White Paper

Systematic Verification of Bioanalytical Similarity Between a Biosimilar and a Reference Biotherapeutic: Committee Recommendations for the Development and Validation of a Single Ligand-Binding Assay to Support Pharmacokinetic Assessments

Joseph C. Marini,[1,11] Michael Anderson,[2] Xiao-Yan Cai,[3] John Chappell,[4] Todd Coffey,[5] Dominique Gouty,[6] Aparna Kasinath,[7] Vera Koppenburg,[8] Philip Oldfield,[9] Shannon Rebarchak,[1] and Ronald R. Bowsher[10]

Abstract. For biosimilar drug development, it is critical to demonstrate similar physiochemical characteristics, efficacy, and safety of the biosimilar product compared to the reference product. Therefore, pharmacokinetic (PK) and immunogenicity (antidrug antibody, ADA) assays that allow for the demonstration of biosimilarity are critical. Under the auspices of the American Association of Pharmaceutical Scientists (AAPS) Ligand-Binding Assay Bioanalytical Focus Group (LBABFG), a Biosimilars Action Program Committee (APC) was formed in 2011. The goals of this Biosimilars APC were to provide a forum for in-depth discussions on issues surrounding the development and validation of PK and immunogenicity assays in support of biosimilar drug development and to make recommendations thereof. The Biosimilars APC's recommendations for the development and validation of ligand-binding assays (LBAs) to support the PK assessments for biosimilar drug development are presented here. Analytical recommendations for the development and validation of LBAs to support immunogenicity assessments will be the subject of a separate white paper.

[1] Janssen Research and Development, LLC, 1400 McKean Road, PO Box 1776Spring House, Pennsylvania 19477, USA.
[2] BDS Immunoassay Services, Tampa, Florida, USA.
[3] Merck & Company, Inc, Union, New Jersey, USA.
[4] CPR Pharma Services Pty Ltd, Thebarton, South Australia, Australia.
[5] CMC Biostatistics, LLC, Snohomish, Washington, USA.
[6] BioAgilytix Labs, Durham, North Carolina, USA.
[7] Clinigene International Limited, Bangalore, India.
[8] Sandoz Biopharmaceuticals/Hexal AG, Holzkirchen, Germany.
[9] Philip Oldfield Bioanalytical Consulting, Hudson, Quebec, Canada.
[10] B2S Consulting, Indianapolis, Indiana, USA.
[11] To whom correspondence should be addressed. (e-mail: jmarini@its.jnj.com;)

BACKGROUND

A biosimilar is a similar version of the active substance of a biological therapeutic (the reference or innovator product) that has been approved in highly regulated markets such as Europe or the USA. Based on a comprehensive comparability exercise, the biosimilar must demonstrate similarity to the reference biological therapeutic in physicochemical characteristics, pharmacokinetics, efficacy, and safety. Biosimilars are developed and approved via stringent regulatory pathways, following patent expiration or loss of exclusivity of their originator reference products.

Biosimilars are biological products; like their originator reference products, they are medicinal products produced in living cells or organisms (such as bacteria, yeast, human cell lines, animal cell lines, and plants) and encompass a wide variety of biopharmaceuticals—ranging from vaccines and gene therapy products to proteins, blood components, cells, and tissues. Types of biological products include some peptide therapeutics, monoclonal antibodies, receptor fusion proteins, and substances that are similar to key signaling proteins. Each biological product is engineered to target a specific type of molecule in the body, with the aim of treating a disease or deficiency.

Biological products are biopolymers. As such, biological products are larger in size and more complex than conventional small molecule drugs that can be synthesized chemically, and the inherent variability of the manufacturing process results in biological products which will display a certain degree of batch to batch variability. Since the large majority of biological drugs originate from living organisms (relatively small chemically synthesized peptides are the exception) and because of unavoidable differences in the manufacturing processes, the quality attributes of the biosimilar and the reference product are not expected to be entirely identical. Any differences found have to be explained and justified with regard to the impact on the safety and efficacy of the biosimilar. Therefore, biosimilar sponsors are facing a vast set of scientific challenges to demonstrate that their biological product is indeed "biosimilar" to the reference product.

Since the introduction of the first human biotechnology drug in 1982 (rDNA-derived insulin, Humulin®), great advances in biotechnology have led to over 200 biological products being approved (1). The interest in biosimilar drug development has been spurred by an increasing population of patients with an acute need for affordable high-quality biologics, but has also been complemented by a growing number of reference biological product patents due to expire in the coming years. This fact, combined with passage of the Biosimilar Price Competition and Innovation Act, a provision of the US Patient Protection and Affordable Care Act, in 2010 (2) which created an abbreviated licensure pathway for biological products in the USA, has created additional interest in the development of biosimilar therapeutics. Understandably, global drug manufacturers are focusing on the development and worldwide marketing of these biological products as an immediate goal.

Guidance documents for the development and manufacturing of biosimilars have been published by the European Medicines Agency (EMA) (3–5) and have also been drafted by the US Food and Drug Administration (FDA) (6–8). These guidances provide general principles for handling the scientific challenges of demonstrating that a proposed biological product is biosimilar to a reference product. They also provide a description of comparability protocols in support of chemistry, manufacturing, and controls (CMC); non-clinical; and clinical studies. However, despite these documents, there is still a lack of specific regulatory guidance around the requirements for the bioanalytical testing of biosimilars and reference biological products in comparability studies that support the development of biosimilars.

