White Paper

Rational Selection, Criticality Assessment, and Tiering of Quality Attributes and Test Methods for Analytical Similarity Evaluation of Biosimilars

Kristof Vandekerckhove,¹ Andreas Seidl,² Hiten Gutka,³ Manish Kumar,⁴ Gyöngyi Gratzl,⁵ David Keire,⁶ Todd Coffey,⁷ and Henriette Kuehne^{8,9}

Received 13 February 2018; accepted 26 March 2018; published online 10 May 2018

Abstract. Leading regulatory agencies recommend biosimilar assessment to proceed in a stepwise fashion, starting with a detailed analytical comparison of the structural and functional properties of the proposed biosimilar and reference product. The degree of analytical similarity determines the degree of residual uncertainty that must be addressed through downstream *in vivo* studies. Substantive evidence of similarity from comprehensive analytical testing may justify a targeted clinical development plan, and thus enable a shorter path to licensing. The importance of a careful design of the analytical similarity study program therefore should not be underestimated. Designing a state-of-the-art analytical similarity study meeting current regulatory requirements in regions such as the USA and EU requires a methodical approach, consisting of specific steps that far precede the work on the actual analytical study protocol. This white paper discusses scientific and methodological considerations on the process of attribute and test method selection, criticality assessment, and subsequent assignment of analytical measures to US FDA's three tiers of analytical similarity assessment. Case examples of selection of critical quality attributes and analytical methods for similarity exercises are provided to illustrate the practical implementation of the principles discussed.

KEY WORDS: biosimilars; comparability; CQA; QTPP; tiering.

INTRODUCTION

Medicines of biological origin are complex mixtures containing a diversity of chemical variations of a therapeutic protein and other substances as a result of their production in living systems.

The specific composition of such a product is controlled within carefully defined limits of variation by the design and control of its manufacturing process. Because the details of product manufacture are proprietary knowledge, biosimilar product developers cannot precisely replicate the manufacturing process of a reference product. Due to the inherent manufacturing variability associated with cell culture and the inability to precisely replicate the originator's manufacturing process, even the best reproductions of protein therapeutics can at best be highly similar to a reference product based on current technologies.

The development of a biological product that closely resembles an existing (originator or reference) product requires an approach that is distinct from traditional product development (Fig. 1). Biosimilar products are modeled after a product that has been thoroughly studied in large nonclinical and clinical study programs, and for which usually a wealth of information is (publicly) available, including years of commercial use in large and diverse patient populations. As a result, a first iteration of quality attribute (QA) selection and ranking can be completed prior to product development. Detailed characterization of the reference product supports further attribute selection and ranking, and ultimately yields a precise quality target product profile (QTPP) for product and process development. As development proceeds, the accumulating knowledge from structural and functional characterization studies provides

¹ Strategic Drug Development, Design and Delivery Innovation, IQVIA, Mediaaan 32/2, 1800, Vilvoorde, Belgium.

² Global Analytical Characterization & Bioanalytics, Technical Development Biosimilars, Novartis BTDM, Hexal AG, 82041, Oberhaching, Germany.

³ Oncobiologics Inc., 7 Clarke Drive, Cranbury, New Jersey 08512, USA.

⁴ Biocharacterization Development, Intas Pharmaceuticals Ltd., Biopharma Division, Plot No. 423/P/A, Sarkhej-Bavla Highway, Moraiya, Ahmedabad, 382213, India.

⁵ West-Ward Pharmaceuticals (a Hikma Company), 300 Northfield Rd., Bedford, Ohio 44146, USA.

⁶ CDER/OPQ/OTR/DPA, U.S. FDA, St. Louis, Missouri 63101, USA.

⁷ Idaho College of Osteopathic Medicine, Meridian, Idaho 83642, USA.

⁸ Coherus BioSciences, 333 Twin Dolphin Drive, Suite 600, Redwood Shores, California 94065, USA.

⁹ To whom correspondence should be addressed. (e-mail: hkuehne@coherus.com)

increased insight into the criticality of the various quality attributes, thus supporting refinement of the QTPP and, consequently, appropriately focused process development.

Once satisfactory biosimilar material has been produced, analytical assessment of biosimilarity can proceed. The analytical assessment (and corresponding analytical study design) is viewed as a critical first step in biosimilar development by both US FDA and EMA, since the scope and range of the requisite clinical studies needed for approval of the biosimilar will be, in part, determined by the degree of analytical similarity. The assessment can be broken down into several distinct steps. First, molecular attributes to be considered are defined based on the drug's structure, mechanism of action, safety, and efficacy, and appropriate analytical methods to assess those attributes are selected. Second, the criticality of each attribute is determined based on its effect on clinical outcome, as assessed from a number of angles. Both of these steps build on the work that has ideally been initiated at, or prior to, the start of product development. US FDA specifically expects a third step, in which the attributes are assigned to different tiers of statistical evaluation, each of which is associated with a specific methodology of data analysis and interpretation. The statistical methodology for evaluation of analytical similarity data was discussed in a previous publication (1) and is the subject of recent draft regulatory guidance documents (2,3).

