameter of δ , which is the superiority margin as compared to the placebo. At the worst possible scenario, we may select $M=\delta=T-P$. In this case, the retention rate becomes

$$r = \frac{T-P}{C-P} = \frac{\delta}{C-P} = \frac{M}{C-P},$$

This lead

$$M = r(C-P).$$

Jones et al. [7] suggested that $r = 0.5$ should be chosen, while $r = 0.2$ is probably the most commonly employed for selection of non-inferiority margin without any clinical judgment or statistical reasoning. Thus, the selection of non-inferiority margin depends upon the estimation of the retention rate of the effect of the test treatment relative to the effect of the active control agent.

The 2016 FDA draft guidance recommended two non-inferiority margins, namely M_1 and M_2 should be considered. The 2016 FDA draft guidance indicated that M_1 is based on (i) the treatment effect estimated from the historical experience with the active control drug, (ii) assessment of the likelihood that the current effect of the active control is similar to the past effect (the constancy assumption), and (iii) assessment of the quality of the non-inferiority trial, particularly looking for defects that could reduce a difference between the active control and the new drug. Thus, M_1 is defined as the entire effect of the active control assumed to be present in the non-inferiority study

$$M_1 = C - P, \qquad (1)$$

On the other hand, FDA indicated that M_2 is selected based on a clinical judgment which is never greater than M_1 even for active control drugs with small effects. It should be noted that a clinical judgment might argue that a larger difference is not clinically important. Ruling out the difference between the active control and test treatment is larger than M_1 is a critical finding which supports the conclusion of effectiveness. Thus, M_2 can be obtained as

$$M_2 = (1-\delta_0)M_1 = (1-\delta_0)(C-P), \qquad (2)$$

where

$$\delta_0 = 1 - r = 1 - \frac{T-P}{C-P} = \frac{C-T}{C-P}$$

δ_0 is referred as the ratio of the effect of the active control agent as compared to the test treatment and the effect of the active control agent as compared to the placebo. Thus, δ_0 becomes smaller if the difference between C and T decreases, i.e., T is close to C (the retention rate of T is close to 1). In this case, the FDA suggests a wider margin for the non-inferiority testing.

4.2) ODAC Approved Method

On July 12–13, 2017, the FDA's ODAC unanimously recommended approval of biosimilar versions of bevacizumab (Avastin) and trastuzumab (Herceptin). In both cases, ODAC and FDA reviewers found no clinical meaningful differences between the biosimilars and the reference products. Six of Avastin's indication, including Non-Small Cell Lung Cancer (NSCLC) and mCRC were approved by ODAC members for Amgen's Avastin biosimilar candidate, ABP 215. Furthermore, a variety of retaining percentages for the CI of treatment effect were proposed by the FDA reviewers related to ABP215 biosimilar margin deriving.

To demonstrate the bio-similarity between ABP215 (Amgen) and US-licensed Avastin in patients with NSCLC receiving first-line treatment, a stepwise approach for totality of the evidence was recommended by FDA, which consisted of extensive analytical data, single-dose PK data and comparative clinical study results. In the comparative clinical study in NSCLC, a meta-analysis was conducted, and ratio of the ORR relative risk (RR) was selected by FDA to characterize the difference between ABP215 and US-licensed Avastin in the ODAC briefing document (BLA 761028).

Four randomized studies were involved in this meta-analysis: E4599 [8], JO19907 [9], AVF0757 [10], and AVAiL [11]. Even AVAiL has a different control with the other three studies, it was included in the meta-analysis because of the similar objective for difference in ORR between the experimental arm and the control arm. The meta-analysis gave a RR for pooled ORR of 0.53 with 70% CI [0.49–0.58]. For different considerations of confidence intervals from 70% to 95% with 50% maintenance of the confidence limit (CL) of the meta-analysis, it was deemed that using a 50% preserving for 70% CI [0.49–0.58] with the margin of (0.73, 1.36) was adequate to demonstrate no clinically meaningful difference between experimental product and reference product. Specifications for calculation method for the biosimilar margin will be detailed in Section 6.1.

4.3) Meta-Analysis Based Approach

In general, the meta-analysis in conjunction with the 95%–95% method (or the modified 95%–95% method) could be performed by three steps as follows.

4.3.1) Step 1:

This step is to identify the 95% CI of the estimated treatment effect of control based on historical data. A meta-analysis is usually considered.

