REVIEW ARTICLE

The Path Towards a Tailored Clinical Biosimilar Development

Martin Schiestl[1] · Gopinath Ranganna[2] · Keith Watson[3] · Byoungin Jung[4] · Karsten Roth[5] · Björn Capsius[6] · Michael Trieb[7] · Peter Bias[8] · Julie Maréchal-Jamil[9]

Abstract

Since the first approval of a biosimilar medicinal product in 2006, scientific understanding of the features and development of biosimilar medicines has accumulated. This review scrutinizes public information on development programs and the contribution of the clinical studies for biosimilar approval in the European Union (EU) and/or the United States (US) until November 2019. The retrospective evaluation of the programs that eventually obtained marketing authorization and/or licensure revealed that in 95% (36 out of 38) of all programs, the comparative clinical efficacy studies confirmed similarity. In the remaining 5% (2 out of 38), despite meeting efficacy outcomes, the biosimilar candidates exhibited clinical differences in immunogenicity that required changes to the manufacturing process and additional clinical studies to enable biosimilar approval. Both instances of clinical differences in immunogenicity occurred prior to 2010, and the recurrence of these cases is unlikely today due to state-of-the-art assays and improved control of process-related impurities. Biosimilar candidates that were neither approved in the EU nor in the US were not approved due to reasons other than clinical confirmation of efficacy. This review of the development history of biosimilars allows the proposal of a more efficient and expedited biosimilar development without the routine need for comparative clinical efficacy and/or pharmacodynamic studies and without any compromise in quality, safety, or efficacy. This proposal is scientifically valid, consistent with regulation of all biologics, and maintains robust regulatory standards in the assessment of biosimilar candidates. Note: The findings and conclusion of this paper are limited to biosimilar products developed against the regulatory standards in the EU and the US.

1 Introduction

The regulatory pathway for biosimilars first established in the European Union (EU) in 2004, enabled the development and approval of biosimilar medicines, based on comparative

✉ Martin Schiestl
martin.schiestl@sandoz.com

1   Sandoz GmbH, Biochemiestrasse 10, 6336 Langkampfen, Austria

2   Mylan Pharm. Private Ltd., Bengaluru, India

3   Celltrion Inc., Incheon, Republic of Korea

4   Samsung Bioepis Co. Ltd., Incheon, Republic of Korea

5   Polpharma Biologics S.A., Warsaw, Poland

6   Formycon AG, Martinsried, Germany

7   bioeq GmbH, Holzkirchen, Germany

8   Teva Pharm. Ind. Ltd., Ulm, Germany

9   IGBA, Geneva, Switzerland

analytical, non-clinical, and clinical data that confirmed them to be equally as safe and effective as their reference products [1–3]. The first biosimilar medicinal product, Omnitrope (somatropin), was approved by the European Commission following a positive European Medicines Agency (EMA) opinion in 2006. As of November 1, 2019 there were more than 60 biosimilar products, corresponding to 18 specific drug substances (INN) approved in the EU and 23 biosimilar products to nine drug substances approved in the United States (US). The current market experience demonstrates that all biosimilars approved in the EU and US perform as expected, that is, they provide the same clinical benefit and no unexpected adverse events have occurred [3, 4]. The scientific principles for biosimilar development and review were not invented from scratch but were built on experience and regulatory history of originator biologics, especially the comparability concept for regulating process manufacturing changes that was developed in the 1990s and cumulated in the ICH Q5E guideline [5–7]. A review of monoclonal antibodies approved in the EU revealed an average of 1.8 manufacturing process changes per product per year, and clinical trials are rarely done to support these

Key Points

The contribution of clinical studies for all biosimilar approvals in the European Union and in the USA until November 2019 was evaluated, based on information from European Public Assessment Reports and FDA reviews.

For 95% (36 out of 38) of biosimilar development programs, the comparative efficacy studies[1] added no value to the scientific review process. In the remaining 5% (2 out of 38) of the development programs, the comparative efficacy studies confirmed equivalent efficacy but failed to demonstrate comparable immunogenicity, and subsequently required further optimization of the manufacturing process to improve product quality prior to obtaining approval.