In response to this lack of guidance, a Biosimilars APC was formed within the American Association of Pharmaceutical Scientists (AAPS) Ligand-Binding Assay Bioanalytical Focus Group (LBABFG). The goal of this Biosimilars APC was to provide a forum to discuss the issues surrounding the assays used to quantify biosimilar and reference therapeutics within a biosimilar drug development program and to make recommendations for a systematic verification of bioanalytical similarity between a biosimilar and a reference biotherapeutic in order to introduce global alignment in biosimilar pharmacokinetic (PK) assay practices. The APC for biosimilars recognized that it was critical to demonstrate safety and efficacy comparability of the biosimilar products to their reference products for biosimilar drug development. Therefore, comparable PK and immunogenicity assays were judged to be critical for a biosimilar study. Given the complexity of biological therapeutics, the analytical methods associated with them could potentially impact the inference around similarity or dissimilarity of a biosimilar to a reference compound. Thus, it is imperative to examine all assay similarity issues carefully. Consequently, the key bioanalytical question identified by the APC was whether supporting PK assays for the reference therapeutic and the biosimilar can be conducted using a single common calibrator (*e.g.*, the biosimilar) or should both entities be employed as assay calibrators for estimating concentrations by dose interpolation.

INTRODUCTION

The purpose of this white paper is to discuss the scientific challenges associated with the development, validation, and bioanalysis of comparative bioanalytical PK assays in support of biosimilar drug development programs and to make suitable recommendations to the scientific community and regulatory authorities in an effort to harmonize global biosimilar PK assay practices (see Fig. 1). At a Hot Topic session at the 2012 AAPS National Biotechnology Conference (NBC), the Biosimilars APC presented "The Formation of a Biosimilars Action Program Committee" to introduce the goals of the APC and the gaps identified in guidance. This was followed by a roundtable presentation at the 2013 AAPS NBC by the Biosimilars APC entitled "Recommendations from the LBABFG Biosimilars APC for the Validation of PK Assays in Support of Biosimilar Drug Development." The recommendations presented here arose from productive discussions at the AAPS meetings and through numerous dynamic discussions among the diverse members of the Biosimilars APC.

A PK method (assay) can be described as a procedure for the quantification of drug concentrations in a biological matrix over time after dosing in a non-clinical or clinical study. Analytical PK methods for establishing biosimilar comparability/bioequivalence in non-clinical and clinical studies for regulatory submissions are often developed to be drug specific and are always validated for their intended use. The development phase of an assay consists of initially evaluating the feasibility of the proposed methodology followed by an extensive development of the assay in order to establish the sensitivity, specificity, and selectivity of the method (among other parameters) used to quantify the drug in the relevant matrix (9). The validation phase of an assay then verifies all of established characteristics. Well-developed methods ensure the validation's success. Successful validations pave the way for productive sample analyses in support of regulatory studies for drug development.

Typically, the bioanalytical similarity between a biosimilar and its reference product is assessed during the PK assay development phase. In this phase, the similarity of assay calibration curves and quality control (QC) samples is studied in detail in order to establish target acceptance criteria for validation. Accordingly, the term "bioanalytical similarity" is used to denote that the two biological products demonstrate an acceptable degree of comparable

What	APC consensus for a systematic and statistics based process to confirm bioanalytical similarity
Goal	Use of a single LBA method to support PK/TK assessments of both a Biosimilar and a Reference therapeutic
Who	Investigators who are conducting 'Regulatory-Compliant' bioanalysis
Why	A single LBA method is more operationally practical, reduces the variability of the data related to potential assay differences and allows the bioanalytical scientist and clinical pharmacologist to remain blinded to dosing.
When	For support of non-clinical and clinical studies

Fig. 1. Overview

bioanalytical behavior with respect to their concentration-response relationships in a ligand-binding assay (LBA). Establishment of bioanalytical similarity in the PK assay development phase is recommended for the use and validation of a single common LBA method for a reference therapeutic and a biosimilar. The validation phase then provides the formal proof that the method is capable of equally quantifying the biosimilar and the reference product. Use of comparative statistics in the assay development and validation phases establishes a mathematically derived systematic process for assessing bioanalytical similarity. The bioanalysis and in-study validation phases are initiated when the prestudy validation is complete and should be concluded within the period of time for which stability data has been generated (see Fig. 2).

A systematic evaluation that verifies bioanalytical similarity between a biosimilar and a reference biotherapeutic enables LBAs to support regulatory-compliant assessments of PK. Two important conclusions that came out of the APC discussions have guided the recommendations in this paper.

1. PK assessment of any biosimilar and its reference therapeutic would be preceded by the generation of substantial CMC data to support structural comparability. Thus, there is a high likelihood that the reference therapeutic and biosimilar will demonstrate similar analytical performance in a LBA.
2. For practical and operational reasons, the default preference is to employ a common assay calibrator to estimate serum/plasma concentrations of both the reference and the biosimilar. Thus, the overall objective becomes one of demonstrating that the concentration-response relationship for the reference biological product and biosimilar are acceptably similar to enable generation of reliable data for both protein entities.