When we perform the meta-analysis approach, usually more than one study is included. Clinical variables of intervention/exposure, control condition and participants in the beginning should be work through. And then move on to the study design, outcome and follow-up period. Results of subgroups in the trials rather than the whole population could be pooled when there is variability within different subgroups. Statistical heterogeneity should be also considered when pooling reported results. If few studies are included but has high statistical heterogeneity, result pooling should be avoided. The proportion of total variance in the pooled studies, I^2 statistic could measure the heterogeneity. The Cochrane manual Higgins J. et al. [12] informed the reference ranges for I^2 statistic: I^2 values of 0–40% represent heterogeneity which might not be important. Other measurements include Cochrane's Q and τ^2 could be considered simultaneously. Considerations for model choice based on the significance level of a heterogeneity test, for example, picking a fixed effects model (Mantel-Haenszel method) when the p-value for the test of heterogeneity is more than and equal to 0.10 and a random effects model (DerSimonian-Laird method) when $P < 0.10$. For the computation of effect size, an absolute measurement (risk different) and three relative measurements (risk ratio, odds ratio and hazard ratio) are commonly used in meta-analyses. All metrics should be considered additive on different scales, risk different (RD) is additive on an original scale, RR on a log scale, odds ratio (OR) on a logit scale and hazard ratio (HR) on a log scale. More specific information for performing meta-analysis could be referred to the previous published paper [13].

In this step, the modified 95% CIs of treatment effect, which are narrower than the normal 95% CI could be accepted in practical use. For the case of Avastin biosimilar study in NSCLC patients, 70% CI of treatment effect was chosen to support the demonstration of similarity sufficiently.

4.3.2) Step 2:

This step is to select the suitable preserving percentage for the treatment effect with 95% (or modified 95%) CI calculated from step 1. Usually, 50%~80% could be reasonable selection for the retention to derive the clinical margin.

4.3.3) Step 3:

This step is to define the largest clinical acceptable difference between the test product and active control. Calculation of largest clinical acceptable difference is based on step 1 and step 2. In non-inferiority designs, the lower bound of the confidence interval of the estimated active control effect based on historical study data is typically chose as M_1, which could be determined in step 1. The selection of M_2 is based on clinical judgment regarding how much of the active comparator treatment effect need to be preserved to demonstrate sufficiency for drug approval, which could be related to step 2. Methodologies for deriving of M_1 and M_2 could be referred to Section 4.1.

5) Clinical Trial Simulation

If meta-analysis related approach was not available due to lack of historical information, a simulation-based procedure could be considered to assist equivalence margin selection with following multiple steps.

Firstly, choose the objective margin. When both the 95%–95% method or modified 95%–95% method recommended by FDA or ODAC could not conducted in some comparative clinical study, other similarity margin could be considered, like margin *in vivo* PK/PD bio-similarity testing of 80% to 125%, similarity limit (δ) or EAC of $1.5_{\alpha R}$ for analytical similarity assessment. For example, in case study 2 of bio-similarity study which compared HLX04 to Avastin in patients with mCRC, the equivalence margin of 80% to 125% has been selected for demonstrating no clinical meaningful difference between the biosimilar candidate and the reference drug.

Secondly, perform the simulation procedure in terms of reasonable distribution and model to validate the accuracy of the alternative objective margin on the basis of the first step. Simulation procedures for oncology clinical trials could be performed based on the exponential distribution and Cox model. Characteristics of the exponential distribution, formulas of the survival time and the hazard function of Cox model are specified as the following Table 1 and Table 2.

Table 1. Characteristics of the Exponential Distribution

Characteristic	Exponential distribution
Parameter	Scale parameter $\lambda>0$
Range	$[0, \infty)$
Hazard function	$h_0(t) = \lambda$
Cumulative hazard function	$H_0(t) = \lambda t$
Inverse cumulative hazard function	$H_0^{-1}(t) = \lambda^{-1} t$
Density function	$f_0(t) = \lambda \exp(-\lambda t)$
Survival function	$S_0(t) = \exp(-\lambda t)$
Expected (Mean) survival time	$\mu = E(T) = \frac{1}{\lambda}$
Median survival time	$Med(T) = \frac{\ln(2)}{\lambda}$
Variance	$Var(T) = \frac{1}{\lambda^2}$

Table 2. Formulas of the Survival Time and the Hazard Function of Cox Model of Exponential Distribution

Characteristic	Cox-exponential model
Survival time	$T = -\frac{\log U}{\lambda \exp(\beta' x)}$
Hazard function	$h(t\|x) = \lambda \exp(\beta' x)$

U is a random variable following a uniform distribution with $U \sim U[0,1]$.