This experience allows a proposal for a tailored clinical biosimilar development paradigm without routine need for comparative efficacy and/or pharmacodynamic studies, while maintaining regulatory robustness.

The challenge for regulators and industry is to apply their collective experiences and retrospective knowledge of biosimilar development and regulatory reviews to waive comparative efficacy studies and rely on data from analytical and in vitro functional assays and clinical pharmacokinetic equivalence studies that incorporate an immunogenicity assessment.

changes [8]. However, the biosimilar regulatory framework was initially developed with the conservative stance that comparative efficacy studies should typically be expected. Since the physiological function of a protein is defined by its structure, high analytical similarity is the fundamental requirement for biosimilars. This is consistent with the consensus regulatory opinion in the EU and US that subsequent clinical studies cannot compensate for the absence of an analytical match. Therefore, the role of clinical studies is to confirm that this match was established. The required extent of this clinical confirmation is subject to opinion and consequently regulators adopted a more flexible approach. Soon the EMA, followed by the US Food and Drug Administration (US-FDA), stated in their biosimilar guidelines that comparative efficacy studies could be waived under certain circumstances, specifically when suitable pharmacodynamic (PD) markers exist [9, 10].

---

[1] The term 'comparative efficacy studies' as used in this paper is short for 'comparative clinical efficacy studies in patients or healthy volunteers.'

The evolution of this regulatory requirement is illustrated by the approval history of filgrastim biosimilars. The first filgrastim biosimilar—Ratiograstim—was approved by the European Commission in 2008 and was supported by comparative efficacy studies. But in 2009, the EU approved Zarzio, another biosimilar of filgrastim, without requiring a comparative efficacy study [11, 12]. The clinical confirmation of biosimilarity for Zarzio was achieved by a series of comparative pharmacokinetic (PK)/PD studies and a single-arm, open-label safety/immunogenicity study. Interestingly, the same biosimilar required an additional comparative efficacy study to obtain US-FDA approval (Zarxio; filgrastim-sndz) in 2015 [13]. As the US-FDA has gained more experience assessing biosimilars, it has also adopted a more flexible approach to comparative efficacy studies in certain cases. Finally, in 2018, both the EU and US-FDA approved pegfilgrastim biosimilars without any confirmatory efficacy trials [14–16]. The approval history of filgrastim biosimilars shows the growing confidence of EU and US regulators to tailor clinical development programs without the need for comparative efficacy trials where suitable biomarkers exist. However, the option for waiving comparative efficacy trials for many biosimilar candidates including most monoclonal antibodies (mAb) is limited because suitable biomarkers, which would typically be required, are often not available or described in the literature.

To this point, recent publications have elaborated on whether confirmatory efficacy studies are always necessary, even in the absence of suitable PD markers, and have refueled the discussion on how to further tailor clinical development programs for biosimilars [17, 18]. This article provides a retrospective evaluation of the clinical development programs for all biosimilar products either approved, refused, or withdrawn in both the EU and US between April 2006 and November 2019. This information supports our proposal that a more tailored clinical biosimilar development program without the need for comparative efficacy trials could be adopted without jeopardizing scientific rigor or compromising regulatory assessment or standards.

2 Methods

We assessed all clinical trial results from biosimilar PK and comparative efficacy trials that were disclosed in European Public Assessment Reports (EPAR) and US-FDA drug review summaries. The information was retrieved from the EMA and US-FDA websites for all biosimilars from April 11, 2006 (EU marketing authorization date of Omnitrope) to November 1, 2019. This data was complemented with other publicly available sources including scientific literature and company statements.

Programs that eventually received marketing authorization and/or licensure in at least the EU or US were classified as below, while programs that failed achieving at least one approval are discussed individually.

The PK studies were reviewed and classified into the following categories:

(a) PK studies that met their PK equivalence margins;
(b) PK studies that initially failed to meet their PK equivalence margins but obtained approval either after repetition of studies and/or with justification and post-hoc analysis.

