IPRP International Pharmaceutical Regulators Programme

20 December 2020

The Basics of the Clinical Comparability Exercise for Biosimilar Monoclonal Antibodies: Training Manual for Regulatory Reviewers

IPRP Biosimilars Working Group

Contents

1. DISCLAIMERS

- This document reflects the views of subject matter experts participating in the IPRP Biosimilars Working Group (BWG) and should not be construed to represent the official views of any given regulatory authority participating in the IPRP.
- This material is a compilation of publicly available information on the current approach for the clinical comparability exercise for biosimilars, particularly monoclonal antibodies.
- This material does not include any specific recommendations of the IPRP BWG and the views and opinions expressed in this material are those of the individuals who serve in his/her personal capacity and do not necessarily reflect the official policy or position of any agency or organization.
- Names of products or manufacturers used in this material are only examples to help readers' understanding and do not reflect any support of IPRP, WHO, or other organizations for licensing/authorization or ensuring quality/safety/efficacy of products.
- This material does not create any specific rights for anyone to use commercially. It is not protected under copyright and is accessible by anyone who wants to use it.
- This material is intended to help regulatory reviewers who already have review experience and a certain level of understanding of biotherapeutics, before they begin to review clinical aspects of biosimilars.
- This material could be used as an initiation step in the training of biosimilarity assessments and as a complementary tool and interactive course, such as hands-on training.
- This material includes relevant guidelines currently published and cases of biosimilars approvals. These examples may not entirely reflect the current practices of the respective agencies and consultation with the agency is recommended for the most updated advice.

2. CONCEPTS FOR THE CLINICAL COMPARABILITY ASSESSMENT FOR SIMILARITY/BIOSIMILARITY

Executive Summary

Role of clinical studies in biosimilar development programs

Regulatory authorities are in general agreement that the role of comparative clinical studies in development programs for "similar biologic products"/"biosimilars" is to provide clinical contextual information about the impact of differences that may be observed between the biosimilar and its reference biologic in analytical, structural, and functional data, and also non-clinical data, if non-clinical studies have been performed. There is also general agreement that the extent and type of clinical studies that may be needed in a biosimilar development program is dependent on product- and program-specific information and the nature and extent of residual uncertainty based on the available data in the development program. However, generally, there has been a default expectation among regulatory authorities for comparative clinical study information to support the demonstration of biosimilarity. Some regulatory authorities are explicit about having a minimum expectation for comparative pharmacokinetic (PK)/pharmacodynamic [PD] (if PD is relevant and

available) studies and other authorities are less definitive about the clinical comparability studies that may be required for a given product or development program.

Comparative PK/PD Studies

Regarding comparative PK/PD studies, regulatory authorities generally utilize common principles for the design and analysis of PK studies, following the "average bioequivalence" approach used for small molecule generic drugs and using standard bioequivalence acceptance criteria (i.e., 90% confidence interval [CI] for the ratio of the geometric means is within 80-125%). When feasible and ethical, single-dose cross-over studies in healthy subjects are preferred because such a study is likely to be the most sensitive for detecting potential differences between the biosimilar product and the reference product. However, many biologics may have characteristics that call for a different study design (e.g., parallel group design for products that have long half-lives, or limits on acceptable population or dose(s)). Additionally, studies that include PD endpoints entail product- and program-specific considerations, as well as clinical considerations, and expectations may differ somewhat among regulatory authorities.

Clinical Comparability Studies

Regulatory authorities are in agreement that comparative clinical studies evaluating efficacy endpoints and safety/immunogenicity are intended to support conclusions that there are no clinically meaningful differences between a biosimilar and its reference product and are not intended to re-prove efficacy. Although regulatory authorities are in general agreement about the overarching goals and principles for comparative clinical efficacy and safety studies, detailed expectations/ recommendations among regulatory authorities may differ on a case-by-case basis, depending on country-specific statutory requirements, and product- and program-specific facts and issues.

Table 1: Summary of Approach to Comparative PK or PK/PD and Comparative Clinical Efficacy and Safety Studies, by Regulatory Authority

	Health Canada	MFDS	PMDA	FDA, United States	EMA
PK or PK/PD studies					
Subjects	Healthy Volunteers preferred if feasible and relevant	Healthy volunteers, if ethical, or the most sensitive model/ patient group	Depends on product and clinical context	Healthy volunteers preferred if feasible; otherwise the most sensitive population, dose, and route of administration in patients	Healthy volunteers, if feasible. Otherwise in a sensitive model/population. Supportive PK data from clinical studies in patients are encouraged.
Dose	For Healthy: A low therapeutic dose or a sub-therapeutic dose on the linear part of the dose response curve For Patients: an approved dose of reference drug	Dose should be selected within the recommended therapeutic dose range of the reference drug. If SQ and IV routes available, SQ should be tested; IV route may not be necessary	Use dose approved for the original biologic unless there is a scientific rationale to use a different dose within the dose range	Use a dose that is adequately sensitive to allow for the detection of differences in PK/PD, if feasible.	The doses in the single dose PK biosimilar comparability study in healthy volunteers may be lower than the recommended therapeutic doses.
Design	Depends on product and linearity of PK, clinical context	Depends on product characteristics. PK endpoints in a single-dose study: Primary: AUC(0-inf) and Cmax. For IV route, only AUC(0-inf); Secondary: Tmax, volume of distribution; half-life Repeat dose study: Primary: Truncated AUC after 1st dose through the second dose (AUC0-t) and AUC over dose interval at steady state (AUCτ); Secondary: Ctrough and Cmax at steady state	Depends on product properties PK endpoints: AUC, Cmax	Single-dose, randomized, crossover design is generally preferred for a product with a short half-life, a rapid PD response, and low anticipated incidence of immunogenicity. Parallel group design may be appropriate for products with a long-half life or where repeated exposures may lead to time—related changes with exposure to drug.	A single dose cross-over study is preferable. A parallel group design may be necessary with substances with a long half-life and/or a high risk of immunogenicity. If the reference product can be administered both intravenously and subcutaneously, the evaluation of subcutaneous administration will usually be sufficient as it covers both absorption and elimination. For a single dose study, the primary parameters are the AUC(0-inf) for intravenous administration and AUC(0-inf) and usually Cmax for subcutaneous administration. In a multiple dose study, the primary parameters should be the truncated AUC after the first administration until the second administration (AUC0-t) and AUC over a dosage interval at steady state (AUC□).

	Health Canada	MFDS	PMDA	FDA, United States	EMA
Acceptance Limits	Should be pre-defined and justified and include parameters reflecting absorption and elimination (generally 90% CI of GMR within 80-125%)	Generally standard BE 90% CI of Geometric Mean Ratio (GMR) within 80-125%, unless otherwise noted or justified by company	Should be pre-defined and justified	Generally, 90% CI of GMR within 80-125%, but acceptable limits may vary among products	Limits for the main PK parameters should be pre-defined and justified. The criteria used in standard clinical bioequivalence studies, initially developed for chemically derived, orally administered products, may be a reasonable basis for planning
PD considerations	Use of a particular PD marker should be clinically relevant and scientifically justified.	One dose should be within the steep part of the dose-response curve. Generally, clinical efficacy trial would also be expected unless PD marker(s) is established surrogate for efficacy	Include PD marker if appropriate; may be helpful if PK studies are technically difficult to conduct.	Include PD marker(s) if relevant. A multiple-dose design may be appropriate when the PD effect is delayed or not parallel to the single-dose PK profile.	It is recommended that pharmacodynamic (PD) markers are added to the pharmacokinetic studies whenever feasible. The PD markers should be selected on the basis of their relevance to the clinical outcome.
Immunogenicity		Measure anti-drug antibody levels at time of PK sampling		Generally, samples can be taken and stored for future analysis, if immunogenicity assays have not yet been developed. If the reference product has a high potential for immune-mediated toxicity, immunogenicity assays and testing should be developed in advance to allow for real-time assessment.	Anti-drug antibodies should be measured in parallel to PK assessment using appropriate sampling time points.
Reference	Guidance Document: Information and Submission Requirements for Biosimilar Biologic Drugs, November 2016, pages 15-19.	Guidelines on the Evaluation of Biosimilar Products, English version, Revision 1 (http://www.mfds.go.kr/eng/wpge/m_37/de011024l001.do)	Guideline for the Quality, Safety, and Efficacy Assurance of Follow-on Biologics (https://www.pmda.go.jp/files/000153851.pdf)	Guidance for Industry: Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product, December 2016, https://www.fda.gov/media/88622/download	Similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-similar-biological-medicinal-products-containing-biotechnology-derived-proteins-active_en-2.pdf Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-similar-biological-medicinal-products-containing-monoclonal-antibodies-non-clinical_en.pdf
Comparative Clinical Efficacy and Safety Studies					

	Health Canada	MFDS	PMDA	FDA, United States	EMA
Subjects	Most sensitive population to rule out clinically meaningful differences.			Population should allow for an assessment of clinically meaningful differences. Typically, in one of the approved indications, but could be different if scientifically justified.	The study population should generally be representative of approved therapeutic indication(s) of the reference product and be sensitive for detecting potential differences between the biosimilar and the reference. Occasionally, changes in clinical practice may require a deviation from the approved therapeutic indication
Design	Equivalence trials generally preferred. Other designs (i.e., non-inferiority) should be justified and discussed with Health Canada prior to study initiation	Equivalence trial preferred rather than non-inferiority trial, unless adequate scientific and clinical justification. Double-blind preferable or at minimum observer-blind.	Design and margins should be prespecified and justified. Use clinically established endpoints. If surrogate endpoints are available, primary efficacy endpoints are not always required, with thorough justification. However, a clinical study to evaluate safety and immunogenicity may still be needed.	Typically, an equivalence design with a symmetric margin. However, in some cases an asymmetric interval or non-inferiority design may be justifiable. Discuss with FDA before initiating the study(ies).	In the absence of surrogate markers for efficacy, it is usually necessary to demonstrate comparable clinical efficacy of the biosimilar and the reference medicinal product in adequately powered, randomised, parallel group comparative clinical trial(s), preferably double-blind, by using efficacy endpoints. In general, an equivalence design should be used. The use of a non-inferiority design may be acceptable if justified on the basis of a strong scientific rationale and taking into consideration the characteristics of the reference product, e.g. safety profile/tolerability, dose range, dose-response relationship
Equivalence or Non-Inferiority Margin	Choice of endpoints and margin should be clinically justified and pre-specified.	Should be pre-defined and appropriately justified in a range that could rule out clinically meaningful differences. May be justifiable to use efficacy endpoints and timing that are different from historical trials with the reference product.		The approach to choosing the margin, testing the hypothesis, and general study conduct should generally follow the Guidance for Industry: Non-Inferiority Clinical Trials to Establish Effectiveness November 2016. https://www.fda.gov/media/78504/download. Discuss with the appropriate FDA review division prior to initiating the study.	Comparability margins should be pre-specified and justified on both statistical and clinical grounds by using the data of the reference product
Safety / Immunogenicity	Suitably sensitive population with an adequate number of patients and adequate duration to rule out clinically meaningful differences in safety / immunogenicity. Should include incidence and magnitude of anti-drug antibody (ADA)	Required observation period for immunogenicity testing should be adequate duration to observe for clinically significant antibody formation. Comparison of frequency and pattern of ADA, clinical effects arising from immune reactions,	Although in some cases comparative efficacy can be assessed in a comparative PD study using a surrogate endpoint, a clinical study to evaluate safety and immunogenicity may still be needed. ADA and neutralizing	Typically, comparative safety and immunogenicity is descriptive. Immunogenicity testing should generally follow the Guidance for Industry: Immunogenicity Assessment for Therapeutic Protein Products. August 2014.	Comparative safety data should normally be collected pre-authorisation, their amount depending on the type and severity of safety issues known for the reference product. The duration of safety follow-up pre-authorisation should be justified.

	Health Canada	MFDS	PMDA	FDA, United States	EMA
	response, time-course of ADA development, ADA persistence, and impact of ADA on safety, efficacy and PK	Impact on PK, safety, and efficacy would be needed if there is increased ADA formation with the biosimilar.	antibodies should be evaluated. Class, affinity, and specificity of the antibodies should be analysed. Consider whether ADA are related to impurities or specific carbohydrate antigens.	https://www.fda.gov/media/85017/download	
Reference	Guidance Document: Information and Submission Requirements for Biosimilar Biologic Drugs, November 2016, pages 16-19.	Guidelines on the Evaluation of Biosimilar Products, English ver. Rev. 1 (http://www.mfds.go.kr/eng/wpge/m_37/de011024l001.do)	Guideline for the Quality, Safety, and Efficacy Assurance of Follow-on Biologics (https://www.pmda.go.jp/files/000153851.pdf)	Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, April 2015 https://www.fda.gov/media/82647/download pages 18-20.	Similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-similar-biological-medicinal-products-containing-biotechnology-derived-proteins-active_en-2.pdf Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-similar-biological-medicinal-products-containing-monoclonal-antibodies-non-clinical_en.pdf

1.1. Role of clinical studies in development programs for similar biologic products / biosimilars

1.1.1. Health Canada

The purpose of the clinical program is to show that there are no clinically meaningful differences between the biosimilar and the reference biologic drug. The clinical program should begin with a PK/PD study(ies) which may be followed by an additional clinical trial(s). Differences observed between the biosimilar and reference biologic drug, such as differences in immunogenicity, should be addressed. If differences cannot be addressed; the sponsor should consider whether the biosimilar submission route is still appropriate or whether the traditional new drug submission route would be more suitable.