ONE ASSAY OR TWO SEPARATE ASSAYS FOR BIOANALYTICAL SIMILARITY OF BIOSIMILAR AND REFERENCE

A well-developed, robust method should be capable of quantifying both the biosimilar and reference compounds comparably within established assay limits. The use of a different PK assay between the reference and biosimilar could add complexity to bioanalysis planning and to data interpretation (see Table I). Therefore, the recommendation is to use one PK assay for comparison of the reference and biosimilar.

During the developmental stage for the one PK assay approach, it is recommended to explore the extent of bioanalytical similarity between the biosimilar and reference product calibration curves (see "Assay Calibrator: Reference or Biosimilar" section). The decision for using either the biosimilar or reference therapeutic as the standard curve calibrator should be based on attainment of acceptable results for bioanalytical similarity. Subsequent to investigation of calibration curve similarity, accuracy and precision need to be evaluated systematically where QC samples from both the reference and the biosimilar are analyzed together with the selected calibration curve (see "Accuracy and Precision" section).

If results from the calibration curve similarity, accuracy, and precision testing performed above are satisfactory, with no outstanding concerns identified, the APC's recommendation is that the biosimilar sponsor may assess PK using a *single assay* with one calibration curve to evaluate non-clinical or clinical study samples treated with the reference or biosimilar. However, if the results from that testing are not satisfactory or a concern is identified, then the two drugs may not be considered bioanalytically similar and must be investigated. Upon investigation, if no tangible cause for the difference in the performance of the two drugs is identified, two separate assays may be used (with appropriate justification) in the interest of progressing the drug development program.

Typically, challenges associated with a two-assay scenario are in relation to the congruence of data coming from two different assays. In addition, segregation of samples between the reference and biosimilar calibration curve on assay plates may be particularly challenging when blinded PK analysis is expected. To support this, suitable protocols and SOPs with clear identification of the proposed unblinding should be written and approved in advance to sample analysis. In addition, timely sharing of assay-related information with the PK scientist and the non-clinical/clinical teams often helps in better study
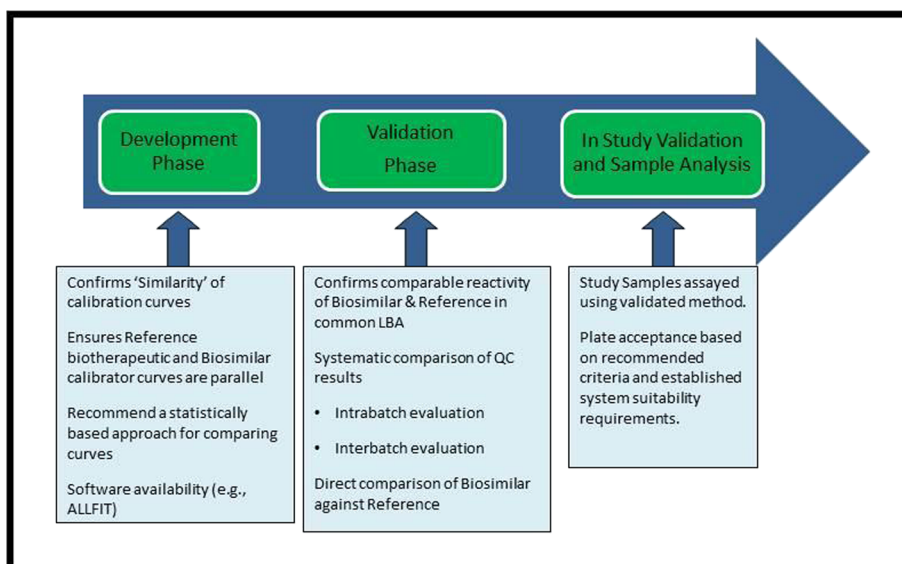
Fig. 2. Biosimilar PK assay process

designs and assists in interpretation of the PK data obtained from two different assays.

Generally, comparative clinical PK studies are either built on a cross-over model or a parallel group design depending on the drug product, its half-life, and risk of immunogenicity. The recommended flexibility in approach with one or two separate PK assays should also carefully consider the intended clinical design, and the choice should be evaluated with a clear understanding about the nature of the drug.

SECTION 1: DEVELOPMENT OF A PK ASSAY TO QUANTIFY BIOSIMILAR AND REFERENCE THERAPEUTICS IN SUPPORT OF BIOSIMILAR DRUG DEVELOPMENT

During the development phase of a PK assay used to evaluate the bioanalytical similarity between a biosimilar and its reference, specific parameters recommended by standard bioanalytical guidance documents, white papers, and other bioanalytical assay publications should be assessed (9–15).

Table I. Advantages and Disadvantages of One Assay *vs.* Two Assays

One-assay approach		Two-assay approach	
Pro	Con	Pro	Con
Conservative approach to use biosimilar curve for quantification of both biosimilar and reference drug concentrations	Need to demonstrate minimal absolute difference in %RE between biosimilar QCs and reference QCs	Concentration of biosimilar and reference drug will be calculated from curve of the same respective material, eliminating variability due to differences between curve material and QC material	Need to perform cross-comparison of QCs against both curves to establish a true comparison between QCs that would demonstrate bioanalytical similarity
No "between-assay" variability, *i.e.*, minimization of the potential impact of assay bias on the comparison of the biosimilar and the reference product			Two assays with different properties: different reagents, assay characteristics (selectivity, sensitivity, *etc.*) • Introduction of additional variability that might reduce the reliability of the comparison
Blinded study sample analysis possible			Blinded analysis would require all samples to be run through both assays
Need to develop and validate one assay			Need to develop and validate two assays

QC quality control

The assay data obtained during development should be interpreted carefully before moving into the validation phase.