Each comparative efficacy study was reviewed and classified as follows:

(a) Efficacy studies that met primary endpoints and showed comparable safety/immunogenicity;
(b) Efficacy studies that failed to meet their primary endpoints, but obtained approval either with post-hoc analysis and/or additional scientific justification;
(c) Efficacy studies that failed to demonstrate comparable immunogenicity, and required further optimization of the manufacturing process to improve product quality prior to obtaining approval.

A biosimilar development program is a set of activities undertaken with the goal of obtaining a biosimilar marketing authorization in the EU or licensure in the US. Each development program relates to a specific drug substance (INN, e.g. infliximab, filgrastim) but can result in multiple marketing authorizations in the EU for the same product with different trade names. To avoid duplicate counting of individual clinical studies the data in the results section are based solely on the number of biosimilar development programs, not the total number of products granted marketing authorizations/licensures.

The number of study participants in this review represents the healthy volunteers or patients who were dosed with the biosimilar candidate or reference product at least once in the respective study. In cases where the number of dosed subjects was not available, the number of randomized study participants was used.

3 Results and Discussion

Across the review period there were 45 biosimilar drug substance (INN) development programs reviewed. Of these, 42 programs obtained marketing authorizations in the EU and 23 were licensed in the US. Three out of 45 programs did not get an EU marketing authorization nor US licensure and will be discussed later in the text. All of the 42 approved programs had conducted at least one PK study (Table 1), while 38 programs conducted comparative efficacy and safety studies (Table 2). The source data and references to all programs cited in Tables 1 and 2 are provided in Supplementary Table 1 (see electronic supplementary material).

3.1 Evaluation of Pharmacokinetic Studies

Table 1 summarizes the PK studies that were part of EU and/or US approved biosimilar development programs. In six (14%) out of the 42 programs, at least one PK study failed to show bioequivalence in relevant study endpoints. The drug substance (INN) and number of programs that did not meet relevant PK study endpoints were adalimumab (2), filgrastim (1), pegfilgrastim (2), and etanercept (1). The probable reasons for not showing bioequivalence included problems in study design or underestimated variability of the serum concentrations [12–15, 19–22]. For one adalimumab program that had a PK study failure, the company demonstrated that inadequate stratification of the study population with respect to their propensity for developing anti-drug antibodies influenced serum product levels and bioequivalence [23].

Clinical PK is the only test to assess and compare the combined impact of protein, device, and formulation on systemic exposure. This test is especially important as both device and formulation may differ between the biosimilar and its reference product. In 14% of the finally successful development programs, at least one PK study did not meet the primary endpoints, which shows its discriminatory

Table 1 Summary of clinical PK studies of development programs that led to marketing approval

Category	No. (%) of corresponding biosimilar development ($N=42$)	No. of studies	No. of patients/healthy volunteers
a. PK studies that met their PK equivalence margins	36 (86%)	82	8788
b. PK studies that initially failed to meet their PK equivalence margins but obtained approval either after repetition of studies and/or with justification and post-hoc analysis	6 (14%)	7	995

PK pharmacokinetic

power. These failures were shown to be due to methodological issues and not related to different product quality. Accordingly, meeting PK equivalence is a strong confirmation of biosimilarity. On the other hand, a failure of PK endpoints requires investigation to determine the root cause, which could be either a study design issue or a true difference that might be clinically relevant.

To this point, following a review of their own biosimilar experience to date, EU regulators concluded that PK studies will remain a sine qua non in biosimilar development [3].

3.2 Evaluation of Comparative Efficacy Studies

Evaluation of clinical comparative efficacy studies of development programs that led to marketing approval are summarized in Table 2.

Of the 42 approved biosimilar development programs, 38 had a clinical efficacy study. Four biosimilar programs—two pegfilgrastims, one enoxaparin, and one teriparatide—were approved in the EU without comparative efficacy studies, but were instead approved with a data package that included clinical studies conducted with suitable biomarkers. It is important to note that enoxaparin and teriparatide are not regulated as biologics in the US and follow-on versions were approved not as biosimilars but as drugs in the US. The majority of the biosimilar development programs with comparative efficacy studies [33 (87%); Table 2, category a] met primary efficacy endpoints and showed comparable safety/immunogenicity.