Non-clinical and clinical study requirements are applicable to biosimilars that have been demonstrated to be similar to the reference biologic drug based on the results of the comparative structural and functional studies included in the chemistry and manufacturing data package. If similarity has not been established, reduced non-clinical and clinical data cannot be justified and the product cannot be considered a biosimilar.

Specific study requirements for drug classes (e.g. insulin and growth hormone) may differ depending on the class and on various clinical parameters such as therapeutic index.

Clinical data should be generated based on the product for which market authorization is sought. Chemistry and manufacturing changes introduced during the clinical development phase or at the end of the clinical development program may result in the need for additional bridging data.

From: Guidance Document: Information and Submission Requirements for Biosimilar Biologic Drugs, November 2016 pages 14-15.

1.1.2. Ministry of Food and Drug Safety (MFDS), Republic of Korea

Pivotal clinical data should be generated using the product derived from the final manufacturing process. If the manufacturing process of the drug products used in clinical studies is different from the final manufacturing process for which marketing authorization is sought, such differences should be justified, and additional data may be required.

The clinical comparability exercises include pharmacokinetic, pharmacodynamic, and efficacy studies. If the comparability can be demonstrated by confirmatory pharmacokinetic/pharmacodynamic data, an efficacy study may be omitted.

From Guidelines on the Evaluation of Biosimilar Products, English version, Revision 1 (http://www.mfds.go.kr/eng/wpge/m_37/de011024l001.do)

1.1.3. *Pharmaceutical and Medical Devices Agency (PMDA), Japan*

The purpose of the clinical studies is to demonstrate the comparability between a "follow-on" biologic and the original biologic. In the guideline, "comparability" means that they are highly similar and that existing knowledge is sufficiently predictive to ensure that any differences in quality attributes have no adverse impact on the drug product or on its safety or efficacy.

The quality attributes of the follow-on biologic of interest, the results of the comparative studies of relevant quality attributes between the follow-on biologic and the original biologic, and the findings of non-clinical studies should be considered to conduct clinical studies.

Even though high similarity in quality has been demonstrated through comparability studies on the quality attributes, an analysis of all data from the PK, PD or PK/PD studies might not demonstrate the comparability of clinical efficacy. In this case, it is necessary to conduct clinical studies to verify that the efficacy of the follow-on and originator biologics in respect of the indications of the product for which approval is sought are comparable. The type and contents of necessary clinical studies will vary widely according to available information and the properties of the original biologics.

From: Guideline for the Quality, Safety, and Efficacy Assurance of Follow-on Biologics (https://www.pmda.go.jp/files/000153851.pdf)

1.1.4. *Food and Drug Administration, United States (FDA, US)*

The role of clinical studies in biosimilar development programs is in support of the statutory requirement for information demonstrating that: "there are no clinically meaningful differences between the biological product and the reference product in terms of the safety, purity, and potency of the product. The nature and scope of the clinical study or studies will depend on the nature and extent of residual uncertainty about biosimilarity after conducting structural and functional characterization. The frequency and severity of safety risks and other safety and effectiveness considerations (e.g., poor relationship between pharmacologic effects and effectiveness) for the reference product may also affect the design of the clinical program. The scope of the clinical program and the type of clinical studies (i.e., comparative human PK, PD, clinical immunogenicity, or clinical safety and effectiveness) should be scientifically justified by the sponsor."

Reference: Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, April 2015 https://www.fda.gov/media/82647/download. Pages 13-14.

1.1.5. *European Medicines Agency (EMA)*

Generally, the aim of clinical data is to address slight differences shown at previous steps. Clinical data cannot be used to justify substantial differences in quality attributes

Reference: Guideline on similar biological medicinal products (https://www.ema.europa.eu/en/similar-biological-medicinal-products)

Efficacy trials of biosimilar medicinal products do not aim at demonstrating efficacy per se, since this has already been established with the reference product. The purpose of the efficacy trials is to confirm comparable clinical performance of the biosimilar and the reference product.

Reference: Similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues (https://www.ema.europa.eu/en/similar-biological-medicinal-products-containing-biotechnology-derived-proteins-active-substance-non).

Further product-specific guidance on non-clinical and clinical issues is available on the following webpage for a number of product types, including monoclonal antibodies: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/multidisciplinary/multidisciplinary-biosimilar

1.2. Comparative Human Pharmacokinetic/Pharmacodynamic Studies

1.2.1. Health Canada

Pharmacokinetic (PK) / Pharmacodynamic (PD) Studies: Study Design and Analysis

Comparative PK studies should be conducted to rule out differences in PK characteristics between the biosimilar and the reference biologic drug.

- PK studies should be carried out in healthy subjects when appropriate as they are usually considered to be a homogeneous and sensitive population. A low or sub-therapeutic dose residing on the linear part of the dose response curve should be considered if studies are performed in healthy subjects.
- Studies should be conducted in the patient population when the PK or PK/PD in the patient population is known to be substantially altered by the disease states for which authorisation is requested or due to ethical and safety concerns for conducting PK studies in healthy volunteers. The dose(s) used in the PK studies in a relevant patient population should be within the therapeutic dosing range specified in the product monograph of the reference biologic drug.
- The following factors should be taken into consideration during comparative PK study design (e.g. when choosing between cross-over versus parallel-group study):
    - half-life of the biologics
    - linearity of PK parameters
    - where applicable, the endogenous levels and diurnal variations of the protein under study
    - conditions and diseases to be treated
    - route(s) of administration, and
    - indications for which the biosimilar sponsor is applying.
- Acceptable criteria for the determination of similarity in comparative pharmacokinetics should be defined and justified prior to the initiation of PK study(ies).

General principles of study design and statistical methods for comparative bioavailability studies should be considered when assessing the similarity of the PK parameters between the biosimilar and the reference biologic drug. The PK comparison should not be limited to parameters reflecting absorption only. Parameters representing elimination (e.g. clearance and terminal

half-life) should also be compared. Data should not be excluded from the analysis unless the exclusion can be justified and is considered acceptable by Health Canada.

Regarding PD studies, as for all other studies in the biosimilar developmental program, these studies should be comparative in nature.

- Parameters investigated in PD studies should be clinically relevant. Use of a particular PD marker should be scientifically justified. PD markers should be relevant to the mechanism of action of the drug but may not need to be established surrogates for efficacy.
- In general, the principles regarding study design, conduct, analysis and interpretation that are relevant to equivalence trials with a clinical outcome as the primary endpoint are applicable to equivalence trials with a PD marker as the primary outcome.
- PD studies should be combined with PK studies, in which case the PK/PD relationship should be characterized.

Refer to Guidance Document: Information and Submission Requirements for Biosimilar Biologic Drugs, November 2016, pages 15-19.

1.2.2. MFDS

Pharmacokinetic (PK)/Pharmacodynamic (PD) Studies: Study Design

In principle, PK studies should generally be performed for all proposed routes of administration and dose should be selected within the recommended therapeutic dose range of the reference product.

PK studies should be comparative in nature to demonstrate the comparability of the biosimilar product and should be designed to enable detection of potential differences between the biosimilar product and the reference product.

In general, this is achieved effectively by performing single-dose PK studies in a sensitive and homogenous study population and by using a sensitive enough dose to detect differences to reach its maximum value.

- The choice of single-dose studies, steady-state studies, or repeated determination of PK parameters and the study population should be justified.
- If the cross-over design is adopted, it is necessary to demonstrate that the half-life, antibody formation, and other characteristics do not affect the PK profiles.
- If the parallel design is selected, careful attention should be paid to avoid potential imbalances between groups.
- In PK studies for demonstrating comparability, healthy volunteers could be considered as a sensitive and homogenous population for study if considered ethical.
- When a patient group is selected as a subject in a PK study, the most sensitive model/patient group that is able to minimize any major inter-individual or time-dependent variation should be selected.

Pharmacokinetic endpoints to be considered depend on the study design.

- Single-dose PK study

  - Primary endpoint parameters: $AUC_{(0\text{-}inf)}$ and C_{max} for intravenous administration, only $AUC_{(0\text{-}inf)}$ is the primary endpoint parameter.
  - Secondary endpoint parameters: t_{max}, volume of distribution, and half-life

- In a repeat-dose PK study
  - Primary endpoint parameters: The truncated AUC after the first administration until the second administration $AUC_{(AUC(0\text{-}t))}$ and AUC over a dose interval at steady state $AUC_{(AUC\tau)}$
  - Secondary endpoint parameters: C_{trough} and C_{max} at steady state.

In the comparative PD studies, PD effects should be investigated in a suitable patient population using one dose within the steep part of the dose-response curve in order to detect potential differences between the biosimilar product and the reference product in the most sensitive manner.

If it is possible to use PD markers well established in healthy volunteers, the comparative evaluation of PD effects may be conducted using healthy volunteers.

There are pharmacodynamic surrogate markers that are associated with the clinical efficacy as shown below:

- Granulocyte-colony stimulating factor (G-CSF): Absolute neutrophil counts.
- α-interferon: Initially decreased viral concentrations in patients with chronic hepatitis C.
- Insulin: Euglycaemic clamp test.
- β-interferon: Magnetic resonance imaging (MRI) scans of the lesions.

In general, for the demonstration of the efficacy of biosimilar products, clinical trials should be conducted. In the following cases, however, a comparative PK/PD may be alternatively performed.

- A reference product with well-established pharmacodynamic and pharmacokinetic characteristics.
- More than one PD surrogate marker that is indicative of the efficacy.
- A reference product with well-established dose-exposure relationship, PD parameters and response-efficacy relationship.

Pharmacokinetic (PK)/Pharmacodynamic (PD) Studies: Study Analysis

- At the time of specimen collection, the level of anti-drug antibody is measured in conjunction with pharmacokinetic studies.
- If the approved route of administration of the reference product is either an intravenous or subcutaneous route, absorption and elimination should be observed.
- Once the comparability in respect to absorption and elimination of the subcutaneous route is demonstrated, it may not be necessary to conduct the comparability exercise for the intravenous route.
- To demonstrate the pharmacokinetic comparability between the biosimilar product and the reference product, acceptance range should be defined and then justified. Unless otherwise noted, acceptance range of 80-125% may be used as they have been used for standard bioequivalence studies.

- If there are attempts to broaden the margin of comparability, the justification should be made considering their potential effects on the efficacy and safety.
- Test materials (drugs or metabolites) should be detected within the range of quantification based on the optimal specificity, sensitivity, precision and accuracy. Furthermore, time-dependent changes should also be evaluated.
- If the active ingredient of a biosimilar product is an endogenous protein and the concentration of the endogenous protein is measurable, the concentration-time profile of the administered exogenous protein may be substantially affected. In these cases, it would be mandatory to describe valid methods for the purpose of minimizing the effects of endogenous protein.

Refer to Guidelines on the Evaluation of Biosimilar Products, English version, Revision 1 (http://www.mfds.go.kr/eng/wpge/m_37/de011024l001.do)

1.2.3. PMDA

Pharmacokinetic (PK)/Pharmacodynamic (PD) Studies: Design and Analysis

- In principle, the sponsor should conduct the comparability exercise for cross-over PK studies which are carefully designed to evaluate comparability between the follow-on and original biologics. However, since a cross-over study may not always be applicable to clinical studies for biologics with a long half-life (e.g., antibodies, PEG-binding proteins) or biologics that may produce antibodies in humans, the clinical study should be designed according to the properties of the follow-on biologic (e.g., parallel-group design).
- Depending on the original biologic and/or target disease, it may be appropriate to conduct a clinical study in healthy adults, while a clinical study enrolling patients is sometimes more appropriate.
- It is necessary to conduct a clinical study using the same route of administration as that in the approved indications of the original biologic. Where multiple routes of administration are allowed, in principle, each route of administration should be studied.
- Clinical studies should be conducted using the approved dosage of the original biologic, while a scientifically rational dosage within the dosage range of the original biologic may also be chosen.
- While key parameters of a PK study include the area under the blood concentration curve (AUC) and maximum concentration (Cmax), the acceptable range of data from the comparability exercise (comparability margin) should be determined before the study. In this case, the margin of the acceptable range set should be fully justified.
- If possible, it is necessary to select PD marker(s) for clinical efficacy and to conduct the comparability studies between the follow-on biologic and original biologics using the appropriate PD marker. A comparative study with PD marker is particularly useful, if PK studies are technically difficult to conduct.

From: Guideline for the Quality, Safety, and Efficacy Assurance of Follow-on Biologics (https://www.pmda.go.jp/files/000153851.pdf)

1.2.4. FDA, US

Pharmacokinetic (PK)/Pharmacodynamic (PD) Studies: Study Design

Clinical pharmacology studies build on the comparative analytical studies in the stepwise approach to support a demonstration of biosimilarity and are normally a critical part of demonstrating biosimilarity by supporting a demonstration that there are no clinically meaningful differences between the proposed biosimilar product and the reference product.

- Single-dose, randomized crossover study-design is generally preferred and recommended for a product with a short half-life (e.g., shorter than 5 days), a rapid PD response (e.g., the time of onset, maximal effect, and disappearance in conjunction with drug exposure), and a low anticipated incidence of immunogenicity.
- A parallel group design may be appropriate for products that have a long half-life or where repeated exposures may lead to time-related changes associated with exposure to the drug.
- Include pharmacodynamic measure(s), if relevant. For PD similarity assessments, a multiple-dose design may be appropriate when the PD effect is delayed or otherwise not parallel to the single-dose drug PK profile.
- Publicly available information on the safety and immunogenicity profile of the RP should be considered for safety and immunogenicity assessments. Generally, samples can be stored for future analysis if assays are not yet developed for immunogenicity. If the RP has high potential for immune-mediated toxicity, assays capable of detecting binding antibodies (and their neutralizing potential) should be developed in advance to allow for real time assessment. Sampling/visits should take into account the expected appearance and resolution of safety signals or immune responses.
- Use a population, dose(s), and route of administration that are adequately sensitive to allow for the detection of differences in PK/PD profiles, if feasible. Clinical PK and PD studies should be conducted in healthy subjects if the product can be safely administered to them.