This white paper provides possible methodologies distilled from the combined experience of a group of subject matter experts from different organizations regarding the selection of quality attributes and their criticality assessment, as a prerequisite to the design of an analytical similarity study program meeting the current expectations of leading regulatory agencies. It also discusses the assignment of proposed analytical similarity measures to the different tiers of assessment recommended by US FDA (2,4). This document was prepared under the auspices of the AAPS Biosimilars Focus Group's Chemistry Manufacturing and Controls Analytical (CMC-A) subcommittee and is intended to share current practices within the CMC-A community to aid the development of biosimilar products.

SELECTION OF ATTRIBUTES AND ANALYTICAL PROCEDURES

The main purpose of any analytical similarity assessment is to identify and evaluate similarity in all attributes that could impact a drug's purity, potency, safety, and efficacy. Therefore, as a first step in the analytical similarity exercise, the QAs are defined based on the drug's structure, mechanism of action, safety, and efficacy. QAs for a biotherapeutic can be identified and defined using several or all of the

following resources:

- i. Any publicly available information from the reference product manufacturer that provides information on QAs that are tested and tightly controlled within a predefined range for the reference product. Notably, Certificates of Analysis (COAs), when available, may only be accessible for the reference drug product (DP) not for the originator drug substance (DS). Release

tests that are only performed on DS may not be listed on the DP COA.

- ii. Reported information on the reference product and similar products from other sources, *e.g.*, information from standard setting organizations (product-specific or molecular type, *e.g.*, monoclonal antibodies), regulatory guidelines (*e.g.*, ICH Q6B), pharmacovigilance reports, European public assessment reports (EPARs) on the reference product, and health agency communications to the reference product sponsor (post marketing commitments or findings).
- iii. Literature reports on the reference product and related products (research papers, mechanism of action, and clinical findings).
- iv. Results from detailed analytical characterization of multiple lots of the reference product.

Case study example: attribute selection

Avastin (bevacizumab) is known to exert its clinical effect by binding and neutralization of vascular endothelial growth factor (VEGF), preventing association of VEGF with its receptor by steric hindrance. Original research did not reveal a contribution of Fc-related bioactivity (antibody-dependent cell-mediated cytotoxicity (ADCC), complement-dependent cytotoxicity (CDC)) to the observed clinical effect of bevacizumab (5). However, bevacizumab may also bind with membrane-bound VEGF (expressed by certain carcinomas, or in VEGF ligand-receptor complexes) and can hypothetically mediate Fc-related effector functions. Salvador *et al.* (6) report *in vitro* CDC activity for bevacizumab in human ovarian carcinoma SKOV-3 cells. This example illustrates how all plausible biological functions of the molecule of interest, even those reported to be clinically irrelevant for the reference product, must be included in the list of attributes for similarity assessment. Publically available information on the development of ABP215, Amgen's biosimilar version of bevacizumab, confirms this expectation (7).

After defining QAs for inclusion in the analytical similarity assessment, appropriate techniques to assess those QAs are selected, based on a variety of considerations:

- i. Suitability of the analytical method to measure existing differences between reference product and biosimilar product.
- ii. Ability of the method to yield quantitative data (*e.g.* qualitative vs. quantitative methods such as SDS-PAGE vs. capillary gel electrophoresis with sodium dodecyl sulfate (CE-SDS)).
- iii. Orthogonality. Each attribute may be interrogated by one or more analytical methods. Conversely, the same method may inform on several attributes. For example, aggregates are observed, *e.g.*, by SEC, AUC (analytical ultracentrifugation), or AF4 (asymmetrical flow field flow fractionation), to name a few. At the same time, SEC can also be used to differentiate other size variants, such as dimers or clips.

As the biosimilar product manufacturing process and the overall biosimilar product understanding matures, test methods dedicated to testing particular QAs are developed, qualified, validated, and implemented. In addition, as the analytical tools are developed and optimized, the improved

Fig. 1. Overview of CQA selection process

assays can be used to provide feedback for the manufacturing process to achieve a high degree of similarity. While a formal validation study is not strictly required for methods used in similarity assessments, their suitability for the intended purpose should be demonstrated. Sponsors also need to consider the composition of the test samples for analytical similarity studies: regulatory agencies usually expect analysis of the drug product, except for quality attributes that cannot be adequately measured in the drug product, *e.g.*, due to low protein concentration or the presence of interfering excipients. In such cases, appropriate deformation or other sample processing strategies have to be developed.