Another three (8%) programs initially failed primary endpoints in the comparative efficacy studies but still obtained biosimilar approval after post-hoc analysis and/or scientific justification (Table 2, category b). Two of these were trastuzumab biosimilar development programs. According to their EPARs, the comparative efficacy studies did not formally meet the primary endpoint and superiority of biosimilar against the reference product could not be excluded [24, 25]. For the US-FDA assessment, this was only raised in the ABP980 program, whereas the efficacy trial of the SB3 program met the primary endpoint, presumably because the US-FDA and EMA had different expectations for the analysis plan of the same study [26, 27]. During the EMA assessment, both companies argued that the issue was caused by a temporary reduction in antibody-dependent cellular cytotoxicity (ADCC) potency in a substantial number of reference product batches used in the clinical trial [24, 25]. In the reference product arms, 40% (program SB3) and 20% (program ABP980) of the patients received batches with low ADCC. The argument was made that this numerically impacted the primary endpoint, shifting the confidence interval to where it was slightly above the predefined equivalence margin. This explanation was supported by post-hoc analysis, which excluded patients treated with low ADCC reference product batches, for example. The shift in ADCC activity was observed in reference product batches that had expiry dates between 2018 and 2019 and was confirmed with in vitro functional assays and by quantifying the amount of afucosylated glycans present in the Fc domain of mAb because this glycan moiety is known to impact ADCC function [28, 29]. These findings point to a presumably unintended variation in the reference product by the reference product manufacturer and the importance in general of strict adherence to ICH Q5E and adequate manufacturing controls for all biologics. While some level of batch-to-batch variability is inevitable, it is important that this variability remain within acceptable ranges to avoid any detrimental impact on clinical outcome [30]. The trastuzumab biosimilar examples show that for products with substantial differences in critical quality attributes (CQAs) linked to a contributory mechanism of action, comparative efficacy studies can be sensitive enough to detect a difference in clinical endpoints. This example demonstrates the value of efficacy studies in confirming biosimilarity, however, it also shows that physicochemical methods and functional assays are able to detect differences in functional attributes with much greater sensitivity. In this case, a reduction in ADCC activity was caused by an increase in the amount of afucosylated glycans, which likely led to a detectable difference in efficacy outcomes.

Table 2 Summary of clinical comparative efficacy studies of development programs that led to marketing approval

Category	No. (%) of corresponding biosimilar development programs ($N=38$)	No. of studies	No. of patients
a. Efficacy studies that met primary endpoints and showed comparable safety/immunogenicity	33 (87%)	55	22,338
b. Efficacy studies that failed to meet their primary endpoints, but obtained approval either with post-hoc analysis and/or additional scientific justification	3 (8%)	3	2189
c. Efficacy studies that failed to demonstrate comparable immunogenicity, and required further optimization of the manufacturing process to improve product quality prior to obtaining approval	2 (5%)	2	426

The third case in category b impacted one pegfilgrastim biosimilar development program [31]. A comparative efficacy trial in breast cancer patients compared the biosimilar candidate with EU-sourced and US-sourced reference product. The study met its primary endpoint in demonstrating equivalent efficacy between the biosimilar candidate and the EU reference product. However, the primary endpoints did not show equivalence between EU- and US-sourced reference product and consequently between the biosimilar and US-sourced reference product. The biosimilar was approved in the EU because all requirements for comparing the biosimilar and the EU reference product were met, including a PK/PD study with EU-sourced reference product [31]. The biosimilar has not yet been approved in the US within the data cut-off date of this review [32]. However, three other independent pegfilgrastim biosimilar development programs led to approvals in both the EU and US [14, 15, 21, 32, 33]. All three programs confirmed similarity of EU and US reference product by extensive bridging data. Therefore, it is likely that the failure of the bridging study between EU and US reference products of the pegfilgrastim program mentioned above was due to other issues of the comparative efficacy trial, rather than any clinically relevant differences between EU and US reference products.

In two biosimilar development programs, efficacy studies failed to demonstrate comparable immunogenicity and required further optimization of the manufacturing process to improve product quality and to enable biosimilar approval (Table 2, category c).