The following key aspects of method development are expected to influence the outcome of the PK assay used to determine bioanalytical similarity:

1. Assay methodology/platform
2. Critical reagents
3. Assay design: calibrators and quality controls
4. Assay calibrators: reference or biosimilar

Assay Methodology/Platform

Biosimilar drug development requires non-clinical and clinical PK studies to be conducted in a head-to-head comparability design between the biosimilar product and the reference product (5–8).

The authors acknowledge that it may not be feasible or appropriate to use the same assay format, platform, and reagents used in the development of the reference compound due to the lack of historical information and unavailability of the same reagents. Further, it is possible that the assay originally used by the reference has since been updated or replaced.

Therefore, when choosing an appropriate/alternate assay platform, the inherent nature and properties of the molecule under evaluation should be considered first (*e.g.*, its specific interaction with the identified target and complex-forming properties with soluble and membrane-bound receptors in the given biological matrix). Additional parameters such as presence of endogenous versions of the drug, circulating antidrug antibodies, assay sensitivity, required dynamic range, and available sample volumes should be taken into account.

Considering the long duration of the drug development process, with comparative trials spread over several years, it is advisable to build in contingency and select a flexible, "open to development" assay platform for both drugs.

Critical Reagents

Critical reagent generation is one of the preliminary activities during development of PK assays. The nature and stability of the critical reagents used in the assay should be studied in detail during development, and well-characterized reagents should be used in validation and sample analysis studies within their stability period. Common LBA critical reagents are antibodies, receptors, target proteins, and ligands that are specific to the drug in the assay. In cases where these reagents are not available commercially, they not only require a long lead time for generation, but are often accompanied by lot to lot variability challenges. Considering tight drug development timelines, it is common for biosimilar manufacturers to begin generation of drug-specific critical reagents well in advance. These reagents may be generated by using either the reference or the biosimilar, depending upon the immediate availability of the drug. Irrespective of the drug chosen for the generation of critical reagents for an assay, it is recommended that the PK bioanalytical similarity method development program evaluate the specificity of the reagents to both the biosimilar and the reference therapeutic

since reagents that are not specific to the biosimilar and reference therapeutic may have an adverse effect on the final results.

Assay Design: General Considerations for The Preparation Of Calibrators And Quality Controls

Generally, LBAs measure the concentration of a drug in an indirect manner using heterogeneous pairing reagents. The concentration response relationship for these assays is inherently non-linear or sigmoidal (16,17).

A standard calibration curve is prepared by spiking known concentrations of the biosimilar or reference therapeutic in the relevant biological matrix. This curve should consist of at least six calibrator points and, with or without dilutions, should cover the entire estimated drug concentration range. The standard curve calibration points should be placed approximately equidistant on the logarithmic scale from each other. Anchor points (*i.e.*, points below the lower limit of quantification (LLOQ) and above the upper limit of quantification (ULOQ)) may be used for improved curve fitting in development and carried into validation and routine sample analysis as well. The drug concentration *vs*. the mean response data in a PK assay is typically fitted using a four- or five-parameter logistic (PL) model. Weighting of the curve may also be used as appropriate. During method development, it is recommended to test all possible fits of the curve (across multiple runs) with various weighting options to arrive at a response that is robust and evaluable in validation. Mean bias (%RE), total error, and %CV should be determined using the back-calculated concentrations obtained for both biosimilar and reference curves.

QC samples in PK assays are, in principle, used to simulate anticipated study samples. Consistent responses from QC samples impart an added level of confidence to the study sample data and ensure that the assay is performing to expectation. Quality control samples are prepared by spiking known concentrations of the biosimilar and reference in the related biological matrix. It is recommended to evaluate QC samples in at least five levels including low, mid, high, LLOQ, and ULOQ concentrations of both drugs separately. Ultrahigh (dilution) QCs should be evaluated as per the requirements of the study.

Assay Calibrator: Biosimilar or Reference

A critical assumption for use of a common calibrator is that the biosimilar and reference biotherapeutics have undergone detailed characterization and have previously been judged to be comparable with respect to their physicochemical and biological characterization. It should be noted that bioanalytical similarity in the PK assay does not indicate that the reference and biosimilar possess CMC similarity. Likewise, bioanalytical similarity, *i.e.*, statistical PK assay similarity (parallel calibration curves) does not imply biological similarity of the two drugs. Moreover, two therapeutics may possess complete structural similarity but exhibit a difference in their formulation concentration such that there is a subtle difference in their relative potency (*i.e.*, ratio of EC_{50} values). In such cases, establishment of analytical

similarity will be complicated as the two standards will demonstrate a systematic bias.