Pharmacokinetic (PK)/Pharmacodynamic (PD) Studies: Study Analysis

- Dose-exposure considerations: mAbs have both Target-mediated Drug Disposition (TMDD) which is dose-dependent and specific to the mAb and antigen, and target-independent nonspecific cellular uptake, which is not. Therefore dose-dependent and nonlinear clearance is often observed at low-dose levels; linear and predictable clearance is expected above the saturable dose-range.
- Consider integrity and interpretability of bioanalytical methods: Selection and operating characteristics of assay(s). For mAbs, typically ligand-based.
- Safety and Immunogenicity assessment: Publicly available Safety/Immunogenicity profile of the RP can inform the duration of follow up for safety signals or immunogenicity.
- Statistical Comparison: Needs 1) criteria/criterion, 2) confidence interval(s), 3) acceptable limit(s).
- FDA recommends log-transformation of exposure measures before statistical analysis and an average equivalence statistical approach; i.e., calculation of a 90% confidence interval for the ratio between the geometric means of the parameters of the proposed biosimilar and the RP.

- Acceptable limits may vary among products. For mAbs, a typical acceptable limit for the confidence interval of the ratio is 80-125%.

Reference: Guidance for Industry: Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product, December 2016, https://www.fda.gov/media/88622/download

1.2.5. EMA

Pharmacokinetic (PK)/Pharmacodynamic (PD) Studies: Study Design & Study Analysis

The design of a PK study depends on various factors, including clinical context, safety, PK characteristics of the reference product (target-mediated disposition, linear or non-linear PK, time-dependency, half-life, etc.) as outlined in the Guideline on the clinical investigation of the pharmacokinetics of therapeutic proteins (CHMP/EWP/89249/2004) and, as applicable, the Guideline on the investigation of bioequivalence (CPMP/EWP/QWP/1401/98 Rev. 1/ Corr). Furthermore, bioanalytical assays should be appropriate for their intended use and adequately validated as outlined in the Guideline on bioanalytical method validation (EMEA/CHMP/EWP/192217/2009).

The biosimilar comparability limits for the main PK parameters should be defined and justified prior to conducting the study. The criteria used in standard clinical bioequivalence studies, initially developed for chemically derived, orally administered products, may be a reasonable basis for planning comparative pharmacokinetic trials for biologicals in the absence of specific criteria. However, the interpretation of bioequivalence studies for biologicals is less straightforward than for small molecules. In the latter case the molecules are considered identical, whilst for biologicals, PK is used to detect possible differences in the interaction with the body between the originator and the biosimilar. This means that observing 90% CIs of ratios of biosimilar to reference product within a pre-specified, justified acceptance range may not, by itself, be sufficient. The location and the width of the confidence interval should also be taken into account in the interpretation of similarity. For example, statistically significant differences in 90% CIs within the justified acceptance range regarding relevant PK parameters would need to be explained and justified as not to preclude biosimilarity. On the other hand, if the 90% CI crosses the prespecified boundaries the applicant would need to explain such difference and explore root causes. Correction for protein content may be acceptable on a case-by-case basis if pre-specified and adequately justified, with the results from the assay of the test and reference products being included in the protocol. As a principle, any widening of the conventional equivalence margin beyond 80-125% for the primary parameters requires thorough justification, including an estimation of potential impact on clinical efficacy and safety.

The existence of target-mediated clearance in addition to non-target-mediated clearance may affect the number of studies needed. In case target-mediated clearance is not relevant, one comparative PK study may be sufficient. If the reference product/mAb is eliminated both by target-mediated and non-target-mediated mechanisms, comparable PK should be demonstrated where each mechanism of clearance predominates: preferably one study in healthy volunteers for non-target-mediated clearance and one supportive study in patients, which can be part of the efficacy trial, to investigate comparability in target-mediated clearance. If distinct therapeutic areas are involved for one particular mAb (e.g. autoimmunity and oncology), separate PK studies may be needed if different target-mediated clearance exists for different therapeutic areas.

Although the comparison of target-mediated clearance is of major importance in the biosimilarity exercise, it may not be feasible in patients due to major variability in target expression, including

variability over time. However, since in vitro studies are expected to show comparable interaction between the biosimilar and its target(s) (including FcRn for a mAb), the absence of a pivotal PK study in the target population is acceptable, if additional PK data are collected during the efficacy, safety and/or PD studies as this allows further investigation of the clinical impact of variable pharmacokinetics and possible changes in the PK over time. This can be achieved by determining the PK profile in a subset of patients or by population pharmacokinetics.

A single dose cross-over study with full characterisation of the PK profile, including the late elimination phase, is preferable. A parallel group design may be necessary with substances with a long half-life and/or a high risk of immunogenicity. The doses in the single dose PK biosimilar comparability study in healthy volunteers may be lower than the recommended therapeutic doses.

PK studies are not always feasible in healthy volunteers. In this case, the PK needs to be studied in patients as part of a multiple dose study, if a single dose study is not feasible. A sensitive model/population, i.e. that has fewer factors that cause major inter-individual or time-dependent variation, should be explored.

If the reference product can be administered both intravenously and subcutaneously, the evaluation of subcutaneous administration will usually be sufficient as it covers both absorption and elimination. Thus, it is possible to waive the evaluation of intravenous administration if biosimilar comparability in both absorption and elimination has been demonstrated for the subcutaneous route. Omission of the PK study of intravenous administration needs to be justified, e.g., in cases when the molecule has an absorption constant which is much slower than the elimination constant (flip flop kinetics).

In a single dose PK study, the primary parameters are the $AUC_{(0\text{-}inf)}$ for intravenous administration and $AUC_{(0\text{-}inf)}$ and usually Cmax as co-primary parameter for subcutaneous administration. Secondary parameters such as t_{max}, volume of distribution, and half-life, should also be estimated. In a multiple dose study, the primary parameters should be the truncated AUC after the first administration until the second administration ($AUC_{0\text{-}t}$) and AUC over a dosage interval at steady state ($AUC\tau$). Secondary parameters are Cmax and C_{trough} at steady state. Characterisation of the full concentration-time profile at steady state is especially important in case of non-linear PK of the reference mAb (e.g. many anticancer mAbs with cellular targets exhibit dose- or time-dependent PK or immunogenicity-related changes in distribution or elimination kinetics).

In any PK study, anti-drug antibodies should be measured in parallel to PK assessment using appropriate sampling time points.

It is recommended that pharmacodynamic (PD) markers are added to the pharmacokinetic studies whenever feasible. PD markers are especially valuable if they are sensitive enough in order to detect small differences, and if they can be measured with sufficient precision. The use of multiple PD markers, if they exist, is recommended. With regard to pharmacodynamic evaluation, there is often a lack of specific PD endpoints. The emphasis may then have to be on non-clinical PD evaluations, e.g. *in vitro* testing.

- The PD markers should be selected on the basis of their relevance to the clinical outcome. In certain cases, comparative PK/PD studies may be sufficient to demonstrate clinical comparability of the biosimilar and the reference medicinal product, provided that the following conditions are met: At least one PD marker/biomarker is an accepted surrogate marker and can be related to patient outcome to the extent that demonstration of similar effect on the PD marker will ensure a similar effect on the clinical outcome.

- There may be PD-markers that are not established surrogates for efficacy but are relevant for the pharmacological action of the active substance and a clear dose-concentration-response or a time-response relationship has been demonstrated. A comparative single or repeat dose study in the saturation part of the dose-concentration-response curve is unlikely to discriminate between different activities, should they exist, and a dose in the linear part of the dose-response curve may result in treating a patient with a too low dose. It is also acknowledged that dose-response data may not exist for the reference mAb, and that exposing patients to a relatively low dose of the mAbs, in a worst-case scenario, might also sensitize them to develop anti-mAb antibodies, and, consequently, may make them treatment resistant. However, for some reference mAbs clinical conditions may exist where such studies are feasible.

When PD markers are planned as pivotal evidence to establish similarity, it is recommended to discuss such approach with regulatory authorities. This should include a proposal of the size of the proposed equivalence margin and its clinical justification as regards lack of a clinical meaningful difference as well as of the measures for demonstration of a comparable safety profile.

1.3. Comparative Clinical Efficacy and Safety Studies

1.3.1. Health Canada

Comparative Clinical Efficacy Trial(s): Study Design and Analysis

In most cases, a comparative clinical trial(s) is important to rule out clinically meaningful differences in efficacy and safety between the biosimilar and the reference biologic drug. A clinical efficacy trial may not always be necessary, e.g. where there is a clinically relevant PD endpoint. In such cases, a scientific justification is needed and safety as well as comparative immunogenicity data are still required.

- The comparative clinical trial should be adequately sensitive to rule out clinically meaningful differences within predefined comparability margins. In some instances, evaluation of more than one sensitive population may be necessary.
- Careful consideration should be given to the design of the study(ies) including the choice of primary efficacy endpoint(s) and clinical comparability margin. Each of these aspects are important and should be justified on clinical grounds.
- In line with the principle of similarity, equivalence trials are generally preferred. If non-inferiority trials are considered, they should be clearly justified, and sponsors are advised to consult with Health Canada prior to study initiation.
- Efforts should be made to ensure that comparative clinical studies have a sufficient number of patients treated for an acceptable period of time in order to rule out clinically meaningful differences in safety between the biosimilar and the reference biologic drug.
- A suitable population should be selected in which to compare immunogenicity. In selecting an appropriate population, factors such as immunocompetence, prior or concomitant use of

immunosuppressant therapies, and historical data with respect to the immunogenicity of the reference biologic drug should be considered.

The nature, severity and frequency of adverse events should be compared between the biosimilar and the reference biologic drug.

The purpose of the comparative immunogenicity study(ies) is to rule out clinically meaningful differences in immunogenicity between the biosimilar and the reference biologic drug. Of most concern are those antibodies that have the potential to impact safety and/or efficacy; for example, by altering PK, inducing anaphylaxis, or by neutralising the product and/or its endogenous protein counterpart. For each treatment arm, the comparative study(s) should characterise the incidence and magnitude of the anti-drug antibody (ADA) response, the time-course of ADA development, ADA persistence, and the impact of ADA on safety, efficacy and PK.

From Guidance Document: Information and Submission Requirements for Biosimilar Biologic Drugs, November 2016, pages 16-19.

1.3.2. MFDS

Comparative Clinical Studies: Study Design and Analysis

To adopt posology and route of administration and to accept extrapolation of indications of reference product, it is recommended to design the efficacy trial with equivalence study rather than non-inferiority study.

- Non-inferiority test could only be considered if provided with valid scientific evidence and when safety and tolerance, dosage range, dose-response relationship of the reference product and others are justifiable.
- Non-inferiority design could be applied when the likelihood of superiority in efficacy is excluded with certainty.
- Comparability margin should be pre-defined and appropriately justified. The margin should be selected within the range that would not show clinical differences from the reference product.
- Similar efficacy of the biosimilar product and the reference product should be demonstrated in an adequately powered, randomized, and parallel group clinical trial (equivalence trials). Such clinical studies should preferably be double-blind or at a minimum observer-blind.
- Product-specific guidelines, which recommend clinical designs for the demonstration of clinical efficacy by product type, could guide the choice of clinical endpoints. However, in certain circumstances, different methods (the choice of clinical endpoints, time points of analysis of endpoints) for biosimilar comparability exercises may be applied.
- The antibody-testing strategy, including the selection, assessment, and characterization of assays, identification of appropriate sampling time points, sample volumes and sample preparation/storage as well as selection of statistical methods for data analysis should be described in detail. Antibody assays need to be validated for their intended purpose.
- The required observation period for immunogenicity testing should be specified in the manner of allowing observation of clinically significant antibody formation. The period

usually depends on the intended duration of therapy and the expected time of antibody development.

Safety: A comparison of the safety profile between the biosimilar product and the reference product should be made based on the types, incidence and severity of AEs.

Immunogenicity: It is required to make a comparison of the frequency and pattern of antibody formation and clinical effects arising from the immune reactions between the biosimilar product and the reference product before authorization.

If there is an increase in the formation of antibodies against the biosimilar product as compared with the reference product, it would be mandatory to assess its potential effects on the pharmacokinetics, safety and efficacy.

1.3.3. PMDA

Comparative Clinical Studies: Study Design and Analysis

To evaluate the comparability of the efficacy of the follow-on biologic with that of the original biologic, comparative clinical studies should be appropriately designed and justified.

- It is necessary to determine the necessary and adequate number of patients to be enrolled, and pre-specify the margins defining clinical comparability (comparability margin) using clinically established endpoints.
- Where appropriate surrogate endpoints are available, the use of primary endpoints will not always be required. However, the choice of surrogate endpoints should be thoroughly justified on the basis of supportive data or literature, etc.