β' is the vector of regression coefficients.

The distribution of time-to-event data could be described by either a density function or a hazard function with the formulars given in Table 1. The Cox proportion hazard model is parametrized by the hazard function which could measure the correlation between covariates and time-to-event. If baseline hazard function is constant, it will be easy for the translation of the regression coefficients from hazard to survival time. In addition, Bender et al. [14] showed that exponential distribution is norm to use because it is more efficient and powerful than its alternatives even without censoring, so we considered exponential distribution for Cox model. In particular, mandatory efficacy endpoints often used Overall survival (OS) or progression free survival (PFS) with the event defined as death or documented progression of disease respectively. For example, in case study 2 we considered overall survival time as the random variable generated to simulate Cox models.

Thirdly, define characteristics of clinical trial design, like ratio of two arms, strata randomization variables, estimated parameter of distribution, accrual time and follow-up time, etc., should be pre-proposed during the simulation procedure. For example, in case study 2, some parameters of the clinical trial simulation were defined as follows:

(i) 337 patients in bevacizumab (Avastin) Group and 338 patients in HLX04 Group.
(ii) Strata randomization variables were chemotherapy regimen (XELOX or mFOLFOX6 chemotherapy), site of original tumor (left or right), ECOG scores (0 or 1).
(iii) Time to Death in bevacizumab (Avastin) Group followed exponential distribution with lambda (ln (2) / 21.7).
(iv) Time to Death in HLX04 Group followed exponential distribution with lambda (ln (2) / 20.7).

The first patient was randomized in April 2018 and the last patient was randomized in April 2019. The cut-off data was in April 2021 which was used to calculate exposure time.

Simulation was conducted 5000 times. Cox analysis was conducted for individual trial simulation. In the simulation procedure for case study 2 of HLX04-mCRC03 trial, HR = 1.025 (90% CI: 0.869, 1.209) was used. As indicated in the FDA draft guidance [1], if the true geometric mean ratio (GMR) (i.e., HR) is greater than 1/8 σ_R half of the equivalence margin, then the proposed biosimilar product cannot be demonstrated to be similar to the reference product. Thus, to ensure the demonstration of bio-similarity between HLX04 and Avastin, a simulation study was conducted by restricting the HR to Q1 (0.94) and Q3 (1.11) of the margin.

The margin of 0.8 to 1.25 for drug product with intra-subject CV been 20% to 30% and 0.7 to 1.43 for drug products with intro-subject CV been > 30% in Avastin NSCLC biosimilar studies were proposed. For the margin selection of HLX04 on mCRC patients, a simulation study was conducted with 5,000 runs been performed to study the similarity margin between 0.8 to 1.25 and 0.74 to 1.35 (A margin suggested by the 2017 ODAC meeting on Avastin Biosimilar evaluation) since the intro-subject CV of US-Avastin is greater than 30%. The simulation results indicated that even with the margin of 0.8 to 1.25 (the most stringent margin under this simulation study), we had 73.7% probability of meeting the margin and hence demonstrated the equivalence. Thus, we claim that HLX04 is similar to US-Avastin for a similarity margin within the range of 0.8 to 1.25 and 0.74 to 1.35 in this simulation study. Results for this simulation procedure are specified in Table 3.

Table 3. HLX04 Clinical Trial Simulation Results under Different Similarity Margins

USE HR 90%CI 1.025 [0.869, 1.209]		to get the Q1, Q3 (0.94, 1.11)	
		HR (HLX04/Avastin) = 1.025	
Similarity Margin	#. Simulation	#. HR (0.94, 1.11)	Success Probability of 90%CI met the margin
(0.74, 1.35) NSCLC Avastin ORR Biosimilar Margin	5000	3068	1
(0.76, 1.32)	5000	3068	0.9958
(0.78, 1.28)	5000	3068	0.8533
(0.80, 1.25)	5000	3068	0.737

6) Case Studies-Avastin (Bevacizumab) Biosimilar Studies

6.1) A Meta-Analysis based on Clinical Studies in NSCLC Patients

Based on the result of the meta-analysis conducted by FDA reviewers specified in Section 4.2, we re-calculated the margin as follows and obtained the same results.