A PK assay is critical to support non-clinical and clinical comparability studies for biosimilar development. Therefore, it is recommended that the PK assay used for the comparability studies should evaluate the performance of both the reference and the biosimilar in a head-to-head manner. To evaluate the bioanalytical similarity of the reference and biosimilar therapeutics, a comparison of the biosimilar and reference calibration curves should be done in method development. The recommendation is to use an assay design that minimizes potential experimental bias in positional, operator, and/or environmental effects. The calibration points should be analyzed with the same set of assay reagents in the same biological matrix under the same assay conditions in at least three independent runs made by multiple operators over multiple days. At least three independent calibration curve preparations each of the biosimilar and reference should be analyzed in duplicate on the same plate (see Fig. 3). In this design, the total number of wells per calibrator concentration per drug is six. The average response for each concentration should be calculated from the independent responses of the six wells. As a best practice, an assessment of the similarity of the calibrator curves' responses should be made using the statistical methods and target acceptance criteria shown in detail in "SECTION 3: RECOMMENDED STATISTICAL ANALYSIS FOR ESTABLISHING BIOANALYTICAL SIMILARITY" section. Upon establishment of the calibration curve similarity between the reference and the biosimilar, either the reference or the biosimilar may be chosen as calibration standard for subsequent experiments. For logistical reasons, it is expected that the biosimilar will be the most likely choice.

It is reasonable to evaluate EU, US, and other globally authorized reference products under the same assay conditions during development and/or validation to compliment head-to-head trials with the drug products. However, a recently released EMA draft revised Guideline on Similar Biological Medicinal Products (18) allows the evaluation of the biosimilar drug with a non-European Economic Area (EEA)-authorized product on a case-by-case basis, with caveats for quality, safety, and efficacy studies. In this regard, it is recommended that laboratories developing and validating comparative PK assays take appropriate decisions about the reference product upon consultation with their non-clinical and clinical teams in order to support their respective trials.

SECTION 2: VALIDATION OF A PK ASSAY TO QUANTIFY BIOSIMILAR AND REFERENCE THERAPEUTICS IN SUPPORT OF BIOSIMILAR DRUG DEVELOPMENT

In general, bioanalytical method validation formally establishes the results obtained in method development and demonstrates that a particular assay is fit for its intended use (*i.e.*, reliable and reproducible quantification of the drug in the relevant matrix). Selective, sensitive, and validated analytical methods for the quantitative evaluation of drugs are critical for the successful conduct of non-clinical and/or biopharmaceutics and clinical pharmacology studies (11). While bioanalytical method validation for a comparative assay should also be based on specific requirements defined in a validation plan and should adhere to all validation recommendations made by standard guidance documents and white papers for bioanalytical assays, this paper should be read as supplementary to the information available in standard guidance for the validation of LBAs.

The recommendation is to fully validate the PK assay used to quantify biosimilar and reference therapeutics in support of biosimilar drug development. Partial or cross-validation may be applied for previously validated and established methods on a case-by-case basis. Key assumptions for recommendations in this section are that the reference and biosimilar calibration curve have demonstrated similarity in the development phase and that the calibrator curve is made from a single compound (either the biosimilar or the reference). Departure from these key assumptions should be handled on a case-by-case basis with the final goal of demonstrating that the validated method is fit for its intended purpose.

Accuracy and Precision

The accuracy of a comparative PK assay should be judged on the performance of QC samples at a minimum of five concentration levels (low, mid, high, LLOQ, and ULOQ) each made independently from the reference and the biosimilar spiked into the relevant matrix. The QC samples should be evaluated against a single calibration curve (preferably using the biosimilar reference standard) as previously determined in "SECTION 1: DEVELOPMENT OF A PK ASSAY TO QUANTIFY BIOSIMILAR AND REFERENCE THERAPEUTICS IN SUPPORT OF

Fig. 3. Balanced assay design for evaluation of reference and biosimilar comparability. The biosimilar and reference standard curves can be alternated on different assay plates to mitigate potential positional effects.

BIOSIMILAR DRUG DEVELOPMENT" section. This assessment should be made on the same plate.

Three independently prepared sets of QCs (minimum five concentration levels) from the reference, three independently prepared sets of QCs (minimum five concentration levels) from the biosimilar, and one independently prepared set of the calibration standards should be analyzed on the same plate and considered one independent run. Plate layouts should be designed efficiently for maximum utilization (an example is shown in Fig. 4).

QC samples and calibrators should be spiked independently using separately prepared starting stock solutions. However, the same dilution pattern should be followed for the preparation of the QC sets using the biosimilar or the reference. Each QC preparation should be added to the plate in duplicate. The mean value from the two wells associated with each QC should be used for reporting.

The mean of back-calculated concentrations for all wells of each standard curve point should be calculated using the nominal concentration of the biosimilar (or relevant drug). Precision of the back-calculated concentration from the duplicate wells should be calculated. Similarly, for each QC point, mean of back-calculated concentrations for all wells should be using respective nominal concentrations of the two drugs. Precision of the calculated concentration from the duplicate wells with QCs should be calculated.

It is recommended that interbatch (a single assay plate containing standard curve calibrators, QC samples, and test samples that are analyzed together) accuracy and precision assessment include at least six independent runs made by two or more operators running two plates over 3 days in a balanced fashion to provide a high degree of statistical assurance and detect potential bias due to unforeseen reasons on a particular day or operator.

Intrabatch Mean, %CV, and Mean Bias (%RE)

From the mean concentrations of each QC of the reference and biosimilar QC sets, respectively, Intrabatch mean, %CV, mean bias (%RE), and total error should be determined. This determination should be made independently for all six (or more) runs for the biosimilar and reference.

Interbatch Mean, %CV, and Mean Bias (%RE)

From the six (or more) intrabatch mean bias (%RE) data points obtained above for the reference and biosimilar QCs, respectively, interbatch %CV, interbatch mean bias (%RE), and interbatch total error should be determined.