Although comparability of efficacy has been demonstrated, in certain cases the safety profile of a follow-on biologic may still differ from that of the original biologic. If necessary, clinical studies to evaluate safety, including an immunogenicity study should be considered, even where comparability has been demonstrated through PK, PD or PK/PD studies and further clinical studies to evaluate efficacy are not required. However, when clinical studies are conducted to compare the efficacy of the two products, the studies may be designed such that safety (types of adverse events and their incidence) can be assessed as well.

- If the results of the impurity profile give a rise to particular concerns about safety, the number of patients should be sufficient to perform a thorough investigation of the safety of the follow-on biologic.
- Repeat dose studies on the follow-on biologic should be considered in the case of chronic administration.

Further, at an appropriate stage of the clinical development, studies should be conducted to evaluate antibody formation and other immunogenicity, thus leading to a scientifically justifiable conclusion. Any antibodies detected should be analysed and identified to assess whether the antibodies neutralize the biological activity or not. It is also preferable to analyse the class, affinity and specificity of the antibodies in a scientifically rigorous way. Any reduction in efficacy or impact on safety arising from antibody formation should be considered. It is suggested that antibody formation against impurities or immune responses to specific carbohydrate antigens of the follow-on biologics should also be fully considered.

From: Guideline for the Quality, Safety, and Efficacy Assurance of Follow-on Biologics (https://www.pmda.go.jp/files/000153851.pdf)

1.3.4. FDA, US

Comparative Clinical Studies: Study Design and Analysis

A comparative clinical study will be necessary to support a demonstration of biosimilarity if there is residual uncertainty about whether there are clinically meaningful differences between the proposed product and the reference product.

- Choice of Study Population: Should allow for an assessment of "clinically meaningful differences" between the proposed product and the reference product. Typically, one of the licensed populations, but could be different, if scientifically justified.
- Choice of Endpoint(s): Should be ones that can assess for "clinically meaningful differences," and may be different from the ones used as primary endpoints in the RP's clinical studies if scientifically supported. PD measures are more sensitive than clinical endpoints and, therefore, may enable more precise comparisons of relevant therapeutic effects.
- Sample Size and Duration: Adequate to allow for the detection of "clinically meaningful differences" between the two products. PD measures may facilitate the conduct of a smaller study of limited duration.
- Study Design: Typically, an equivalence design with a symmetric equivalence margin, however, in some cases an asymmetric interval could be reasonable (for example, if the dose used in the clinical study is near the plateau of the dose-response curve and there is little likelihood of dose-related toxicity at higher doses, or if otherwise supported by previous study data. In some cases, a non-inferiority design may be sufficient (for example, the approved doses of the reference product are known to pharmacodynamically saturate the target and it would be unethical to use lower than clinically-approved doses).

Sponsors may propose and provide adequate scientific justification for the choice of study design, study population, study endpoint(s), estimated effect size for the RP and proposed margins. These proposals should be discussed with FDA before initiating the study(ies).

In analysing results, consider what the nature and extent of the residual uncertainty is that remains about biosimilarity based on data from comparative structural and functional characterization, human PK/PD studies, and other information in the application. What aspects of the clinical study would address those uncertainties, and how sensitive the clinical study would be to address the observed differences?

For example, where is the administered dose in the dose-response curve for the drug, for the studied endpoint? Based on the differences observed, would the study population be expected to exhibit any clinical correlates if they were to be present, based on known mechanism(s) of action or other characteristics in that population?

Reference: Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, April 2015 https://www.fda.gov/media/82647/download pages 18-20.

1.3.5. EMA

Comparative Clinical Studies: Study Design and Analysis

In the absence of surrogate markers for efficacy, it is usually necessary to demonstrate comparable clinical efficacy of the biosimilar and the reference medicinal product in adequately powered, randomised, parallel group comparative clinical trial(s), preferably double-blind, by using efficacy endpoints. The study population should generally be representative of approved therapeutic indication(s) of the reference product and be sensitive for detecting potential differences between the biosimilar and the reference. Occasionally, changes in clinical practice may require a deviation from the approved therapeutic indication, e.g. in terms of concomitant medication used in a combination treatment, line of therapy, or severity of the disease. Deviations need to be justified and discussed with regulatory authorities.

In general, an equivalence design should be used. The use of a non-inferiority design may be acceptable if justified on the basis of a strong scientific rationale and taking into consideration the characteristics of the reference product, e.g. safety profile/tolerability, dose range, dose-response relationship. A non-inferiority trial may only be accepted where the possibility of significant and clinically relevant increase in efficacy can be excluded on scientific and mechanistic grounds. However, as in equivalence trials, assay sensitivity has to be considered. It is recommended to discuss the use of a non-inferiority design with regulatory authorities.

CHMP has issued disease-specific guidelines for development of innovative medicinal products. In the development of a biosimilar medicinal product, the choice of clinical endpoints and time points of analysis of endpoints may deviate from the guidance for new active substances. Therefore, comparability should be demonstrated in appropriately sensitive clinical models and study conditions. The applicant should justify that the chosen model is relevant and sensitive to detect potential differences with regard to efficacy and safety. Nevertheless, deviations from endpoints recommended in disease-specific guidelines need to be scientifically justified. Differences detected between the efficacy of the biosimilar and reference products should always be discussed as to whether they are clinically relevant. Generally, the aim of clinical data is to address slight differences observed at previous steps and to confirm comparable clinical performance of the biosimilar and the reference product, not to demonstrate efficacy per se, since this has already been established with the reference product. Clinical data cannot be used to justify substantial differences in quality attributes. The correlation between the “hard” clinical endpoints recommended by the guidelines for new active substances and other clinical/pharmacodynamic endpoints that are more sensitive to detect clinically meaningful differences may have been demonstrated in previous clinical trials with the reference product. In this case, it is not necessary to use the same primary efficacy endpoints as those that were used in the marketing authorisation application of the reference product. However, it is advisable to include some common endpoints (e.g. as secondary endpoints) to facilitate comparisons to the clinical trials conducted with the reference product. Comparability margins should be pre-specified and justified on both statistical and clinical grounds by using the data of the reference product (see ICH topic E9 Statistical principles for clinical trials and CHMP guideline CPMP/EWP/2158/99 on the choice of the non-inferiority margin). As for all comparative clinical trial designs, assay sensitivity (see ICH topic E10) has to be considered.

Clinical safety is important throughout the clinical development programme and is captured during initial PK and/or PD evaluations and also as part of the pivotal clinical efficacy study. Comparative safety data should normally be collected pre-authorisation, their amount depending on the type and severity of safety issues known for the reference product. The duration of safety follow-up

preauthorisation should be justified. Care should be given to compare the type, severity and frequency of the adverse reactions between the biosimilar and the reference product, particularly those described in the SmPC of the reference product. The applicant should provide in the application dossier an evaluation of the specific risks anticipated for the biosimilar. This includes in particular a description of possible safety concerns that may result from a manufacturing process different from that of the reference product, especially those related to infusion-related reactions and immunogenicity.

As regards immunogenicity assessment, applicants should refer to existing CHMP guidance (EMEA/CHMP/BMWP/14327/2006 Rev 1, EMA/CHMP/BMWP/86289/2010). Comparative assessment of unwanted immune responses against the biosimilar and the reference mAb are normally undertaken as part of the clinical study establishing similar clinical efficacy and safety, using the same validated assay(s). A population PK approach with sparse sampling and determination of drug concentration together with anti-drug antibody detection is acceptable. However, for some mAbs, antibodies can be better detected in healthy volunteers, who develop a strong immune response after a single dose within a few days.

Investigation of unwanted immunogenicity is especially important when a different expression system is employed for the biosimilar mAb compared to the reference mAb which might, for example, yield in relevant quality attributes that have not been detected in the reference product (e.g. new post-translational modification structure) that could result in a higher immunogenicity. This is particularly important if there is limited experience with this expression system in humans. It is recommended that such approaches are discussed in advance with regulatory authorities.

3. Clinical Comparability Assessment: Scenarios and Case Studies

1.4. Human PK/PD Studies

1.4.1. Statistical approaches to assess comparability of results from human PK/PD studies in atypical scenarios

When would use of something other than 90% CI of the geometric mean ratio (GMR) of test/reference falling within the 80-125% criteria be appropriate? [1],[2]

Choice of Bioequivalence (BE) Limits

Historically, the bioequivalence standard employed by US FDA is that two products are deemed bioequivalent if the 90% confidence intervals of the geometric mean ratios of the test/reference Cmax and AUC fall within the bioequivalence limits of 80-125%. To obtain geometric means, the data are log-transformed prior to conducting an analysis of variance (ANOVA), then back-transformed before calculating the test/reference ratios. The 2-one-sided tests (TOST) procedure is used to verify that the bioavailability of the test product is not more than 20% less than the reference, and that the

[1] Guidance for Industry: Statistical Approaches to Establishing Bioequivalence. 2001. https://www.fda.gov/media/70958/download

[2] Davit et al., "Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration." Ann Pharmacother 2009; 43:1583-97.

bioavailability of the reference product is not more than 20% less than the test product. This was based on clinical judgment that for drugs that are not narrow therapeutic index drugs, up to a 20% difference in the concentration of the drug in the blood is not likely to be clinically significant. Numerically this is expressed as a limit of 80% on the test mean/reference mean ratios of Cmax and AUC. Since by convention the data are expressed as test/reference ratios, the second statistical test is the reciprocal of 80%, which is 125%. This approach is known as "average bioequivalence."

However, if the true average result of the test product is close to 20% below or 25% above the reference product, one or both of the 90% confidence interval (CI) limits is likely to fall outside the bioequivalence limits. In fact, if the mean Cmax and AUC of the two products truly differ by more than 12 to 13%, they are unlikely to meet the bioequivalence limits.[2] In a retrospective analysis of the results for 2,070 bioequivalence studies over 12 years,[2] the generic drug differed from its reference by less than 10% in C_{max} and $AUC_{0-\tau}$ over 90% of the time (91.5% for C_{max} and 97.6% for $AUC_{0-\tau}$). The actual average percent differences for C_{max}, $AUC_{0-\tau}$, and AUC_{∞} results were close to 4%.

As noted in the Guidance for Industry: Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product[3], to establish PK and/or PD similarity, the calculated confidence interval should fall within an acceptable limit, and the 80-125% criteria for the ratio is an appropriate starting point. However, the confidence interval and acceptable limits can vary among products, and sponsors are advised that justification for use of alternative limits will be expected.

For generic small molecule drugs that have high within-subject variability or have a narrow therapeutic index, FDA has used a reference-scaled average bioequivalence (RSABE) approach to expand or reduce the BE limits.[4,5] However, there are a number of considerations that may be different for a given biologic product that may make RSABE inappropriate or infeasible (e.g., long half-life making the crossover design impractical), and different regulatory authorities also utilize different reference scaling approaches. Therefore, any sponsor considering such an approach should discuss their proposal with the respective regulatory authorities prior to conducting such a study.

Choice of Confidence Interval (CI)

Confidence intervals convey the most likely range of the unknown population average or percentage. This means, for example, a 90% CI covers the true value in 90 of 100 studies performed, and similarly a 95% CI covers the true value in 95 of 100 studies performed. The width of the CI may be affected by the sample size and standard deviation of the groups in the study. The width is also affected by the selected level of confidence, with higher confidence level generally being associated with an increase in the confidence interval width (i.e., a higher confidence that the CI includes the true value comes at a "cost" of a wider total range of values). While a p-value provides information on the strength of evidence against the null hypothesis, it is not useful for determining clinical significance or relevance. The advantage of the CI is that it provides a range of plausible values of the effect size estimate, and thus CI's can be selected based on clinical judgment regarding what a clinically significant or clinically relevant difference in a parameter would be, how precise the estimate needs to be, or the impact of a desired CI level or CI width on the sample size needed for a study.

[3] Guidance for Industry: Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product, https://www.fda.gov/media/88622/download

[4] Davit BM et al., "Implementation of a Reference-Scaled Average Bioqeuivalence Approach for Highly Variable Generic Drug Products by the US Food and Drug Administration." AAPS Journal. December 2012. 14(4). DOI: 10.1208/s12248-012-9406-x.

[5] Yu LX et al., "Novel Bioequivalence Approach for Narrow Therapeutic Index Drugs." CP&T March 2015; 97(3); doi:10.1002/cpt.28.

By convention, efficacy trials, which are often studying the difference between the groups being compared, typically use a p-value of ≤0.05 for statistical significance (i.e., α), and therefore the associated CI (=1-α) is 0.95. As a result, the 95% CI is the most commonly used CI for efficacy trials. For a clinically relevant PD marker that is used as a surrogate efficacy endpoint, it is desirable to utilize a similar level of confidence for the equivalence comparison. However, if the resulting wider CI interval is not acceptable (e.g., because the interval would then be likely to contain values that would be considered clinically meaningful), the resolution will likely entail increasing the sample size for the study and/or decreasing the CI level.[6]

As mentioned above, the average bioequivalence approach typically employs a 90% CI. This is derived from carrying out two one-sided tests (TOST) of hypothesis at the 5% level of significance; i.e., a 5% statistical error is allowed at both the upper and lower bioequivalence limits for a combined total error of 10%.[7,8] As noted by Davit et al. in their publication from 2009[8], FDA adopted the TOST approach after experience with the standard 2-tailed test with P<0.05 demonstrated issues related to the null hypothesis being rejected in scenarios that were inappropriate. For example, products with very small variance in results could show nearly the same means and be rejected as having a statistically significant difference; and products with large variance in results could show a large difference in means that was not statistically-significant and be deemed equivalent. Therefore, if a PD marker has a large or very small variance, use of the average bioequivalence approach using TOST and a 90% CI may be more feasible.