First, the upper margin is generated from reciprocal of upper bound of 70% CI for RR and then maintaining 50% of the upper margin is computed to give the upper limit margin for the RR. Then to construct a lower margin that is symmetrical, the reciprocal of the upper margin is taken. Based on an original scale, computing margins for RR: 1/0.58 = 1.7241, the upper margin is 1/2 * (1.7241 − 1) + 1 = 1.3621 and lower margin is 1/1.3621 = 0.73. Thus, an equivalence margin of 0.73 to 1.36 will demonstrate clinical similarity for the intended indication while maintaining 50% of the 2-sided 70% CI lower bound of the effect size based on the historical ORR data.

6.2) A Biosimilar Comparative Clinical Study of HLX-04 and Avastin (Bevacizumab) in Patients with mCRC

The reference product Avastin (bevacizumab) in combination with fluoropyrimidine-based chemotherapy is indicated as the first or second line (with or without the addition of prior bevacizumab) treatment for patients with mCRC. The studies approved by FDA for those indications are based on multiple primary endpoints like OS, PFS or ORR. It was first approved by US FDA for marketing in 2004. In China, bevacizumab was approved of import registration in February 2010 and the approved indications only include mCRC and NSCLC. Most of the comparative clinical studies were conducted in NSCLC patients, and the sensitive efficacy outcome of ORR was selected for the primary endpoint.

6.2.1 Why choose the margin of 80% to 125%? To the end of deriving appropriate margins for sensitive endpoint of ORR on bevacizumab, extensively systematic review about largely randomized studies on relative indications has been conducted previously.

The methods or treatment regimens used are briefly summarized for the three reported Avastin studies (AVF2107g trial [15], ML18147 trial [16] and TREE study [17, including TREE-1 and TREE-2 trials]) in patients with mCRC. AVF2107g was a randomized, active-controlled phase III trial. Approximately 900 patients with histologically confirmed, previous untreated, bi-dimensionally measurable mCRC were enrolled and randomized to three groups: bevacizumab plus irinotecan, fluorouacil and leucovorin (IFL) (n=402) group (experimental group), bevacizumab plus fluorouacil/leucovorin (n=110) group (experimental group) and irinotecan, fluorouacil and leucovorin (IFL) plus placebo (n=411) group (active comparator group). However, the chemotherapies used in AVF2107g trial (IFL or fluorouacil/leucovorin) were quite different from which used in HLX04-mCRC03 trial (XELOX or mFOLFOX6 chemotherapy). ML18147 trial, a randomized, open-label phase III intergroup study, assigned 820 patients in a 1:1 ratio to treatment with fluorouacil (infusion or bolus) or capecitabine (oral) plus irinotecan or oxaliplatin with or without bevacizumab. Although the same chemotherapies were used in both ML18147 and HLX04-mCRC03 trials, target patients were different. Unlike the patients (previously untreated, first line) enrolled in HLX04

trial, the mCRC patients enrolled in ML18147 trial had experienced disease progression after first line treatment with standard chemotherapy and bevacizumab. Three Regimens of Eloxatin Evaluation (TREE-1) study were initiated to investigate the tolerability of oxaliplatin with the combination of three different fluoropyrimidine regimens, which were later modified with the addition of bevacizumab (TREE-2). Three chemotherapy treatment arms, mFOLFOX6, bFOL (bolus intravenous administration of FU plus low-dose leucovorin) and CapeOX (same as XELOX) regimens were independently conducted in both TREE-1 (without bevacizumab) and TREE-2 (with bevacizumab) cohorts, whereas XELOX or mFOLFOX6 chemotherapy could be separately combined with HLX04, a proposed biosimilar product or the reference biologic product (bevacizumab) in HLX04-mCRC03 intergroup trial.

In summary, both AVF2107g trial and TREE study were conducted with different interventions from HLX04-mCRC03 trial. Any inconsistency between the treatment arms will pose a risk of not achieving the objective of reasonable equivalence margins, so their results could not contribute to appropriate equivalence margin deriving for the comparative clinical study between HLX04 and bevacizumab. For the ML18147, since the target patient population to be studied were different from HLX04-mCRC03, conflicting data regarding whether prior exposure to any regimen will impact response rates of study population, as a result, results from ML18147 trial could not be used for equivalence margin deriving.