Dedicated tests methods can be implemented for use in the control strategy of the biosimilar product (if appropriately validated), as well as the similarity assessment. As an example, product variants such as methionine oxidation, asparagine deamidation and aspartate isomerization may initially be characterized by liquid chromatography coupled to detection by tandem mass spectrometry (LC-MS/MS), while at a later stage of development and manufacturing, other analytical methods more suited for routine analytical control (*e.g.*, reverse phase chromatography (RPC), hydrophobic interaction chromatography (HIC), or cation exchange chromatography (CEX)) may be developed, qualified, and validated for quality control (QC), stability, and similarity assessment of specific critical variants. The implementation of such a specific and validated method in a quality control setting would be ultimately dependent upon the criticality of the attribute being measured and the need for routine analytical control (in turn depending on the risk of the attribute failing acceptable limits when the process is operated within its design space). Table 1 shows a non-exhaustive list of quality attributes to be considered for similarity assessment of a therapeutic protein, as appropriate for the product under evaluation; together with possible analytical techniques that are conventionally used for this purpose by biosimilar product developers.

Expectations for comparison of levels of process-related impurities (*e.g.*, host-cell protein (HCP), host-cell deoxyribonucleic acid (DNA), Protein A) differ between jurisdictions. Some agencies, such as EMA, do not expect comparative

analysis, arguing that these impurities are process-dependent and are therefore not expected to be similar, although they must be controlled according to acceptable standards (8). US FDA, on the other hand, does instruct sponsors to compare levels of process-related impurities, and to perform a risk-based assessment in case of differences in process-related impurities (9).

CRITICALITY ASSESSMENT

Once a comprehensive list of QAs is established, there should be a risk assessment of each QA for potential patient impact. The relationship between the attribute and the product's clinical performance (PK, PD, efficacy, and safety) should be rigorously evaluated, using prior knowledge in combination with sound scientific judgment. Various sources of "prior knowledge" are available, several of which are the same as those previously discussed for attribute selection. Examples are

- Knowledge of the product's specific mode of action;
- Publicly available information on the reference product, including scientific literature and regulatory approval history;
- Knowledge of similar products, in particular information on effector functions and interactions with biological targets that are relevant for the proposed biosimilar, and information on the relationship between structural features and clinical outcomes; and
- The sponsor's own experimental data, *e.g.*, results of biological characterization (*e.g.*, bioassays, *in vivo* models or clinical studies) of isolated product variants (or variants obtained by enzymatic processing, forced degradation, etc.).

Different models have been used successfully in the criticality assessment of product quality attributes. Examples of such models have been presented at conferences, by representatives of companies who have received approval for their biosimilar products in the USA (10–12). Based on these examples, and the collective experience of the authors,

Table I. Proposed Quality Attributes and Possible Techniques for Analytical Similarity Assessment of a Protein Therapeutic

Category	Attribute (method or methods to query that attribute)
Primary structure	<ul style="list-style-type: none"> • Primary sequence (<i>e.g.</i>, UPLC peptide map, LC-MS/MS, amino acid analysis, Edman degradation, carboxypeptidase sequencing) • Disulfide structure (<i>e.g.</i>, LC-MS of non-reduced protein digest) • Intact mass (<i>e.g.</i>, LC-MS) • Isoelectric point (<i>e.g.</i>, IEF gels, cIEF, iCE) • Extinction coefficient (<i>e.g.</i>, UV/AAA or UV/RI)
Secondary and tertiary structure	<ul style="list-style-type: none"> • Low-resolution secondary structure or indirect tertiary structure measurements (<i>e.g.</i>, CD, DSC, FTIR, fluorescence) • High-resolution measurements of higher order structure (<i>e.g.</i>, 2D-NMR, HDX-MS, X-ray crystallography)
Glycosylation	<ul style="list-style-type: none"> • Glycosylation (<i>e.g.</i>, HILIC, MS (MALDI, ESI), exoglycosidase sequencing, HPLC-FLD, HPAEC-PAD, CE-LIF) • Glycosylation site mapping/site occupancy (<i>e.g.</i>, peptide mapping by LC-MS)
Dose	<ul style="list-style-type: none"> • Protein content (<i>e.g.</i>, UV A280, RP-HPLC) • Deliverable volume (extractable volume)
Particulates	<ul style="list-style-type: none"> • Sub-visible particles (<i>e.g.</i>, light obscuration, MFI, NTA)
Function	<ul style="list-style-type: none"> • Biological activity (<i>e.g.</i>, for mAb: proliferative bioassay, cytotoxicity assay, ADCC, and CDC; other assays may be appropriate for other proteins, <i>e.g.</i>, enzyme kinetics for biosimilar enzyme) • Receptor and/or ligand binding (<i>e.g.</i>, SPR, ELISA)
Product variants (product-related substances and impurities)	<ul style="list-style-type: none"> • High molecular weight species (<i>e.g.</i>, SEC-MALS, AF4/HF5, AUC, DLS) • Covalent dimers (<i>e.g.</i>, SDS PAGE, CE-SDS) • Purity and impurities (oxidation, deamidation, glycation, isomerization, fragmentation, disulfide reduction, <i>e.g.</i>, RP-HPLC, CEX, SEC, IEX, IEF, cIEF, LC-MS) • Amino acid misincorporations (<i>e.g.</i>, LC-MS/MS) • Microsequence heterogeneity (<i>e.g.</i>, LC-MS) • C- and N-terminal modifications (<i>e.g.</i>, LC-MS, Edman degradation)