It should be noted that, while the above discussion is background regarding statistical considerations for comparative PK/PD studies and is not product-specific, parts of the discussion may be more applicable to biological products other than the monoclonal antibodies which are the focus of this training manual. For example, clinically relevant PD markers are available for insulin and hematopoietic growth factors but are not available for many monoclonal antibodies.

1.4.2. PK/PD Studies: Is there a need for studies in special populations?

The objective of a well-designed clinical PK and PD study in a biosimilar development program is to evaluate for similarities and differences in the PK and PD profiles of the proposed biosimilar product and the reference product, so the study population selected should be the most informative for detecting and evaluating differences in PK and PD profiles between the proposed biosimilar and the reference product.[9] Keeping in mind that the products are starting off with the same amino acid sequence and should be analytically highly similar, the question becomes, "What types of differences would we be observing that might need testing in the specific setting of a particular population?"

Plasma proteins larger than 60 kDa are not filtered through the kidneys, and proteins between 45 to 60 kDa are filtered only restrictedly.[10] For reference, a monomer of insulin is a combination of two chains and 51 amino acids and has a molecular mass of 5.8 kDa (although insulin is naturally

---

[6] Jia and Lynn, "A sample size planning approach that considers both statistical significance and clinical significance." Trials (2015) 16:213. DOI 10.1186/s13063-015-0727-9.

[7] FDA Guidance for Industry: Statistical Approaches to Establishing Bioequivalence, January 2001. https://www.fda.gov/media/70958/download

[8] Davit et al., "Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration." Ann Pharmacother 2009; 43:1583-97.

[9] "Guidance for Industry: Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product" December 2016. https://www.fda.gov/media/88622/download

[10] Jia L et al. "An Attempt to Understand Kidney's Protein Handling Function by Comparing Plasma and Urine Proteomes." Plos One 2009; 4(4):e5146. Doi:10.1371/journal.pone.0005146.

produced and stored in hexamer form which is 36 kDa).[11] Monoclonal antibodies (~150 kDa) are too large to be filtered by the kidneys. IgG elimination occurs mostly through intracellular catabolism by lysosomal degradation to amino acids after uptake by either pinocytosis or by a receptor-mediated endocytosis process.[12] Many of the inherent characteristics of a monoclonal antibody that would impact pharmacokinetics and pharmacodynamics would be the same for a proposed biosimilar and its reference product, such as its class and subtype (e.g., IgG1) and its target. There are also heterogeneous characteristics such as those that result from post-translational modification (e.g., charge-which can result in changes in tissue distribution; glycosylation patterns—such as agalactosylation or afucosylation or presence of high mannose forms), where differences between the products could result in uncertainty about whether there are clinically meaningful impacts on exposure (pharmacokinetics) or other functional implications. Again, clinical pharmacology studies should be conducted in the subject or patient demographic group most likely to provide a sensitive measure of the clinical impact of the potential differences between the proposed biosimilar and reference product. That population would typically be healthy volunteers, if it is feasible to safely administer the biologic to them, as they would be expected to have less variability and fewer confounding factors.[13] Comparative PK/PD studies in special populations (pediatric, pregnant, geriatric) would not be expected to be additionally informative, and the results observed for the reference product in these populations would be expected to be similar for the biosimilar.[14]

1.5. Comparative Clinical Efficacy and Safety Studies

1.5.1. Statistical analysis methods to assess comparability of results from comparative clinical studies

The intent of comparative clinical studies in biosimilar development programs is fundamentally different than for clinical studies using an active comparator that are intended to demonstrate the efficacy of a novel molecule. In the latter setting, one is not relying on the similarity or sameness of the molecules, but rather just using the active comparator's known effect to prove that the novel molecule also has a treatment effect. On the other hand, comparative clinical studies in biosimilar development programs are not intended to demonstrate efficacy but rather to support a conclusion that there are no clinically meaningful differences between the proposed biosimilar and the reference product, in the setting where we are relying on analytical data as well that demonstrate high similarity.

[11] Weiss M, Steiner DF, Philipson LH, "Insulin Biosynthesis, Secretion, Structure, and Structure-Activity Relationships." Endotext – NCBI Bookshelf, updated February 1, 2014.
[12] Ryman JT and Meibohm B. "Pharmacokinetics of Monoclonal Antibodies" CPT Pharmacometrics Syst. Pharmacol. (2017) 6:576-588. Doi:10.1002/psp4.12224.
[13] "Guidance for Industry: Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product" December 2016. https://www.fda.gov/media/88622/download
[14] However, note that biosimilar products proposed for marketing in the U.S. are required by the BPCI Act to address the requirements of the Pediatric Research Equity Act. For more information, see the response to Q.I.16 in the Draft Guidance for Industry: New and Revised Draft Q&As on Biosimilar Development and the BPCI Act (Revision 2), December 2018. https://www.fda.gov/media/119278/download

Therefore, while ICH E9[15] principles regarding basic equivalence or non-inferiority trial design and analysis are relevant, for comparative clinical efficacy studies in biosimilar programs, use of a different or modified endpoint or a different study duration, compared with those from the historical reference product efficacy trials, may be justified in order to optimize a study's ability to serve as a sensitive and relevant assay to demonstrate "no clinically meaningful differences." The study design features and outcome variables chosen, can greatly impact the likelihood of study success and will depend on the details of a given situation.

Friedrich et al.[16,17], in the context of meta-analyses, studied the impact of the structure of the outcome variable on overall study results. A common practice with meta-analyses has been to use difference of means (DOM) for continuous outcomes, using mean difference (MD) or standardized mean difference (SMD). Binary outcomes might be presented using either difference (risk difference) and ratio (risk ratio and odds ratio) methods. The authors used ratio of means (ROM) measures and DOM measures for re-analysing 232 meta-analyses pooling continuous outcome measures and noted that similar treatment effect estimates and heterogeneity were obtained whether ROM or DOM was used to describe the continuous outcome. They concluded, therefore, that the choice between difference and ratio outcomes for continuous variables should be determined by other factors, such as the biological effect of the treatment as either additive (favors DOM) or relative/multiplicative (favors ROM) for different control group values, in addition to statistical properties for a given situation.

Sun et al.[18] from FDA's Office of Biostatistics describes these considerations and recommendations in in the specific context of clinical endpoint bioequivalence or non-inferiority trials, as per Table 2 below. As noted in Table 2 and also mentioned by Friedrich et al., a foundational question is whether the inherent properties of the biological effect to be measured are additive or multiplicative. Effects that are additive would favor using a difference of means. In contrast, for effects that are multiplicative, a ratio of means may be a more relevant or interpretable variable to assess because the clinical meaningfulness of a given absolute difference could vary significantly based on where on the response curve it is falling. Is the size of the treatment effect for the outcome expected to be very small (i.e. close to zero)? If so, then a difference of means outcome variable may be more feasible, because very small values in the denominator of the ratio can have a disproportionate impact on the ratio. Notably, there are a number of potential ramifications (mainly related to change in power of the study) associated with the choice of DOM vs. ROM measures if data transformations are used, because these may affect the distribution of the outcome data significantly. For example, a change in location (e.g. reversing the scale so that what reflects response or worsening changes to the opposite direction or using a constant) or scale (e.g., 0 to 10 vs 0 to 100) can affect the power of the study and may require a change in the pre-specified margin.

[15] Statistical Principles for Clinical Trials, E9. Version Step 4, February 1998. Pages 14-15. https://database.ich.org/sites/default/files/E9_Guideline.pdf

[16] Friedrich JO et al. "The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study." BMC Medical Research Methodology 2008, 8:32; doi:10.1186/1471-2288-8-32.

[17] Friedrich JO et al. "Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods." J Clin Epi; 2011, 64:556-564. Doi: 10.1016/j.jclinepi.2010.09.016.

[18] Sun W et al. "Ratio of means vs difference of means as measures of superiority, noninferiority, and average bioequivalence." 2017, J Biopharm Stat. 27(2):338-355. http://dx.doi.org/10.1080/10543406.2016.1265536

Table 2: Summary of recommendations for choice between difference of means (DOM) and ratio of means (ROM) for superiority, non-inferiority (NI) and average bioequivalence (ABE) tests

	Measure			
	Difference of Means (DOM)		Ratio of Means (ROM)	
Condition	Superiority	NI or ABE	Superiority	NI or ABE
(1) Biological effect of the treatment for difference control group values	Additive		Multiplicative	
(2) Units of the outcome of interested	Need to be identical units		Can be different units	
(3) Mean value of reference group	Does not matter		Cannot be very small (e.g., close to zero)	
(4) Sign of means of two treatment groups	Does not matter		Needs to be both positive or both negative Rothmann et al. (2012)	
(5) Direction of the scoring system for the outcome of interested	Does not matter		It doesn't matter because ROM and DOM are identical for superiority.	Power can change significantly as the direction of the scores changes.
(6) Shift in scoring systems with the same margins				
Location shift	Does not matter	Does not matter	Does not matter	Power can change
Scale shift with a positive scale factor	Does not matter	Power can change	Does not matter	Does not matter
Scale shift with a negative scale factor	Does not matter	Power can change	Does not matter	Power can change
Combined location & scale shift	Does not matter	Power can change	Does not matter	Power can change
Other shift (non-location or -scale)	Power can change	Power can change	Power can change	Power can change

Source: Table 3, Sun et al., J Biopharm Stat. 2017, 27(2):338-355. http://dx.doi.org/10.1080/10543406.2016.1265536

Sun et al. illustrates these issues using a case study of an adhesion study submitted in an abbreviated new drug application (ANDA) for a proposed generic patch to an extremely well-adhering reference drug patch (see Table 3 below). This case study illustrates the differing NI test results when transforming the data using a scale shift (multiplying by 10), and location shifts by reversing the scale, adding 1, or adding 100.

This case also illustrates the problem of a ROM test in cases where the reference drug has very large response (in this case, close to perfect adhesion). Despite very good adhesion with the test patch, the study had very low power to pass NI using the ROM NI test using the 95% CI lower bound of the "standard" BE criteria (i.e. 80%). This issue led FDA to revise recommendations in an updated draft guidance on adhesion studies, replacing the original ROM NI test with a DOM NI test and an NI margin of 0.15 if FDA's scoring system (0 to 4) is used. These considerations have direct practical implications for biosimilar comparative clinical studies in which use of the 80-125% BE criteria is contemplated for a given clinical outcome measure.

Table 3: Noninferiority test results for an ANDA patch adhesion study and the impact of different scoring systems under ROM or DOM measures

Scoring Systems (the smaller, the better)	TEST LSmean (Std. error)	RLD LSmean (Std. error)	ROM H_1: $\mu_T/\mu_R<1.25$		DOM H_1 : $\mu_T - \mu_R<0.15$	
			95%UB of $\mu_T - 1.25\mu_R$	Pass or Fail NI	95%UB of $\mu_T - \mu_R$	Pass or Fail NI
Original: (0, 1, 2, 3, 4)	0.04 (0.01)	0.03 (0.01)	0.02 > 0	Fail NI	0.03 <0.15	**Pass NI**
Scale shift: *10 (0, 10, 20, 30, 40)	0.40 (0.12)	0.30 (0.09)	0.20 > 0	Fail NI	0.27 > 0.15	Fail NI
Location shift: +1 (1, 2, 3, 4, 5)	1.04 (0.01)	1.03 (0.01)	-0.23 < 0	**Pass NI**	0.03 <0.15	**Pass NI**
Location shift: +100 (100, 101, 102, 103, 104)	100.04 (.01)	100.03 (0.01)	-24.98 < 0	**Pass NI**	0.03 < 0.15	**Pass NI**
Scoring Systems (the larger, the better)	TEST LSmean (Std. error)	RLD LSmean (Std. error)	ROM H_1 : $\mu_T/\mu_R>0.8$		DOM H_1 : $\mu_T - \mu_R> -0.15$	
			95% LB of $\mu_T - 0.8\mu_R$	Pass or Fail NI	95%UB of $\mu_T - \mu_R$	Pass or Fail NI
Location shift: reversing (4, 3, 2, 1, 0)	3.96 (0.01)	3.97 (0.01)	0.77 > 0	**Pass NI**	-0.03> -0.15	**Pass NI**

Source: Table 2, Sun et al., J Biopharm Stat. 2017, 27(2):338-355. http://dx.doi.org/10.1080/10543406.2016.1265536

In summary, the historical reference product efficacy trial design features and outcomes may or may not be the optimal choices for a comparative clinical study in a biosimilar development program. The choice between DOM and ROM outcome variables for a study should take into consideration a number of factors (as listed in Table 2 above), most of which are statistical and would be best evaluated by a biostatistician.

Case Study: Choice of Endpoint, Margin and Confidence Interval:

MVASI (ABP215, bevacizumab-awwb), Advisory Committee Meeting July 13, 2017[19,20,21]

MVASI is a biosimilar to U.S.-licensed Avastin approved on September 14, 2017. The clinical program for MVASI included a comparative clinical study (Study 20120265; NCT01966003) with an equivalence design that utilized a margin with an asymmetric interval.