In conclusion, we considered PK equivalence margin of 80% to 125% as the objective biosimilar margin.

6.2.2 Is 80% to 125% a reasonable margin? To validate the accuracy of alternative equivalence margin of 80% to 125% for HLX04-mCRC03 trial as described earlier, a simulation study was performed to investigate the performance, properties and adequacy of statistical models for time-to- event data, regarding the proportional hazards model of Cox in pre-specified situations was suitable for HLX04 trial. Exponential distribution was applied to generate appropriate survival times for both HLX04 plus XELOX or mFOLFOX6 chemotherapy group and bevacizumab plus XELOX or mFOLFOX6 chemotherapy group. Characteristics of exponential distribution and formulas of survival time and hazard function of Cox model for exponential distribution were specified in the part of simulation methods.

To avoid error of reporting measurement for each individual study and to gain a more precise estimate of treatment effects or risk factor, a meta-analysis was performed to combine median OS of multiple studies, containing two reported studies (i) Clinical trial number: NCT01878422 (ii) NO16966 [18] and HLX04-mCRC03 trial (Clinical trial number: NCT03511963) with follow-up until one year after database locked. As the confidence interval of median OS was not reported in NO16966 trial, we estimated the 95% confidence interval for median OS of bevacizumab followed exponential distribution.

The confidence interval for median OS of the exponential distribution was derived by modifying the confidence interval for the mean of exponential distribution. The algorithm was descripted in previous publication [19].

Fixed effect model was performed in the meta-analysis based on three suitable studies for bevacizumab mentioned above by reason of $\tau^2 = 0$ and I2 = 0%, which indicated that all studies have no inconsistency, and the test of heterogeneity (Cochrane's $Q = 0.59$, $p = 0.7453$) also suggested no presence of heterogeneous results. Results for pooled median OS of 21.6571 month (≈ 21.7 month) with 95%CI [19.9681, 23.3461] was showed in Table 4.

Table 4. Meta-analysis results based on data from three randomized trials

Study	Treatments	Patients	OS (95%CI)
NCT01878422	bevacizumab + CT(FOLFIRI or FOLFOX4)	176	20.8 [15.9, 23.2]
NO16966	bevacizumab + FOLFOX4 or XELOX	699	21.3 [18.4, 24.0] *
HLX04-mCRC03	bevacizumab + mFOLFOX6 or XELOX	337	22.4 [20.1, 25.3]
Pooled results			21.6571[19.9681, 23.3461]

*: the 95% CI calculated by modifying the confidence interval for the mean of exponential distribution.

The simulation results informed that with the similarity margin of 80% to 125%, we have 73.7% probability to fall in margin limits and hence to demonstrate no clinical meaningful difference between HLX04 and Avastin. As a result, similarity margin of 80% to 125% could maintain the highly clinical judgement on comparable clinical study.

7) Discussion and Concluding Remarks

Considerations for selection of similarity (or equivalence) margins should contain clear statements on their explicit intentions, which were not only relevant to clinical meaningful differences (which usually determined based on clinical judgement by preserving a percentage of the largest acceptable margin), but also to be constrained by sample size and the power. As specified in ICH E9 [20], these margins were the largest difference that can be judged as being clinically acceptable and should be smaller than differences observed in superiority trials of the active comparator.

For case study 2, four similarity margins were considered for demonstrating no clinical meaningful differences between HLX04 and bevacizumab. The much narrower similarity margin of 80% to 125% was selected for the more stringent clinical judgment.

In addition, acceptable cost-effective assessment was considered for that we could get the largest sample size with the lowest cost to pay. In case study 1, if one desired to use a similarity margin of 0.7368 to 1.3572 with 80% power, 50% preservation of 70% confidence interval for RR (0.4902,

0.5833), a sample size of 608 was needed. Controversially if a similarity margin of 0.7732 to 1.2933 was selected with 50% preservation and 95% confidence interval for RR (0.4537, 0.6303), the larger sample size of 856 was expected to reach the 80% power. A similarity margin of 0.7368 to 1.3572 could maintain the higher efficient clinical judgment with fixed costs.

In conclusion, the key elements of considerations for margin definition were relevant to clinical meaningful difference, reasonable statistical analysis, balance between costs and effects and most importantly, to communicate with the regulatory agencies before conducting the clinical trial.