Other quality attributes, such as general pharmaceutical properties (pH, osmolality, appearance, and color), may also be measured for the proposed biosimilar and reference product. These tests are not required for demonstration of analytical similarity and are therefore excluded from this table

2D two dimensional, AAA amino acid analysis, ADCC antibody-dependent cell-mediated cytotoxicity, AF4 asymmetrical flow FFF, AUC analytical ultracentrifugation, CD circular dichroism, CDC complement-dependent cytotoxicity, CE capillary electrophoresis, CEX cation exchange, cIEF capillary IEF, DLS dynamic light scattering, DSC differential scanning calorimetry, ELISA enzyme-linked immunosorbent assay, ESI electrospray injection, FLD fluorescence detection, FTIR Fourier transform infrared spectroscopy, HDX hydrogen deuterium exchange, HF5 hollow fiber flow field flow fractionation, HILIC hydrophilic interaction chromatography, HPAEC high-performance anion exchange chromatography, HPLC high-performance liquid chromatography, iCE imaged capillary isoelectric focusing, IEF isoelectric focusing, LC liquid chromatography, LIF laser-induced fluorescence, MALDI matrix-assisted laser desorption/ionization, MALS multi-angle light scattering, MFI microflow imaging, MS mass spectrometry, MS/MS tandem mass spectrometry, NMR nuclear magnetic resonance, NTA nanoparticle tracking analysis, PAD pulsed amperometric detection, PAGE polyacrylamide gel electrophoresis, RI refractive index, RP reversed phase, SDS sodium dodecyl sulfate, SEC size exclusion chromatography, UV ultraviolet

we conclude that most models appraise at least two out of three dimensions of criticality. First and foremost, the relationship between the attribute of interest and the clinical performance of the product is evaluated, to determine how the attribute may affect clinical outcome. The second dimension commonly employed assesses the evidence supporting the attribute's clinical impact, which can range from purely academic (no data available) to unequivocal proof derived from (multiple) clinical studies or other clinical experience. Finally, a risk score may be factored into the assessment of attribute criticality, such as the risk of the attribute truly affecting clinical performance. One example is when a product variant is found to be biologically inactive (or less active), and therefore viewed as critical (risk of reduced efficacy). If the product variant in question is only formed in minute quantities (independently from any clearance during downstream processing), it may be judged to be clinically irrelevant based on its low abundance. An additional consideration for industry may be the likelihood of the

attribute being dissimilar based on knowledge of process performance and level of process control. Notably, certain quality attributes may have a critical impact on product efficacy or safety if not controlled within a predefined range, but are so tightly controlled by process design, in-process, and quality control testing that the risk of the attribute being out of range is negligible, reducing the overall criticality ranking for this attribute. Thus, a tightly controlled process can mitigate product risk, may impact the need for release testing as part of the control strategy, and may also impact the tier of an attribute in a similarity assessment. One such example is protein dose (*i.e.*, the absolute amount of therapeutic protein administered to the patient) which can be tightly controlled by manufacturing processes. However, protein dose remains a critical attribute for similarity assessment because of potential impact on the patient who may be receiving the reference product and/or the biosimilar product as part of their treatment. Importantly, consideration of the product control strategy for determination of attribute criticality is a practice

that is not endorsed by all companies and is not accepted by regulatory agencies.