Although some bioanalytical laboratories may consider additional experiments to establish lot-to-lot similarity for the two drugs in the developed PK assay, caution should be exercised in interpretation of this data since the results are not intended to establish biosimilarity and therefore should not be confused with typical CMC comparability studies. Nevertheless, these experiments are useful for confirming that the PK assay is able to equally quantify different lots of the biosimilar and the reference.

Target Acceptance Criteria

The recommended target acceptance criteria described below are based on the established Industry standards for bioanalytical PK method validation and are not unique to biosimilar bioanalytical similarity assessments.

For the intrabatch QC data:

As target acceptance criteria for comparative PK analysis, routine (SOP-governed) laboratory system suitability acceptance criteria should be applied to the intrabatch calibration curve and QC data for the reference and biosimilar.

For the interbatch QC data:

The following acceptance criteria should be applied for both reference and biosimilar QCs:

1. Interbatch comparability data between the reference and biosimilar should be used for comparison.
2. Each concentration of the biosimilar QC should be compared to the respective concentration of the same level reference QC.
3. The interbatch mean bias (%RE) should be within ±20% of the nominal concentration for the high, mid, and low QCs and within ±25% for the ULOQ and LLOQ QCs.
4. The interbatch %CV should be ≤20% for the high, mid, and low QCs and within ≤25% for the ULOQ and LLOQ QCs.

Plate 1	Biosimilar						Reference					
	1	2	3	4	5	6	7	8	9	10	11	12
A	Std 1	Std 2	Std 3	Std 4	Std 5	Std 6	Std 7	Std 8	Std 9	Std 10	Std 11	Std 12
B	Std 1	Std 2	Std 3	Std 4	Std 5	Std 6	Std 7	Std 8	Std 9	Std 10	Std 11	Std 12
C												
D	ULOQ-1a	ULOQ -1b	ULOQ-1a	ULOQ -1b	ULOQ-2a	ULOQ-2b	ULOQ-2a	ULOQ-2b	ULOQ-3a	ULOQ-3b	ULOQ-3a	ULOQ-3b
E	HQC-1a	HQC-1b	HQC-1a	HQC-1b	HQC-2a	HQC-2b	HQC-2a	HQC-2b	HQC-3a	HQC-3b	HQC-3a	HQC-3b
F	MQC-1a	MQC-1b	MQC-1a	MQC-1b	MQC-2a	MQC-2b	MQC-2a	MQC-2b	MQC-3a	MQC-3b	MQC-3a	MQC-3b
G	LQC-1a	LQC-1b	LQC-1a	LQC-1b	LQC-2a	LQC-2b	LQC-2a	LQC-2b	LQC-3a	LQC-3b	LQC-3a	LQC-3b
H	LLOQ-1a	LLOQ-1b	LLOQ-1a	LLOQ-1b	LLOQ-2a	LLOQ-2b	LLOQ-2a	LLOQ-2b	LLOQ-3a	LLOQ-3b	LLOQ-3a	LLOQ-3b

Fig. 4. Accuracy and precision—suggested plate layout. The biosimilar and reference standard QCs can be alternated on different assay plates to mitigate potential positional effects. The calibration curve can be prepared using the biosimilar or reference drug (see "Assay Calibrator: Biosimilar or Reference")

5. The interbatch % total error should be ≤30% for the high, mid, and low QCs and ≤40% for the ULOQ and LLOQ QCs.
6. At least 67% of the QC samples should comply with the criteria, and not all QCs should fail at the same level.

In addition, for confirming bioanalytical similarity between the biosimilar and reference QCs, the following acceptance criteria should also be applied:

1. The absolute difference between biosimilar mean bias (%RE) and the reference mean bias (%RE) for the QCs calculated from the interbatch data should be ≤20% for the high, mid, and low QCs and ≤25% for the ULOQ and LLOQ QCs.
2. The 90% confidence interval for the difference between the biosimilar interbatch mean bias (%RE) and the reference interbatch mean bias (%RE) should be contained entirely within ±30% for high, mid, and low QCs and within ±35% for the ULOQ and LLOQ QCs.

Based on the above analysis, conclusions on range and sensitivity should be made for the biosimilar and the reference. A Microsoft Excel® Workbook entitled Biosimilarity Target Acceptance Criteria (Supplemental 1) is provided as an ancillary file to this publication. This Excel® Workbook can be validated and used as a tool to assess the above recommended target acceptance criteria.

Selectivity

Selectivity of the assay to quantify the biosimilar and the reference in the presence of other unrelated substances in the study samples should be determined independently. It is recommended to evaluate the LLOQ QC sample and a blank sample from at least 10 animals or 10–15 individuals from the relevant population (healthy or diseased in case of humans) for both drugs. The mean bias (%RE) of the observed concentration for the LLOQ QCs for the respective drugs should be ±25% of the nominal value for 80% of the samples tested. Although there are no recommended criteria for the blank samples, these samples must be assessed for response and recovery shown to be <LLOQ QC for 80% of the samples tested. Inclusion of representative lipemic, hemolyzed, or other types of expected samples should be considered during selectivity evaluation of the biosimilar and the reference.