Overall Response Rate (ORR) was accepted by FDA, United States as the primary endpoint for the comparative clinical study because it is a consistent measure of the treatment effect in this clinical setting and as it is not altered by subsequent therapy, as may be the case for overall survival. FDA chose to use the ratio of the ORR relative risk (RR) to characterize the difference between ABP215 and E.U.-approved bevacizumab used as the comparator after a scientific bridge including U.S.-licensed Avastin was completed. In FDA's determination of the similarity margin to be used for the ratio of RR, data from the published results of four randomized studies were included in a meta-analysis to evaluate the treatment effect of bevacizumab in combination with platinum-doublet chemotherapy in the first-line treatment of patients with nonsquamous non-small cell lung cancer (NSCLC) to rederive the similarity margin. The four studies were Study AVF0757 (Johnson et al, 2004[22]), Study E4599 (Sandler et al., 2006[23]), Study JO19907 (Niho et al.,2009[24]), and the AVAIL study (Reck et al., 2010[25]). The control arm in three of the studies was paclitaxel plus carboplatin. Cisplatin plus gemcitabine was used in the AVAIL study. As shown in Table 2, there was a consistent bevacizumab treatment effect on ORR risk ratio, ranging from 0.43 to 0.63 in all four studies, which justified the selection of the first-line treatment of nonsquamous NSCLC population as adequately sensitive to support a demonstration of no clinically meaningful differences between MVASI and US-licensed Avastin.

[19] FDA, United States Advisory Committee Proceedings, July 13, 2017 FDA Briefing Information p. 43-46 https://www.fda.gov/advisory-committees/oncologic-drugs-advisory-committee/2017-meeting-materials-oncologic-drugs-advisory-committee

[20] He et al. "Statistical Considerations in Evaluating a Biosimilar Product in an Oncology Clinical Study." CCR 2016; 22:5167-69.

[21] Casak et al. "FDA's Approval of the First Biosimilar to Bevacizumab." CCR 2018; 24:4365-4370.

[22] Johnson 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; 22:2184–91.

[23] Sandler et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 2006; 355:2542–50.

[24] Niho et al. Randomized phase II study of first-line carboplatin-paclitaxel with or without bevacizumab in Japanese patients with advanced non-squamous non-small-cell lung cancer. Lung Cancer 2012; 76:362–7.

[25] Reck 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). Ann Oncol 2010; 21:1804–9.

Table 4: FDA Analysis/Estimation of Historical Treatment Effect in ORR in the Four Randomized Bevacizumab NSCLC Trials

Author	Study	Chemotherapy ORR (CR+PR)/N* (%)	Bevacizumab + Chemotherapy ORR (CR+PR)/N* (%)	Risk Ratio	70% CI
Johnson et al. (2004[4])	AVF9757	6/32 (18.8)	11/34 (32.4)	0.58	(0.37, 0.92)
Sandler et al. (2006[5])	E4599	59/392	133/381 (34.9)	0.43	(0.37, 0.50)
Nishio et al. (2009[6])	JO19907	20/59 (33.9)	68/121 (56.2)	0.60	(0.49, 0.74)
Reck et al. (2010[7])	AVAIL	71/327 (21.7)	114/329 (34.7)	0.63	(0.55, 0.72)
He, et al. 2016	Meta-analysis	156/810 (19.3)	326/865 (37.7)	0.53	(0.49, 0.58)

*N: ITT population; ORR: Overall Response Rate, CR: Complete Response, PR: Partial Response

The meta-analysis performed by FDA, United States of the four clinical studies described in Table 4 yields a RR of 0.53 and an ORR for bevacizumab with chemotherapy of 38%. Based on these assumptions, and assuming a 10% drop-out rate using a symmetrical similarity margin in the comparative clinical study, the sample sizes needed for a comparative clinical study were calculated with 80%, 85%, or 90% power using a range of 70% to 95% CI for the observed ORR in the meta-analysis (Table 5).

Table 5: ORR Sample Size and Confidence Interval Consideration Based on FDA Meta-Analysis of NSCLC Studies

Confidence Coefficient %	CI for ORR from Meta-analysis	Similarity Margin Based on Maintaining 50% of CL	Sample Size		
			80% power	85% power	90% power
70	0.49, 0.58)	(0.73, 1.36)	608	683	768
75	(0.49, 0.59)	(0.74, 1.35)	632	711	799
80	(0.48, 0.60)	(0.75, 1.34)	662	744	837
85	0.47, 0.60)	(0.75, 1.33)	702	789	887
90	(0.47, 0.61	0.76, 1.31)	758	852	958
95	(0.45, 0.63)	0.77, 1.29)	856	962	1082

CI: Confidence Interval, CL: Confidence level

The confidence intervals in the second column of the table above consider the variability of the bevacizumab treatment effect from the meta-analysis. Since a proposed biosimilar product should be highly similar to the reference product based on analytical and PK data, using a 70% CI of the estimated effect on ORR was considered sufficient to support a demonstration that no clinically meaningful differences exist with a feasible corresponding sample size.

1.5.2. Choice of primary endpoints or choice of study/indication population for comparative clinical studies

The choice of primary endpoint in comparative clinical studies ideally would be based on which endpoint and study population would be most sensitive for detecting differences between the test and reference product, if differences exist. Considerations for choice of endpoint and study population may include but are not limited to[26]:

- observed quality attribute differences and uncertainty regarding the impact on different mechanism(s) of action related to a specific population and endpoint.
- The population with the dose/regimen that would be most sensitive to detect differences.
- If there are disease-specific safety or immunogenicity characteristics in a certain population that would make a study in that population more informative.
- Whether the efficacy endpoint used for a certain population would be more sensitive for detecting differences.

Additionally, although the endpoint does not necessarily need to be the same one used in the historical reference product trials, ideally it should be one for which there is a good basis for knowing the effect of the reference product.

Case Study: Considerations on Choice of Endpoint and Population:

Truxima (CT-P10, rituximab-abbs), Advisory Committee Meeting October 10, 2018[27]

Truxima is a biosimilar to U.S.-licensed Rituxan approved on November 28, 2018. The clinical program for Truxima included a comparative clinical study (Study CT-P10 3.4; NCT02260804) that enrolled patients with Low Tumor Burden Follicular Lymphoma (LTBFL) in order increase the sensitivity to detect differences between products that could be influenced by the presence of concurrent background chemotherapy. This study population was justified to be more sensitive than a population with Advanced Follicular Lymphoma and was scientifically justified as follows:

- *The use of monotherapy in treatment-naïve LTBFL settings eliminates the potential impact of chemotherapy in the assessment of efficacy, PK/PD, safety and immunogenicity.*
- *The effect size is sufficiently large in LTBFL patients, allowing for the detection of clinically meaningful differences between a proposed biosimilar and a reference product in a comparative clinical study. A large effect size for ORR, the primary endpoint was previously reported (Ardeshna et al., 2014[28]).*
- *U.S.-licensed Rituxan, as a first-line single-agent therapy for previously untreated patients with indolent NHL including LTBFL, has been highly active and well tolerated and is an accepted treatment modality in follicular lymphoma.*
- *Single-agent U.S.-licensed Rituxan treatment, consisting of U.S.-licensed Rituxan induction and maintenance led to a significant increase of the time to commencement of the new treatment and higher improvements in quality of life (QoL) compared to watchful waiting.*

[26] Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, April 2015 https://www.fda.gov/media/82647/download

[27] FDA, United States Advisory Committee Proceedings, October 10, 2018. FDA Presentation slide 40, Celltrion Briefing Information, p.79. https://www.fda.gov/advisory-committees/advisory-committee-calendar/meeting-oncologic-drugs-advisory-committee-10102018-10102018

[28] Ardeshna et al. Rituximab versus a watch-and-wait approach in patients with advanced-stage, asymptomatic, non-bulky follicular lymphoma: an open-label randomised phase 3 trial. Lancet Oncol 2014; 15:424-35.

- *As tumor burden, B-cell microenvironment, and Fc receptor binding may impact response, LTBFL is sensitive in detecting any potential clinically meaningful differences in therapeutic effect.*

1.5.3. What is scientifically needed to make the data of a clinical study performed with one regulatory authority's approved RP relevant/interpretable for an application in a second country, which may or may not have the same version of the RP?

MFDS

Reference: 'Questions and Answers on the Biosimilar Products Ⅰ" Dec 2018. General Chapter, Q10 '

- MFDS accepts Non-Korean reference product if a sponsor provides appropriate bridging data to Korean reference product. The bridging study should include:
  - Physicochemical and biological comparability data with 3 batches of each Korean, non-Korean reference product and biosimilar candidate product.
  - If necessary, comparative PK/PD study may be requested.

US FDA

Refer to the "Guidance for Industry: Questions and Answers on Biosimilar Development and the BPCI Act" [29] December 2018, in the response to Q.I.8.:

- Analytical data that compares all three products (proposed biosimilar, non-US licensed comparator, and US-licensed reference product), likely to include, but not limited to:
  - Comparative physicochemical characterization
  - Biological assays/functional assays
  - Degradation profiles under stressed conditions
- Comparative clinical PK/PD (when appropriate)
- Information to address any other factors that may affect the relevance of the comparative data with the non-US licensed comparator product to an assessment of biosimilarity, and to establish an acceptable bridge to the US licensed reference product. This may include:
  - The number of comparator and reference products assessed analytically, whether that is adequate to capture the variability in product quality attributes, and the relationship of those non-US-comparator lots to those used in the clinical studies.

EMA

A single reference medicinal product, defined on the basis of its marketing authorisation in the EEA, should be used as the comparator throughout the comparability programme for quality, safety and efficacy studies during the development of a biosimilar in order to allow the generation of coherent data and conclusions.

However, with the aim of facilitating the global development of biosimilars and to avoid unnecessary repetition of clinical trials, it may be possible for an Applicant to compare the biosimilar in certain clinical studies and in *in vivo* non-clinical studies (where needed) with a non-EEA authorised comparator (i.e. a non-EEA authorised version of the reference medicinal product) which will need to

[29] Guidance for Industry: Questions and Answers on Biosimilar Development and the BPCI Act, December 2018. https://www.fda.gov/media/119258/download

be authorised by a regulatory authority with similar scientific and regulatory standards as EMA (e.g. ICH countries). In addition, it will be the Applicant's responsibility to demonstrate that the comparator authorised outside the EEA is representative of the reference product authorised in the EEA.

For demonstration of biosimilar comparability at the quality level, side-by-side analysis of the biosimilar product (from commercial scale and site) with EEA authorised reference product must be conducted. However, combined use of non-EEA authorised comparator and EEA authorised reference product is acceptable for the development of the Quality Target Product Profile of the biosimilar product.

If certain clinical and *in vivo* non-clinical studies of the development programme are performed with the non-EEA authorised comparator, the Applicant should provide adequate data or information to scientifically justify the relevance of these comparative data and establish an acceptable bridge to the EEA-authorised reference product As a scientific matter, the type of bridging data needed will always include data from analytical studies (e.g., structural and functional data) that compare all three products (the proposed biosimilar, the EEA-authorised reference product and the non-EEA-authorised comparator), and may also include data from clinical PK and/or PD bridging studies for all three products. The overall acceptability of such an approach and the type of bridging data needed will be a case-by-case/product-type decision and is recommended to be discussed upfront with the Regulatory Authorities. However, the final determination of the adequacy of the scientific justification and bridge will only be made during the assessment of the application.

Case Study: Considerations on Bridging Between Reference Product Approved in a Different Jurisdiction:

Erelzi (GP2015, etanercept-szzs), Advisory Committee Meeting July 13, 2016[30]

*Etanercept is a dimeric fusion protein comprised of the extracellular domain of human TNF receptor linked to the Fc region of IgG1. It is complex and heavily glycosylated with both N- and O-linked oligosaccharides. Furthermore, Amgen and Pfizer both manufacture etanercept, with Pfizer being the market authorization holder (MAH) in the EU and the rest of the world, excluding the US and Canada, where Amgen is the license/application holder for Enbrel.[31] The U.S. application for GP2015 contained a clinical study (Study 302) comparing GP2015 and **EU-approved Enbrel** to support the demonstration of "no clinically meaningful differences." Analytical data were provided to support a scientific bridge between GP2015, US-licensed Enbrel and EU-approved Enbrel. As noted in the draft Guidance for Industry: Development of Therapeutic Protein Biosimilars: Comparative Analytical Assessment and Other Quality-Related Considerations, May 2019,[32] all pairwise comparisons are analysed. As noted in the Guidance for Industry: Questions and Answers on Biosimilar Development and the BPCI Act, December 2018, [33] the type of bridging data is likely to also include bridging clinical PK and/or PD study data for all three products, and all three pairwise comparisons would generally be expected to meet pre-specified acceptance criteria.*

[30] US FDA Arthritis Advisory Committee Meeting, July 13, 2016: https://www.fda.gov/advisory-committees/arthritis-advisory-committee/2016-meeting-materials-arthritis-advisory-committee

[31] Hassett B et al., "Manufacturing history of etanercept (Enbrel): consistency of product quality through major process revisions." MABS 2018, 10(1):159-165. https://doi.org/10.1080/19420862.2017.1388483

[32] Draft Guidance for Industry: Development of Therapeutic Protein Biosimilars Comparative Analytical Assessment and Other Quality-Related Considerations, May 2019. https://www.fda.gov/media/125484/download