Quantitative or qualitative assessment techniques can be used to determine attribute criticality. Quantitative models use predefined scoring roles, whereby each assessment dimension is broken down into several numerical scores with an associated definition, ideally in sufficient detail to avoid ambiguity and subjectivity. An aggregate score for the attribute is calculated by simple arithmetic, usually by multiplying the individual scores assigned to each assessment dimension. In a next step, each attribute is assigned to a criticality category (*e.g.*, low-moderate-high, each associated with a predefined scoring range) based on the attribute's aggregate score. Examples of quantitative risk scoring techniques are described in the A-Mab case study (13). A variation of Tool 1 described in this case study is used by Sandoz (11,12), the first company to receive approval for a biosimilar product in the USA and EU. Qualitative approaches equally define different levels within each assessment dimension (although these are not associated with a numerical score), but do not make use of a calculation for determination of the overall criticality "score" (or category). Instead, visual approaches (such as heat maps based on relative rankings) may be used (10). Alternatively, rules can be defined that clarify how the assessment dimensions must be integrated to determine the overall criticality category. Whichever approach is chosen, the end-result will consist of the complete list of quality attributes grouped into a number of criticality categories.

Overall, each criticality assessment model has its strengths and weaknesses. Quantitative models have the advantage of offering a rank order of all quality attributes, meaning that attributes within a specific criticality category are further ranked in order of criticality. This approach allows sub-sampling of attributes from a category for specific purposes, which may prove useful when tiering the attributes. A disadvantage may be that the overall criticality score (and thus criticality category) is very sensitive to the individual scores assigned to each assessment dimension (due to multiplication of dimension scores) and a slight difference in interpretation may therefore result in a different categorization of the attribute. Consider a case where attributes are assessed using the assessment dimensions impact and uncertainty (refer to the A-Mab case study, quality attribute assessment Tool #1 (13), for details on this methodology), with criticality defined in categories of low (scores 0–29), moderate (scores 30–79), and high (scores 80–140). Take a quality attribute for which nonclinical data demonstrate this attribute decreases the biological activity of the product. In this example, an uncertainty score of 3 will be assigned (given the availability of nonclinical data for this molecule), but the impact may be scored differently depending on how critical the impact is judged to be by the company (*e.g.*, low ("acceptable change," score = 4) or moderate ("moderate change," score = 12))—arguably a subjective decision. The resulting aggregate score (12 or 36) can therefore either lead to a designation of low or moderate criticality, which determines attribute control (including the tier it is assigned to for similarity assessment). A qualitative model may be more intuitive and less laborious, but bears the risk of less substantiated judgment: experience shows that qualitative assessments are often based on "expert opinions," whereas quantitative tools tend to promote more

intense research of information sources. However, this distinction is not absolute, and qualitative models can equally rely on substantial documentation of proposed criticality assignments.

We choose not to recommend a specific criticality assessment method, but emphasize that whichever model is preferred by the developer, it must be fit for purpose, compatible with the company's internal practices, demonstrate comprehensive justification of scoring decisions, and ensure an appropriate assessment hierarchy for linked attributes. The latter point refers to several quality attributes that relate to a single critical attribute, whereby each individual quality attribute is independently scored as critical because of its link to the critical attribute. A typical example is aggregation: aggregates are often viewed as highly critical, particularly for therapeutic proteins that intend to substitute endogenous proteins (*e.g.*, erythropoietin, insulin, somatotropin, G-CSF). Different quality attributes may increase the propensity of aggregation if inadequately controlled, such as protein conformation, protein modifications such as oxidation and deamidation, and intermolecular disulfide bonds. The criticality assessment should differentiate between attributes that are directly associated with an adverse clinical impact (*i.e.*, aggregation—which can be measured by various analytical techniques), and attributes that have a relationship with the critical attribute (which should be assessed independently of their potential to promote aggregation). Another example of such linkage is the correlation between (a) ADCC (a direct measure of biological activity) and (b) the quality attributes FcγR3a binding (one of the effector functions involved in ADCC) and afucosylated glycan content (correlated with FcγR binding) (14). The developer should consider all the roles of these linked QAs, their correlation with each other, and the specific assays being used in designing a similarity strategy.

Finally, we note that FDA in its recently published draft guidance document "Statistical Approaches to Evaluate Analytical Similarity" (2) explicitly recommend the use of the impact × uncertainty model described earlier in this paper, although specifying that known clinical impact (low uncertainty) should predominate over unknown but potential clinical impact (high uncertainty). The latter requirement is a departure from Tool 1 published in the A-Mab case study (13), which assigns a higher overall criticality score in case of high uncertainty, but is consistent with the methodology used by Sandoz (see case study).