Dilution Linearity

Dilution linearity of the PK method should be evaluated by assessing the performance of a QC sample for both the biosimilar and the reference drug at the expected C_{max} concentration. The back-calculated concentrations for the diluted QC samples (against the biosimilar curve) should be ±20% of the nominal value. Dilution linearity experiments should be designed to study prozone or hook effect of the biosimilar and reference in the relevant biological matrix.

Stability

Stability of the biosimilar and the reference drug in the relevant biological matrix should be evaluated independently using samples at the low and high QC levels. If prior information regarding stability of the reference in the relevant matrix is available, the biosimilar manufacturer may not repeat this stability assessment for the reference. However, if a clinical study is conducted in a new population, which has not been tested previously by the reference drug manufacturer, it is recommended to test the stability of the reference (in matrix) under all the anticipated transport and storage conditions, *e.g.*, room temperature/bench top, freeze/thaw, +2°C to +8°C, −20±5°C, and ultra-low temperatures (≤−70°C).

Documentation for Validation

Regulated PK assay validation is one of the most critical aspects of any study. Data acquired during validation establishes the reliability of the developed method to consistently quantify the drug under study in the relevant biological matrix in support of a trial/study.

This fit for purpose activity should be SOP and protocol driven. All raw and calculated data should be verified by a Quality Unit and reported in a consolidated validation report. Failure of samples to meet the target acceptance criteria or differences in the validation results between the biosimilar and reference should be thoroughly investigated and adequately documented in the validation report before proceeding into sample analysis.

SECTION 3: RECOMMENDED STATISTICAL ANALYSIS FOR ESTABLISHING BIOANALYTICAL SIMILARITY

Establishing Calibration Curve Bioanalytical Similarity Between the Biosimilar and Reference During Method Development

The standard curve comparison outlined below may be used to compare the biosimilar and reference standard curves in early development. The recommendation below includes a standard curve comparison plate set up as well as recommended acceptance criteria.

Bioanalytical similarity between the reference and biosimilar standard curves can be most appropriately assessed for PK assays by comparing the mean %RE of the standard curve calibrators and QCs. Acceptable differences between biosimilar and reference in method development should be compatible with acceptance criteria for method validation. For some assays, especially when the slopes of the biosimilar and reference curves are visually different or intersect slightly, it may be desirable to make a scientific assessment of similarity instead of a visual assessment. For example, one approach would be to fit the biosimilar and reference to appropriate calibrator curve models (*e.g.*, four- or five-parameter logistic model algorithm) and use the extra sums-of-squares *F* test as a scientific assessment of the similarity of the curves. *P* values obtained from this test that are greater than 0.05 are deemed to indicate similarity of the

curves. This approach may not be appropriate for all curves due to differences in the asymptotes that are only manifested beyond the upper and lower limits of quantitation so no formal acceptance criteria are proposed. For some assays, when the slopes of the biosimilar and reference curves are vastly different or intersect severely, the sponsor should consider using two separate assays. Ensuring that the slope and asymptotes of the biosimilar and reference curves are similar provides additional assurance that the acceptance criteria in method validation can be met successfully and results in increased understanding of potential challenges with the assay meeting the formal acceptance criteria in validation. However, this approach requires specialized software. One choice is ALLFIT, developed by the US National Institutes of Health and available freely from http://abs.cit.nih.gov/allfit/register.html (19).

Establishing QC Bioanalytical Similarity Between the Biosimilar and Reference During Method Validation

Formulas for calculating the intra- and interbatch mean bias (%RE) and %CV and interbatch total error are provided as Supplemental 2. The mean difference between the interbatch mean bias (%RE) for the biosimilar and reference is calculated by subtraction for each QC concentration. When biosimilar and reference are placed on the same plate, the calculation of the two-sided 90% confidence interval is performed using methodology for paired samples. For each batch, the intrabatch mean bias (%RE) for the biosimilar is subtracted from that of the reference, which results in a difference in mean bias for each batch. The overall mean and standard deviation of these differences is then calculated. The confidence interval is then obtained using the formula:

$$\bar{x}_d \pm t_{0.05, N-1} \frac{s_d}{\sqrt[2]{N}}$$

where $\bar{x}_d$ is the mean of differences for the intrabatch mean bias (%RE) across the six runs, $t_{0.05, N-1}$ is the critical value from a t distribution with a one-sided significance level of 0.05 and $N-1°$ of freedom, and $\frac{s_d}{\sqrt[2]{N}}$ is the standard error of the differences (standard deviation divided by the square root of the number of runs).

Although the two-sided 90% confidence interval includes a significance level of 0.05 on each side, the overall significance level of the test is 0.05 according to the two one-sided testing methodology (20). An Excel spreadsheet to perform these calculations is available via download from the AAPS website.

The sample size for method validation of $n=6$ assay runs is to ensure that assays with intermediate precision of no more than 15% CV and a difference in interbatch mean bias (%RE) of less than or equal to 10% between the biosimilar and reference would pass the proposed acceptance criteria with at least 90% probability. The sample size for assays that are more variable or have more bias may need to be assessed on a case-by-case basis.

SECTION 4: IN-STUDY VALIDATION AND SAMPLE ANALYSIS

The validated method should be used for all sample analysis. All assay plates should include the appropriate calibrator curve selected during the development phase. All plates should contain the low QC, mid QC, and high QC which are prepared using the same drug which was selected for the preparation of the standard curve (biosimilar or reference). QCs should be on the plate at least two times, with two replicates of each.