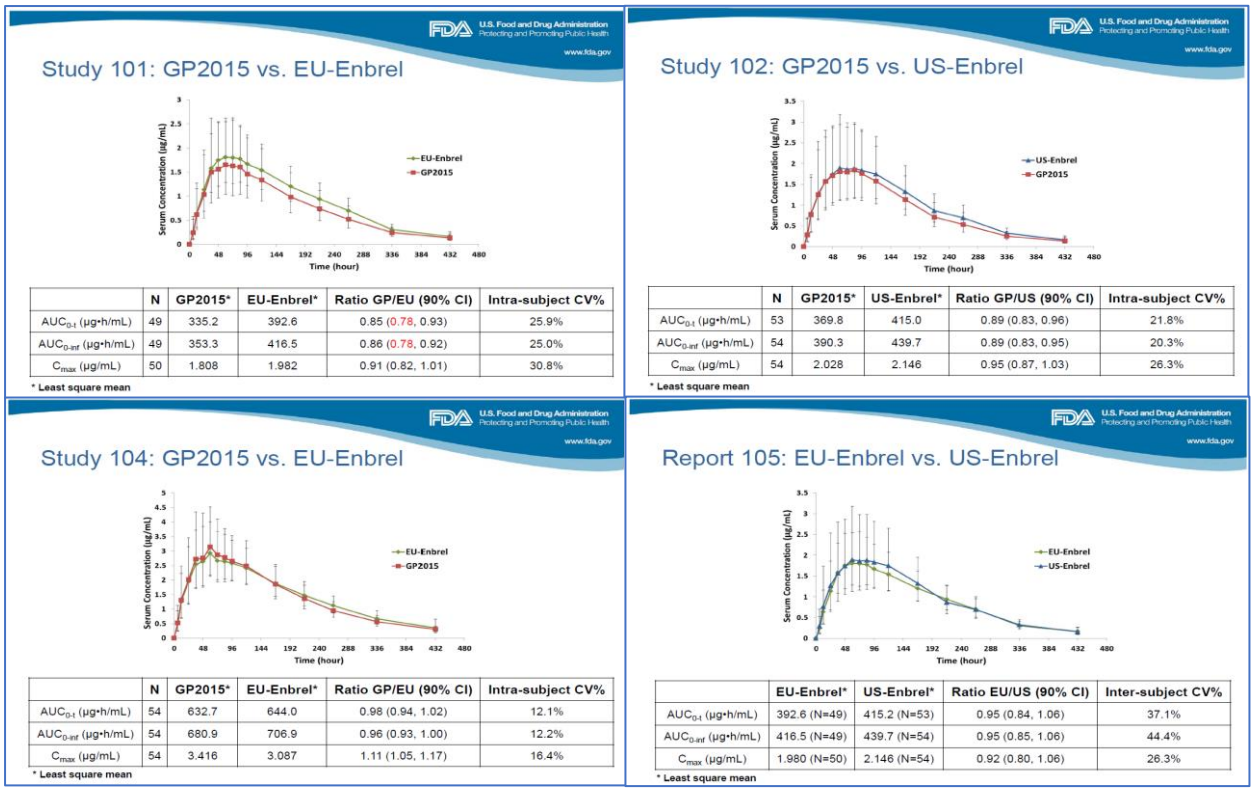
[33] Guidance for Industry: Questions and Answers on Biosimilar Development and the BPCI Act, December 2018. https://www.fda.gov/media/119258/download

Table 6: Rationale for the Three Pairwise Comparisons

Pairwise Comparison	***Analytical Rationale***	***PK/PD Rationale***
GP2015 vs US-licensed Enbrel	***Required to be highly similar to the US reference product for U.S. approval***	***Supports a demonstration of "no clinically meaningful differences" compared to the US reference product***
US-licensed Enbrel vs EU-approved Enbrel	*Analytical comparison is needed to address uncertainty about whether there are any differences between the US-licensed Enbrel and EU-approved Enbrel that might impair the ability of the data acquired with EU-approved Enbrel to support the required conclusions about the GP2015 vs US-licensed Enbrel comparisons.*	*Pairwise comparisons of PK/PD data were requested in this case for a similar reason as the analytical rationale. However, in some instances, pairwise comparison for PK/PD data may not be informative or additionally helpful for this purpose (for example, because there is not residual uncertainty from the analytical data). This should be discussed ahead of time and a scientific justification provided.*
EU-approved Enbrel vs GP2015	*Crosscheck to provide analytical context to the comparative clinical study comparison. GP2015 serves as the anchor and sufficient similarity of GP2015 to EU-Enbrel in this comparison, in addition to the primary comparison to US-Enbrel above, supports a conclusion that clinical data from the study comparing GP2015 with EU-approved Enbrel is likely to be representative of results that would have been obtained with US-reference product.*	*Similar crosscheck as noted in the analytical data rationale, but for PK/PD data. Also see comment directly above.*

In the application, adequate analytical data were provided to establish that GP2015 was highly similar to US-licensed Enbrel, and an adequate analytical bridge was established between GP2015, US-licensed Enbrel, and EU-approved Enbrel. However, there were not 3-way PK data from a single study containing all three products, but rather 2 separate studies were originally conducted: GP2015 vs. EU approved Enbrel (Study 101) and GP2015 vs. US-licensed Enbrel (Study 102). The US-licensed Enbrel and EU-approved Enbrel comparison was provided as a cross-study analysis of results for the reference products in Studies 101 and 102. This approach was further complicated by GP2015 not meeting PK acceptance criteria vs EU-Enbrel in Study 101 (see Figure 1 below), missing the lower bound of the 90% CI for the ratio of the $AUC_{0\text{-}t}$ and $AUC_{0\text{-}inf}$ parameters. If GP2015 were truly lower in exposure compared to EU-Enbrel (but not lower than US-Enbrel, according to Study 102), then this raised uncertainty regarding the equivalent results in the comparative clinical study, Study 302, in moderate to severe chronic plaque psoriasis patients. Study 302 used EU-Enbrel: if GP2015 showed no clinically meaningful differences to EU-Enbrel in this study, but did so at a lower exposure, could this suggest GP2015 was more potent, and if it were compared to US Enbrel, would there be a clinically meaningful difference?

Figure 1: Pairwise Comparisons of GP2015, EU-Enbrel, and US-Enbrel



	N	GP2015*	EU-Enbrel*	Ratio GP/EU (90% CI)	Intra-subject CV%
AUC_{0-t} (µg•h/mL)	49	335.2	392.6	0.85 (0.78, 0.93)	25.9%
AUC_{0-inf} (µg•h/mL)	49	353.3	416.5	0.86 (0.78, 0.92)	25.0%
C_{max} (µg/mL)	50	1.808	1.982	0.91 (0.82, 1.01)	30.8%

* Least square mean

	N	GP2015*	US-Enbrel*	Ratio GP/US (90% CI)	Intra-subject CV%
AUC_{0-t} (µg•h/mL)	53	369.8	415.0	0.89 (0.83, 0.96)	21.8%
AUC_{0-inf} (µg•h/mL)	54	390.3	439.7	0.89 (0.83, 0.95)	20.3%
C_{max} (µg/mL)	54	2.028	2.146	0.95 (0.87, 1.03)	26.3%

* Least square mean

	N	GP2015*	EU-Enbrel*	Ratio GP/EU (90% CI)	Intra-subject CV%
AUC_{0-t} (µg•h/mL)	54	632.7	644.0	0.98 (0.94, 1.02)	12.1%
AUC_{0-inf} (µg•h/mL)	54	680.9	706.9	0.96 (0.93, 1.00)	12.2%
C_{max} (µg/mL)	54	3.416	3.087	1.11 (1.05, 1.17)	16.4%

* Least square mean

	EU-Enbrel*	US-Enbrel*	Ratio EU/US (90% CI)	Inter-subject CV%
AUC_{0-t} (µg•h/mL)	392.6 (N=49)	415.2 (N=53)	0.95 (0.84, 1.06)	37.1%
AUC_{0-inf} (µg•h/mL)	416.5 (N=49)	439.7 (N=54)	0.95 (0.85, 1.06)	44.4%
C_{max} (µg/mL)	1.980 (N=50)	2.146 (N=54)	0.92 (0.80, 1.06)	26.3%

* Least square mean

Ultimately, it was concluded that there were sufficient other data to conclude that the data from Study 302 was relevant to support conclusions that there were no clinically meaningful differences between GP2015 and US-licensed Enbrel. This was based on:

- *Analytical bridging data that did not demonstrate any analytical differences between GP2015, EU-approved Enbrel, or US-licensed Enbrel that would raise uncertainty regarding clinically meaningful differences in PK, potency, or efficacy.*
- *Geometric mean trough serum concentration-time profiles of GP2015 and EU-approved Enbrel were comparable at steady state in Study 302.*
- *Cross-study comparison of EU-Enbrel and US-Enbrel suggested no clinically meaningful differences in PK parameters between them, thus supporting the relevance of the results of EU-Enbrel from Study 302.*
- *Results of Study 101 were not confirmed in a repeat PK study of GP2015 and EU-Enbrel (Study 104).*

1.5.4. What are the factors to consider when deciding whether the data provided are adequate to support a clinical indication of the RP that has not been directly studied (sometimes known as "extrapolation.")?

MFDS

Reference: 'Guidelines on the Evaluation of Biosimilar Products, English version, Revision 1'

If similar quality, efficacy and safety of the biosimilar product and the reference product have been demonstrated for a particular therapeutic indication, extrapolation of these data to other indications

of the reference product for which post-marketing surveillance was completed in Korean market could be possible if all of the following conditions are fulfilled:

- A sensitive clinical test model has been used that is able to detect potential differences between the biosimilar product and the reference product;
- The clinically relevant mechanism of action and/or involved receptor(s) are the same;
- Safety and immunogenicity have been sufficiently characterized.

Refer to Guidelines on the Evaluation of Biosimilar Products, English version, Revision 1 (http://www.mfds.go.kr/eng/wpge/m_37/de011024l001.do)

> ***Example: Truxima (CT-P10, Rituximab)***
> *Truxima was developed as the biosimilar product of MabThera. Mechanism of action is to disrupt B cells by antibody dependent cell cytotoxicity (ADCC) of the rituximab after binding to CD20 expressed B cells. Mabthera was approved for 2 categories of therapeutic indications, which are blood cancer (lymphoma, chronic lymphocytic leukemia) and immune modulatory disease (rheumatoid arthritis, Wegener's granulomatosis and microscopic polyangiitis).*
>
> *Physicochemical and biological activity of Truxima showed comparable to the reference product with minor differences in glycosylation and product-related variants. Sponsor provided complementary data that the minor differences effects little on biological activity. In addition, non-clinical and clinical data fall under equivalence margin of the studies. It was concluded that Truxima is similar to Mabthera with respect to the totality of evidence.*
>
> *In clinical studies, pharmacokinetics data in rheumatoid arthritis (RA) patients and 2 phase 3 clinical trials data for RA and for advanced follicular lymphoma patients were provided, and all data shown in clinical trials, PK equivalence in RA patients, clinical equivalence in RA and AFL patients was demonstrated. The extrapolation of all indications was accepted based on the conclusion that mechanism of action and the binding receptor are same for two categories of indications, sensitive clinical model was used, and no considerable safety and immunogenicity issues were observed.*

US FDA

Considerations include:

- **The mechanism of action (MOA) in each condition of use for which licensure is sought vs the MOA in the condition studied.** *For example, as shown in Table 7 below and described in the Inflectra case study below, of the conditions of use of Remicade, the Crohn's Disease (CD) and Ulcerative Colitis (UC) indications were the only ones thought to possibly use additional mechanisms of action other than blocking of TNFR1 and TNFR2 activity. The small difference in ADCC and FcγRIIIa binding observed with Inflectra would not have raised the potential concern about clinically meaningful differences in the RA, AS, PsA and PsO indications, but could not be ruled out for CD and UC. While this might have made CD or UC patients more important to study, there were other factors such as study duration and endpoints that made these CD or UC trials impractical (see Inflectra case study below).*

Table 7: Known and Potential Mechanisms of Action of Remicade in the US-Licensed Conditions of Use

MOA of Remicade	RA	AS	PsA	PsO	CD, Pediatric CD	UC, Pediatric UC
Mechanisms involving the Fab (antigen binding) region:						
Blocking TNFR1 and TNFR2 activity via binding and neutralization of s/tmTNF	Known	Known	Known	Known	Likely	Likely
Reverse (outside-to-inside) signaling via binding to tmTNF:	-	-	-	-	Likely	Likely
Apoptosis of lamina propria activated T cells	-	-	-	-	Likely	Likely
Suppression of cytokine secretion	-	-	-	-	Likely	Likely
Mechanisms involving the Fc (constant) region:						
Induction of CDC on tmTNF-expressing target cells (via C1q binding)	-	-	-	-	Plausible	Plausible
Induction of ADCC on tmTNF-expressing target cells (via FcγRIIIa binding expressed on effector cells)	-	-	-	-	Plausible	Plausible
Induction of regulatory macrophages in mucosal healing	-	-	-	-	Plausible	Plausible

ADCC: antibody-dependent cellular cytotoxicity; AS: ankylosing spondylitis; CD: Crohn's Disease; CDC: complement-dependent cytotoxicity; MOA: mechanism of action; PsA: psoriatic arthritis; PsO: plaque psoriasis; RA: rheumatoid arthritis; UC: ulcerative colitis; sTNF: soluble TNF; tmTNF: transmembrane TNF

Source: FDA Briefing Document, Arthritis Advisory Committee Meeting for Inflectra, Feb.9, 2016.