Case study example: Sandoz' approach to attribute ranking (15) The Sandoz approach to criticality assessment, as published for its candidate biosimilar product for etanercept, starts with assigning all quality attributes to one of the following categories: (a) product-related variants; (b) process-related impurities; and (c) potency, identity, strength and composition, and appearance and description. The criticality of each attribute is then assessed by different risk assessment tools. Specifically, tools A and B calculate the criticality score as a function of *impact* and *uncertainty*. Tool C is used for attributes that are considered highly critical by default. Tool A calculates the impact and uncertainty scores for product-related variants. The impact ranking assesses the known or potential influence on, or consequences for, the clinical performance, and generates a numerical impact

factor (ranging from 2 to 20), by considering the attribute's effect on the following:

1. Efficacy, either through clinical experience or results using the most relevant potency assay(s)
2. Pharmacokinetics/pharmacodynamics (PK/PD)
3. Immunogenicity (against the drug substance)
4. Safety

In addition to the impact factor, an uncertainty factor is assigned (ranging from 1 to 7). The uncertainty factor depends on the type and source of knowledge available for a quality attribute, e.g., sponsor or literature data on *in vitro* or *in vivo* studies, clinical data relating to the same attribute in a related molecule or even clinical data on the molecule of interest containing the variant in question. The category with highest calculated score determines the overall criticality score for the quality attribute.

Tool B assesses the criticality of process- and excipient-related impurities by a similar approach to that taken for active-ingredients in tool A above. Tool B considers only the relevance to safety/immunogenicity for impact scoring (ranging from 2 to 20). Similarly, uncertainty scoring can take only three different default values: 1 (none/no uncertainty, e.g., generally regarded as safe), 3 (low, e.g., component used in previous processes), or 7 (high, e.g., no information available, new impurity). The Sandoz approach differs from the approach described in tool 1 of the A-Mab case study for calculation of the overall criticality score. The A-Mab case study tool ranks quality attributes with known (regardless whether positive or negative) impact on clinical performance as lower than those for which there is uncertainty with respect to their impact on, e.g., efficacy or safety. This approach is difficult to justify from the perspective of patient interest, safety and well-being: those attributes certain to have an impact on clinical performance should be assigned the highest ranking, to ensure that they are well controlled and receive the highest attention in the similarity assessment and control strategy. In addition, product development would be mainly uncertainty-driven if those attributes with higher uncertainty received greater attention than those for which the impact is known. This uncertainty dilemma in the A-Mab case study tool 1 is caused by the approach of simple multiplication of the criticality sub-scores to calculate the overall criticality score (criticality score = impact score $\times$ uncertainty score). Therefore, if two attributes have exactly the same impact score, the attribute with higher uncertainty would be ranked higher. To overcome this issue, Sandoz modified the equation and implemented a full-factorial function ensuring that those attributes with low uncertainty and limited impact on clinical performance receive the lowest score, followed by those with uncertainty about their low impact, attributes with high uncertainty about their high impact, and the highest scores reserved for attributes with low/no uncertainty about their known high impact. For example, consider two quality attributes for the same product. For attribute 1, clinical evidence exists of its very high impact on safety. This would translate in an impact score of 20 and an uncertainty score of 1. Attribute B is a new glycoform of the product, for which a very high impact on immunogenicity is assumed on theoretical grounds only (no experimental data available); an impact score of 20 and an uncertainty score of 7 are assigned. In this example, the A-Mab approach would rank attribute 2 (total score 140) higher in criticality than attribute 1 (total score 20). Sandoz' approach used for criticality assessment of Erelzi's quality attributes results in

an inversed ranking: attribute 1 receives a score of 140, attribute 2 a score of 90.

The scoring methodology for tool A and tool B are depicted in Fig. 2 below. In contrast to product-related variants and process- and excipient-related impurities, quality attributes related to potency, identity, strength and composition, and appearance and description are considered critical quality attributes by default, unless justified otherwise (tool C). Different criticality categories (and therefore criticality scores) are assigned to these quality attributes depending on their individual impact on safety and efficacy and a written justification included for each

Case study example: Amgen's approach to attribute ranking (10)

Amgen's approach of risk ranking involves relative prioritization of attributes that have the potential to impact potency, pharmacokinetics, safety, and immunogenicity. It also considers the potential risk for a difference based on process design (e.g., cell line differences, formulation, and process steps). Amgen employs a two-dimensional grid (Fig. 3) as a tool for criticality mapping. One axis scores the potential impact on similarity of potency, safety, or pharmacokinetics. The other axis considers risk to similarity due to differences in-process design. For example, attributes which could change during stability are considered of higher risk (e.g., HMWs, LMWs, disulfide modifications) than attributes highly controlled by process design (e.g., HCP, DNA).

TIER ASSIGNMENT (US FDA ONLY)

FDA recommends assigning each quality attribute to one of the three tiers, each of which is associated with a specific methodology of (statistical) assessment of the analytical data. Tiers are assigned based on a risk assessment that takes into consideration potential clinical impact and uncertainty (2). Tier assignment starts with criticality. Each criticality category should be mapped to one of the three assessment tiers defined by US FDA, whereby tier 1 is reserved for the most critical tests with direct impact to the mode of action, tier 2 is used for tests of moderate criticality, and tests for attributes of the lowest criticality rank are assigned to tier 3. Several additional factors should be considered to determine the ultimate tier assignment of a quality attribute or analytical test.