Acceptance Criteria for Routine Sample Analysis

The recommended acceptance criteria described below are based on the established industry standards for bioanalytical PK method validation and are not unique to biosimilar sample analysis.

Standard Curve

All calibrator samples should be analyzed as two replicates, and the mean per concentration should be calculated. At least 75% of the calibrator mean concentrations, or a minimum of six, must be within ±20% of their nominal concentration (25% at the upper and lower limits of the curve).

Quality Controls

All QC samples should be analyzed in duplicate, and the mean per concentration should be calculated. At least four of the six QC sample means must be within ±20% of their nominal concentration.

The %CV of all QC sample means within acceptable range must be ≤20%. At least one mean at any QC concentration level must be within ±20% of its nominal concentration.

Samples

The %CV of the mean back-calculated sample result must be ≤20%. Follow established system suitability requirements for additional sample acceptance criteria.

CONCLUSION

This white paper establishes a systematic process for assessing the degree of bioanalytical similarity between a biosimilar and a reference therapeutic. The two-staged process consists of a development phase and a validation phase. Within each phase, acceptance criteria are provided to permit the evaluation of analytical biosimilarity. Successful completion of this process will provide empirical data that the bioanalytical laboratory can use to support implementation of a single assay for quantification of a biosimilar and reference therapeutic to support PK analysis of a biosimilar drug development program.

The Biosimilars APC of the LBABFG recommends:

1. A single LBA be used to support PK assessments during biosimilar drug development.

2. During the development phase of a LBA, the standard curve generated using biosimilar drug should be bioanalytically similar to a standard curve generated using reference drug.

3. During the validation phase of a LBA, the nominal concentration of QC samples generated using biosimilar drug should be bioanalytically similar to the nominal concentration of QC samples generated using reference drug.

4. The acceptance criteria provided for the validation should be followed to ensure that one bioanalytical method can be used between the biosimilar and reference therapeutic. If any validation parameter does not meet the recommended acceptance criteria, an investigation should be conducted, and a decision should be made to determine if the assay needs to be optimized or if two assays should be used.

REFERENCES

1. Biological approvals by year. Vaccines, Blood & Biologics. U.S. Department of Health and Human Services. Food and Drug Administration. Silver Spring, MD. 2014. http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/BiologicalApprovalsbyYear/ucm385842.htm . Accessed 24 June 2014.
2. U.S. Department of Health and Human Services. Food and Drug Administration. Docket No. FDA-2010-N-0477. Approval pathway for biosimilar and interchangeable biological Products. Public Hearing November 2010.
3. European Medicines Agency. CHMP/437/04. Guideline on similar biological medicinal products. October 2005.
4. European Medicines Agency. EMEA/CHMP/BMWP/42832/2005. Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues. June 2006.
5. European Medicines Agency. EMEA/CHMP/BWP/49348/2005. Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Quality issues. June 2006.
6. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). Biosimilars: Questions and answers regarding implementation of the biologics price competition and innovation Act of 2009. February 2012 (Draft).
7. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). Quality considerations in demonstrating biosimilarity to a reference protein product. February 2012 (Draft).
8. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). Scientific considerations in demonstrating biosimilarity to a reference product. February 2012 (Draft).
9. DeSilva B, Smith W, Weiner R, Kelley M, Smolec JM, Lee B, *et al*. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. Pharm Res. 2003;20:1885–900.
10. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Veterinary Medicine (CVM). Guidance for industry: Bioanalytical method validation. May 2001.
11. European Medicines Agency. EMEA/CHMP/EWP/192217/2009. Guideline on bioanalytical method validation. February 2012.
12. Smolec J, DeSilva B, Smith W, Weiner R, Kelley M, Lee B, *et al*. Bioanalytical method validation for macromolecules in support of pharmacokinetic studies. Pharm Res. 2005;22:1425–31.
13. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, *et al*. Workshop/conference report—quantitative bioanalytical methods validation and implementation: Best practices for chromatographic and ligand binding assays. AAPS J. 2007;9:E30–42.
14. Fast DM, Kelley M, Viswanathan CT, O'Shaughnessy J, King SP, Chaudhary A, *et al*. Workshop report and follow-up—AAPS workshop on current topics in GLP bioanalysis: Assay reproducibility for incurred samples—implications of Crystal City recommendations. AAPS J. 2009;11:238–41.
15. Cai X-Y, Gouty D, Baughman S, Ramakrishnan MS, Cullen C. Recommendations and requirements for the design of bioanalytical testing used in comparability studies for biosimilar drug development. Bioanalysis. 2011;3:535–40.
16. Findlay JW, Smith WC, Lee JW, Nordblom GD, Das I, DeSilva BS, *et al*. Validation of immunoassays for bioanalysis: A pharmaceutical industry perspective. J Pharm Biomed Anal. 2000;21:1249–73.
17. Findlay JW, Dillard RF. Appropriate calibration curve fitting in ligand binding assays. AAPS J. 2007;9:E260–7.
18. European Medicines Agency. CHMP/437/ 04 Rev 1. Guideline on similar biological medicinal products. May 2013 (Draft).
19. Munson, P. ALLFIT Registration. 2007. http://abs.cit.nih.gov/allfit/register.html . Accessed 24 June 2014.
20. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987;15:657–80.