- **The pharmacokinetics, pharmacodynamics and bio-distribution of the product in the different patient populations.**
  - *There are a number of physiologic differences in infants compared to older children and adults which results in faster plasma clearance. For example, total body volume available for distribution is higher in infants, and there is also a faster rate of blood perfusion to the tissues in infants up to 6 months of age when compared to adults.*[34] *However, unless there were a difference in product or formulation characteristics between the biosimilar and its reference product that might result in a clinically meaningful difference in exposure likely being specifically revealed only in infants, a study in infants would not be warranted.*
- **The immunogenicity of the product in different patient populations.**
  - *Is the reference product known to demonstrate higher immunogenicity in a certain patient population? Then this population may be a more sensitive test of differences in immunogenicity. Is a population more likely to be immunosuppressed and be less likely to react to immunogenic proteins? Then this population would likely be a less sensitive test of differences in immunogenicity.*
- Toxicities in each condition of use and patient population and whether any differences expected would be related to the pharmacologic activity or off-target activities of the product.
- Any other factor that may affect the safety or efficacy of the product in each condition of use and patient population for which licensure is sought.

In choosing which condition of use to study, FDA recommends that a sponsor consider choosing a condition of use that would be adequately sensitive to detect clinically meaningful differences

[34] Malik P, Edginton A "Pediatric physiology in relation to the pharmacokinetics of monoclonal antibodies." 2018, Expert Opinion on Drug Metabolism & Toxicology, 14:6, 585-599. DOI: 10.1080/17425255.2018.1482278.

between the two products. Refer to "Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product"[35] April 2015, p.21.

EMA

The reference medicinal product may have more than one therapeutic indication. When biosimilar comparability has been demonstrated in one indication, extrapolation of clinical data to other indications of the reference product could be acceptable but needs to be scientifically justified. In case it is unclear whether the safety and efficacy confirmed in one indication would be relevant for another indication, additional data will be required. Extrapolation should be considered in the light of the totality of data, i.e. quality, non-clinical and clinical data. It is expected that the safety and efficacy can be extrapolated when biosimilar comparability has been demonstrated by thorough physico-chemical and structural analyses as well as by *in vitro* functional tests complemented with clinical data (efficacy and safety and/or PK/PD data) in one therapeutic indication. Additional data are required in certain situations, such as;

- the active substance of the reference product interacts with several receptors that may have a different impact in the tested and non-tested therapeutic indications
- the active substance itself has more than one active site and the sites may have a different impact in different therapeutic indications
- the studied therapeutic indication is not relevant for the others in terms of efficacy or safety, i.e. is not sensitive for differences in all relevant aspects of efficacy and safety.

Immunogenicity is related to multiple factors including the route of administration, dosing regimen, patient-related factors and disease-related factors (e.g. co-medication, type of disease, immune status). Thus, immunogenicity could differ among indications. Extrapolation of immunogenicity from the studied indication/route of administration to other uses of the reference product should be justified.

Case study: Considerations in Extrapolating Conclusions of Biosimilarity to Other Indications:

Inflectra (CT-P13, infliximab-dyyb) advisory committee meeting: February 9, 2016:[36]

The primary mechanism of action (MOA) of infliximab is direct binding and blocking of TNF-receptor-mediated biological activities. Infliximab binds to both soluble (s) and transmembrane (tm)TNF, thus blocking TNF-binding to its receptors TNFR1 and TNFR2 and the resulting downstream pro-inflammatory cascade of events. The scientific literature indicates that this MOA is the primary MOA in rheumatoid arthritis (RA), ankylosing spondylitis (AS), psoriatic arthritis (PsA), and plaque psoriasis (PsO). The data provided by Celltrion showed similar TNF-binding and potency to neutralize TNF-α, supporting the demonstration of analytical similarity pertinent to this MOA. With respect to indications in inflammatory bowel disease (IBD), i.e., Crohn's Disease (CD) and Ulcerative Colitis (UC), in addition to binding and neutralization of soluble TNF, other mechanisms may also be in play, including reverse signalling via binding to transmembrane TNF and mechanisms involving the

[35] Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, April 2015. https://www.fda.gov/media/82647/download

[36] FDA US Advisory Committee Proceedings for Inflectra (CT-P13), February 9, 2016. https://www.fda.gov/advisory-committees/arthritis-advisory-committee/briefing-information-february-9-2016-meeting-arthritis-advisory-committee-aac

Fc region, including induction of ADCC on transmembrane TNF-expressing target cells via FcγRIIIa binding.

As noted at the FDA Advisory Committee on February 9, 2016, the comparative analytical assessment identified ***a small (~20%) difference*** *between the analysed lots of Inflectra (development name CT-P13) and US-licensed Remicade in* ***some NK-cell based ADCC assays****, which were unable to be immediately resolved because* ***of corresponding small differences in afucosylation and FcγRIIIa binding****. These small differences were of unclear clinical significance but* ***would only have been expected to impact the IBD indications****, for which there were no clinical data at the time. Ultimately, concerns regarding any analytical differences were resolved by analysing the ADCC activity of additional lots of CT-P13, US-Remicade, and EU-Remicade, and a control strategy for FcγRIIIa binding strength. The majority of the Advisory Committee was supportive of approval for IBD indications (Vote was 21 yes, 3 no) but emphasized the need for* ***post-marketing clinical data with Inflectra in IBD indications*** *and the GI prescriber and patient community was subsequently slow to adopt use. Post-marketing clinical data in IBD has subsequently further validated the small differences as not being clinically meaningful.*

Although a comparative clinical study in an IBD population would have been the most comprehensive in terms of addressing the question of clinically meaningful differences across all known mechanisms, the historical clinical trial designs in IBD utilized doses and timing of primary endpoint assessments that are in the therapeutic plateau and raised concerns that the clinical outcome measures (e.g. clinical response, clinical remission) utilized would not be sensitive to differences in analytical attributes.

1.5.5. Immunogenicity considerations

US FDA

Refer to the "Scientific Considerations in Demonstrating Biosimilarity to a Reference Product"[37] April 2015. "At least one clinical study that includes a comparison of the immunogenicity of the proposed product to that of the reference product will be expected. FDA encourages that, where feasible, sponsors collect immunogenicity data in any clinical study, including human PK or PD studies. The extent and timing of the clinical immunogenicity assessment will vary depending on a range of factors, including the extent of analytical similarity between the proposed product and the reference product, and the incidence and clinical consequences of immune responses for the reference product."

As noted in the draft Guidance for Industry: Clinical Immunogenicity Considerations for Biosimilar and Interchangeable Insulin Products[38], there may be circumstances in which, for a proposed biosimilar or interchangeable insulin product that is highly similar to its reference product, justification can be provided that there is little or no residual uncertainty regarding immunogenicity, and that like the reference product, the proposed biosimilar or interchangeable product would be expected to have minimal or no risk of clinical impact from immunogenicity, in which case, a comparative clinical study to assess immunogenicity would not be needed.

[37]Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, April 2015 https://www.fda.gov/media/82647/download

[38]Draft Guidance for Industry: Clinical Immunogenicity Considerations for Biosimilar and Interchangeable Insulin Products. November 2019, https://www.fda.gov/media/133014/download

Sponsors should consult with the relevant FDA review division to discuss the data and information that may be needed for their development program.

Comprehensive information regarding general considerations related to immunogenicity for therapeutic protein products may also be found in two guidances: Immunogenicity Assessment for Therapeutic Protein Products, August 2014[39]; and Immunogenicity Testing of Therapeutic Protein Products—Developing and Validating Assays for Anti-Drug Antibody Detection, January 2019[40].

EMA

The principles for the assessment of immunogenicity of therapeutic proteins and monoclonal antibodies have been described in two CHMP guidelines:

- Immunogenicity assessment of biotechnology-derived therapeutic proteins (https://www.ema.europa.eu/en/immunogenicity-assessment-biotechnology-derived-therapeutic-proteins)
- Immunogenicity assessment of monoclonal antibodies intended for in vivo clinical use (https://www.ema.europa.eu/en/immunogenicity-assessment-monoclonal-antibodies-intended-vivo-clinical-use)

The potential for immunogenicity of a biosimilar should be investigated in a comparative manner to the reference product and should follow the principles as laid down in the aforementioned CHMP guidelines unless it can be justified that there is a need for deviation from this approach. The type and amount of immunogenicity data will depend on the experience gained with the reference product and the product class.

Immunogenicity testing of the biosimilar and the reference product should be conducted within the biosimilar comparability exercise by using the same assay format and sampling schedule which must meet all current standards. Analytical assays should be performed with both the reference and biosimilar molecule in parallel (in a blinded fashion) to measure the immune response against the product that was received by each patient. The analytical assays should preferably be capable of detecting antibodies against both the biosimilar and the reference molecule but should at least be able to detect all antibodies developed against the biosimilar molecule. Usually, the incidence and nature (e.g. cross-reactivity, target epitopes and neutralising activity) of antibodies and antibody titres should be measured and presented and should be assessed and interpreted in relation to their potential effect on clinical efficacy and safety parameters.

Duration of the immunogenicity study should be justified on a case-by-case basis depending on the duration of the treatment course, disappearance of the product from the circulation (to avoid antigen interference in the assays) and the time for emergence of humoral immune response (at least four weeks when an immunosuppressive agent is used). Duration of follow-up should be justified based on the time course and characteristics of unwanted immune responses described for the reference medicinal product, e.g. a low risk of clinically significant immunogenicity or no significant trend for increased immunogenicity over time. In case of chronic administration, one-year follow up data will normally be required pre-authorisation. Shorter follow-up data pre-authorisation (e.g. 6 months) might be justified based on the immunogenicity profile of the reference product. If

[39] Guidance for Industry: Immunogenicity Assessment for Therapeutic Protein Products, August 2014. https://www.fda.gov/media/85017/download

[40] Immunogenicity Testing of Therapeutic Protein Products—Developing and Validating Assays for Anti-Drug Antibody Detection, January 2019. https://www.fda.gov/media/119788/download

needed, immunogenicity data for an additional period, up to one-year, could then be submitted post-authorisation.

Increased immunogenicity as compared to the reference product may become an issue for the benefit/risk analysis and would question biosimilarity. However, also a lower immunogenicity for the biosimilar is a possible scenario, which would not preclude approval as a biosimilar. In case of reduced development of neutralizing antibodies with the biosimilar, the efficacy analysis of the entire study population could erroneously suggest that the biosimilar is more efficacious than the reference product. It is therefore recommended to pre-specify an additional exploratory subgroup analysis of efficacy and safety in those patients that did not mount an anti-drug antibody response during the clinical trial. This subgroup analysis could be helpful to establish that the efficacy of the biosimilar and the reference product are in principle similar if not impacted by an immune response.

1.5.6. Interchangeability considerations

MFDS

Reference: 'Notification of authorization and review of biological product (MFDS notification)"

The Reference and biosimilar products are not interchangeable (or substitutional) in Korea. This is a specific difference between biological products and chemical drugs, by regulation.

US FDA

Refer to the "Guidance for Industry: Considerations in Demonstrating Interchangeability with a Reference Product"[41] May 2019.

Section 351(k)(4) of the Public Health Service Act describes the standards by which FDA will determine a biological product to be interchangeable with a reference product. FDA must determine that the information submitted in a 351(k) application or supplement is sufficient to show that the biological product "is biosimilar to the reference product" and "can be expected to produce the same clinical result as the reference product in any given patient" and that "for a biological product that is administered more than once to an individual, the risk in terms of safety or diminished efficacy of alternating or switching between use of the biological product and the reference product is not greater than the risk of using the reference product without such alternation or switch."

Section 351(i) of the PHS Act states that the term interchangeable or interchangeability, in reference to a biological product that is shown to meet the standards described in section 351(k)(4) of the PHS Act, means that "the biological product may be substituted for the reference product without the intervention of the health care provider who prescribed the reference product," subject to state laws.

Thus, the additional information in an application or supplement for a proposed interchangeable is intended to support the efficacy and safety of potential pharmacy-level substitution. The Guidance for Industry lists scientific factors to consider and possible study designs to assess the risk of alternating or switching. Product-specific and clinical context-specific factors will influence the extent and type of information needed to support an application for interchangeability. Sponsors are encouraged to discuss the issue with the relevant FDA review division.

[41] Guidance for Industry: Considerations in Demonstrating Interchangeability with a Reference Product, May 2019. https://www.fda.gov/media/124907/download

EMA

In the EU, interchangeability generally refers to the possibility of exchanging one medicine for another medicine that is expected to have the same clinical effect. This could mean replacing a reference product with a biosimilar (or vice versa) or replacing one biosimilar with another. Replacement can be done by:

- **Switching**, which is when the prescriber decides to exchange one medicine for another medicine with the same therapeutic intent.
- **Substitution** (automatic), which is the practice of dispensing one medicine instead of another equivalent and interchangeable medicine at pharmacy level without consulting the prescriber.

Centralised evaluations do not include recommendations on whether the biosimilar is interchangeable with the reference medicine, and thus whether the reference medicine can be switched or substituted with the biosimilar.

- The decision on whether to allow interchangeable use and substitution of the reference biological medicine and the biosimilar is taken at national level. However, there is broad consensus that any biosimilar approved via the centralised procedure is in principle considered to be interchangeable at the prescriber level.

Israeli Ministry of Health (IMOH)

The Israeli MOH policy regarding interchangeability is as follows:

- Switching is not allowed at pharmacy level.

- Treatment of naïve patients:
  The treating physician will choose between the biosimilar and the reference product, in coordination with the HMO (health maintenance organization).

- Patients on treatment with the reference (originator) product or biosimilar:
  An automatic switching by the treating physician, between the reference product and the biosimilar, is not allowed.

  After the marketing authorization of the biosimilar product, the MAH may submit an application for the approval of interchangeability with the reference product.

  The decision regarding interchangeability will be made after the recommendation of an ad-hock advisory committee and with the agreement of the treating physician. The decision will be based on the existing clinical data on interchangeability and the interchangeability status in different health authorities.