Before discussing these factors, we need to first emphasize that analytical measures, not quality attributes, are assigned to the different assessment tiers. We define analytical measure as the result of an analytical procedure that is proposed as a basis for determination of similarity between the proposed biosimilar and its reference product. For example, oligosaccharides profiling by HPLC will yield quantitative information on all enzymatically released glycan species that are quantifiable by the method, in addition to a chromatographic fingerprint of the glycan population. Each individual glycan, predefined groups of glycans (sialylated, afucosylated, etc.), and the chromatogram itself, all could be defined as analytical measures. This distinction is important: most quality attributes will be measured by several orthogonal analytical techniques (in line with US FDA's recommendations (9)), but the agency does not expect that each individual test (or measure from that test) is subsequently assigned to the same tier.

Tests which only yield qualitative information, such as spectra or other analytical fingerprints (with no possibility of

* The criticality score is calculated by a (undisclosed) full-factorial equation developed by Sandoz.

Fig. 2. Sandoz criticality assessment model: Tool A and Tool B (adapted from (15))

Fig. 3. Amgen risk ranking approach. Source: Karow, 2016 (10)

deconvolution), or are binary in nature (pass/fail, *e.g.*, protein sequence), cannot be assigned to tier 1 or tier 2, regardless of their criticality. Attribute content levels are another factor of consideration. For example, if several or all results from analysis of the proposed biosimilar and/or reference product are below the quantitation limit of the most sensitive analytical method available for measurement of that attribute, then the data cannot be processed by US FDA's recommended statistical tests, necessitating assignment to tier 3 (regardless of the criticality of the associated quality attribute). Tier assignment may also be influenced by analytical considerations. If, for example, different analytical measures controlling the same highly critical quality attribute were all assigned to tier 1, the probability of at least one false negative result (to be understood as the chance that non-equivalence is concluded, when the products are truly equivalent, *i.e.*, a type II error) greatly increases. Therefore, a sponsor should propose to assign only one analytical measure per attribute to tier 1, while other analytical measures of the same attribute are assigned to tier 2, based on criteria such as method performance (sensitivity, precision), known method limitations; and the clinical relevance of the analytical information.

Once the tiers have been assigned and the study conducted, the results are evaluated. An earlier commentary by the AAPS Biosimilars Focus Group's CMC-A subcommittee addresses this final step in the analytical similarity exercise, the statistical evaluation of analytical similarity results (1). In brief, tier 1 tests are assessed using mean equivalence testing while tier 2 tests are evaluated using a quality range approach. Results from tests assigned to tier 3 are compared using various forms of visual displays.

In conclusion, meaningful assignment of analytical similarity data to one of the US FDA's recommended three tiers for assessment requires detailed knowledge of the type of information that is offered by each method, mapped against a predefined and criticality-ranked list of quality attributes. Each analytical measure used to demonstrate analytical similarity should be assigned to the appropriate tier through the application of a set of rules, based on criteria such as attribute criticality, analytical data type, attribute levels, and analytical considerations.

CONCLUSION

The selection and criticality ranking of quality attributes constitutes an essential step in early biosimilar development, informing both process development and controls as well as the design of the pivotal analytical similarity study. The attributes and associated analytical methods are selected based on careful review of the various sources of knowledge for the product or related molecules, including the biosimilar manufacturer's own data. Subsequent appraisal of the known or suspected relationship between each quality attribute and the clinical performance of the product yields a rank order of attribute criticality. A number of approaches to criticality ranking have been described, each with their own set of advantages and disadvantages. Case examples of products approved by US FDA show that varying approaches to criticality ranking have been employed in their development, indicating that successful programs can take different, equally valid, paths.

In case of analytical similarity assessment programs intended to support US registration, individual tests need to be assigned to one of the three tiers, each associated with a specific methodology

of data evaluation, as a final step in program design. Attribute criticality is only one among several factors that determine the ultimate assignment of each analytical test to the appropriate tier.

With the emergence of detailed regulatory standards and the wide availability of increasingly powerful analytical technology, the structural and functional comparison of a proposed biosimilar with its reference product continues to gain in importance. Sponsors are therefore encouraged to spend careful attention to the design of their *in vitro* study program and discuss their proposal with regulatory agencies prior to execution.

ACKNOWLEDGMENTS

The authors wish to thank Thomas Stangler for his contributions to develop CQA assessment concepts and helpful discussions related to this manuscript.

COMPLIANCE WITH ETHICAL STANDARDS

FDA Disclaimer This article reflects the views of the authors and should not be construed to represent FDA's views or policies.