Acknowledgments

This article was supported and funded by HLX04-mCRC03 project (Clinical trial number: NCT03511963) of Shanghai Henlius Biotech, Inc. located on Shanghai, China.

References

1. FDA. Guidance for Industry – Statistical Approaches to Evaluate Analytical Similarity. United States Food and Drug Administration, Silver Spring, Maryland, USA. 2017.

2. FDA. Guidance for Industry – Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Considerations. United States Food and Drug Administration, Rockville, Maryland, USA. 2003.

3. Chow SC, Liu JP. Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition, Taylor & Francis, New York. 2008.

4. FDA. Guidance for Industry – Non-Inferiority Clinical Trials to Establish Effectiveness. United States Food and Drug Administration, Silver Spring, Maryland, USA. 2016.

5. He K, Chen H, Gwise T, Casak S, Lemery S, Keegan P, et al. Statistical considerations in evaluating a biosimilar product in an oncology clinical study. Clin Cancer Res. 2016, 22: 5167-5170.

6. James Hung HM, Wang S-J, Tsong Y, Lawrence J, O'Neill RT. Some fundamental issues for non-inferiority testing in active controlled trials. Stat Med. 2002, 22: 213-225.

7. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: The importance of rigorous methods. BMJ. 1996, 313: 36-39.

8. Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel–carboplatin alone or with bevacizumab for non–small-cell lung cancer. N Engl J Med. 2006, 355: 2542-2550.

9. Niho S, Kunitoh H, Nokihara H, Horai T, Ichinose Y, Hida T, et al. Randomized phase II study of first-line carboplatin-paclitaxel with or without bevacizumab in Japanese patients with advanced non-squamousnon-small-cell lung cancer. Lung Cancer. 2012, 76: 362-367.

10. Johnson DH, Fehrenbacher L, Novotny WF, Herbst RS, Nemunaitis JJ, Jablons DM, et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J Clin Oncol. 2004, 22: 2184-2191.

11. Reck M, von Pawel J, Zatloukal P, Ramlau R, Gorbounova V, Hirsh V, et al. Overall survival with cisplatin–gemcitabine and bevacizumab or placebo as first-line therapy for nonsquamous non-small-cell lung cancer: results from a randomised phase III trial (AVAiL). Annal Oncol. 2010, 21: 1804-1809.

12. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.2. 2021.

13. Morton SC, Murad MH, O'Connor E, Lee CS, Booth M, Vandermeer BW, et al. Quantitative synthesis – An update. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 18-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018.

14. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Stat Med. 2005, 24: 1713-1723.

15. Grothey A, Hedrick EE, Mass RD, Sarkar S, Suzuki S, Ramanathan RK, et al. Response-independent survival benefit in metastatic colorectal cancer: a comparative analysis of N9741 and AVF2107. J Clin Oncol. 2008, 26: 183-189.

16. Bennouna J, Sastre J, Arnold D, Osterlund P, Greil R, Van Cutsem E, et al. Continuation of bevacizumab after first progression in metastatic colorectal cancer (ML18147): a randomised phase 3 trial. Lancet Oncol. 2013, 14: 29-37.

17. Hochster HS, Hart LL, Ramanathan RK, Childs BH, Hainsworth JD, Cohn AL, et al. Safety and efficacy of oxaliplatin and fluoropyrimidine regimens with or without bevacizumab as first-line treatment of metastatic colorectal cancer: results of the TREE Study. J Clin Oncol. 2008, 26: 3523-3529.

18. Cassidy J, Clarke S, Diaz-Rubio E, Scheithauer W, Figer A, Wong R, et al. XELOX vs FOLFOX-4 as first-line therapy for metastatic colorectal cancer: NO16966 updated results. Br J Cancer. 2011, 105: 58-64.

19. Abu-Shawiesh, MOA. Adjusted confidence interval for the population median of the exponential distribution. J Modern Appl Stat Met. 2010, 9: 461-469.

20. ICH. Statistical Principal for Clinical Trials E9. 1998.

21. Chow SC, Shao J. On non-inferiority margin and statistical test in active control trials. Stat Med. 2006, 25: 1101-1113

22. Chow SC. Biosimilars: Design and Analysis of Follow-on Biologics. Chapman and Hall/CRC Press, Taylor & Francis, New York. 2013.