GLOSSARY

2D	Two dimensional
AAA	Amino acid analysis
ADCC	Antibody-dependent cell-mediated cytotoxicity
AF4	Asymmetrical flow FFF
AUC	Analytical ultracentrifugation
CD	Circular dichroism
CDC	Complement-dependent cytotoxicity
CE	Capillary electrophoresis
CEX	Cation exchange
cIEF	Capillary IEF
CQA	Critical quality attribute
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray injection
FFF	Field flow fractionation
FLD	Fluorescence detection
FTIR	Fourier transform infrared
HCP	Host-cell protein
HDX	Hydrogen deuterium exchange
HF5	Hollow fiber flow FFF
HILIC	Hydrophilic interaction chromatography
HPAEC	High-performance anion exchange chromatography
HPLC	High-performance liquid chromatography
iCE	Imaged capillary isoelectric focusing
IEF	Isoelectric focusing
LC	Liquid chromatography
LIF	Laser-induced fluorescence
MALDI	Matrix-assisted laser desorption/ionization
MALS	Multi-angle light scattering
MFI	Microflow imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NMR	Nuclear magnetic resonance
NTA	Nanoparticle tracking analysis
PAD	Pulsed amperometric detection

PAGE	Polyacrylamide gel electrophoresis
QA	Quality attribute
QC	Quality control
RI	Refractive index
RP	Reversed phase
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
UPLC	Ultra-high-pressure liquid chromatography
UV	Ultraviolet

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- Burdick R, Coffey T, Gutka H, Gratzl G, Conlon HD, Huan C-T, *et al.* Statistical approaches to assess biosimilarity from analytical data. *AAPS J.* 2017;19:4–14. <https://doi.org/10.1208/s12248-016-9968-0>.
- U.S. FDA. Guidance for industry (draft): statistical approaches to evaluate analytical similarity. September 2017. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM576786.pdf>. Accessed 1 Feb 2018.
- EMA. EMA/CHMP/138502/2017: Reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development (draft). March 23, 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/03/WC500224995.pdf. Accessed 22 Jan 2018.
- Tsong Y, Dong X, Shen M. Development of statistical methods for analytical similarity assessment. *J Biopharm Stat.* 2017;27:197–205. <https://doi.org/10.1080/10543406.2016.1272606>.
- Wang Y, Fei D, Vanderlaan M, Song A. Biological activity of bevacizumab, a humanized anti-VEGF antibody in vitro. *Angiogenesis.* 2004;7:335–45.
- Salvador C, Li B, Hansen R, Cramer DE, Kong M, Yan J. Yeast-derived β -glucan augments the therapeutic efficacy mediated by anti-vascular endothelial growth factor monoclonal antibody in human carcinoma xenograft models. *Clin Cancer Res.* 2008;14:1239–47.
- U.S. FDA. Oncologic drugs advisory committee (ODAC) meeting briefing document ABP215 (proposed biosimilar to Avastin). July 13, 2017. <https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/UCM566365.pdf>. Accessed 26 Jan 2018.
- EMA. EMA/CHMP/BWP/247713/2012: Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: quality issues (revision 1). May 22, 2014. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/06/WC500167838.pdf. Accessed 20 Dec 2017.
- U.S. FDA. Guidance for industry: quality considerations in demonstrating biosimilarity of a therapeutic protein product to a reference product. April 2015. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291134.pdf>. Accessed 20 Dec 2017.
- Karow, M. Applying risk ranking for similarity into the QTPP for antibody biosimilars. 2016 PDA Biosimilars Conference, Baltimore, MD, USA. June 20–21, 2016.
- Sonderegger C. QTPP for biosimilars: from reference product data to biosimilarity criteria. 2016 PDA Biosimilars Conference, Baltimore, MD, USA. June 20–21, 2016.
- Stangler T. What to control? CQAs and CPPs. 2011 BWP workshop on setting specifications, London, UK. September 9, 2011. http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2011/10/WC500115824.pdf. Accessed 15 Dec 2017.
- CMC Biotech Working Group. A-Mab: a case study in bioprocess development (version 2.1). October 30, 2009. http://cymcdn.com/sites/www.casss.org/resource/resmgr/imported/A-Mab_Case_Study_Version_2-1.pdf. Accessed 1 Dec 2017.
- Reusch D, Tejada ML. Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology.* 2015;25:1325–34.
- U.S. FDA CDER. Quality review BLA761042 - GP2015 (Etanercept), Sandoz. 2016. Application number 761042Orig1s000 chemistry review(s). https://www.accessdata.fda.gov/drugsatfda_docs/nda/2016/761042Orig1s000ChemR.pdf. Accessed 29 Jan 2018.