The first example is a biosimilar somatropin [34]. The comparative efficacy study performed in 1999 confirmed that clinical efficacy endpoints were all met but there were higher rates of immunogenicity with the biosimilar candidate. The root cause analysis revealed a correlation between immunogenicity rate and higher amounts of host cell protein (HCP) impurities in the biosimilar, which were not detected by the commercial HCP assay used for process development and clinical trial batch release. Subsequently, the manufacturing process was optimized to purge the HCP impurities and further clinical studies confirmed comparable immunogenicity rates between the reference product and the biosimilar. The improved product quality together with the confirmatory clinical data enabled approval in 2006.

The second case affected a biosimilar epoetin [35, 36]. A comparative efficacy study, undertaken to complete the data package for gaining approval for subcutaneous (SC) treatment of chronic kidney disease patients, was stopped in 2009 after two patients developed neutralizing antibodies. The root cause analysis revealed that residual tungsten in the syringe, once in contact with epoetin solution, catalyzed the formation of insoluble aggregates, which are thought to increase the risk of immunogenicity. The clinical comparative efficacy trial was successfully repeated with clinical study material filled in low-tungsten syringes, which enabled the approval of the SC administration in the EU in 2016 [37].

Both cases of category c were traced to the presence of elevated levels of process-related impurities and not product-related impurities. The presence of HCP impurities are dependent on a number of factors including cell line, fermentation conditions, and the manufacturing process whilst the ability to detect and quantify them is dependent on the sensitivity and selectivity of the assay. From a regulatory perspective, expectations are that the amount of HCP impurities in the drug substance should be as low as possible. The HCP issue seen with somatropin is unlikely to happen again because the state-of-the-art in HCP control has increased substantially in the last two decades. For example, there is now greater understanding of the risk of HCP impurities, there are better assays for detection and pharmacopeia guidelines available for HCP analysis [38, 39]. Another research group also investigated the issue of residual tungsten originating from syringe manufacturing and how it could induce protein aggregation [40]. Learning from the past and with additional knowledge, a repetition of the epoetin case is unlikely.

In general, the control of process-related impurities and prevention of unwanted immunogenicity is required for all biologics throughout their product life cycle, including manufacturing changes. Extensive knowledge of these impurities and other risk factors for unwanted immunogenicity help to design manufacturing controls and regulatory oversight to achieve comparable low immunogenicity [41, 42].

3.3 Biosimilar Development Programs that Did Not Receive Approval or are Currently on Hold in the EU and/or US

In an attempt to counter the sampling bias of this analysis as it contains all biosimilar programs that received at least one approval in the EU or US, it is also important to evaluate the contribution of clinical data of those programs that did not result in biosimilar approval in any of those regions. We found three published examples of programs that entered the clinical stage of development and are currently on hold, received a negative opinion from the regulators and were not approved, or where the company chose to withdraw their application before the end of the formal review process.

- In the EU, an interferon alfa biosimilar candidate received a negative EMA opinion in 2006. Most importantly, the analytical and functional data was not deemed by the regulators to be comparable between the reference and biosimilar products. Furthermore, Study 002 for this proposed biosimilar showed highly anomalous PK data. There were also uncertainties in PD equivalence,

especially in viral response that could not be resolved. Nonetheless, the primary and secondary endpoints of the comparative efficacy trial were met [43].

- In the EU, an insulin biosimilar candidate was withdrawn by the company in 2012. The analytical and functional data was not deemed by regulators to be comparable between the reference and biosimilar products. In this case, comparative clinical PK studies demonstrated bioequivalence, there were similar PD outcomes, and the supportive comparative efficacy studies showed similar efficacy and safety. However, in addition to the lack of analytical and functional comparability there were other relevant good manufacturing practice (GMP) issues and the clinical trial material was not demonstrated to be representative of the proposed commercial product [44].

- An abatacept biosimilar candidate missed the primary endpoint in a three-arm PK study with US- and EU-sourced reference product. The program is on hold [45]. No further information was found in the public domain.

In summary, in the EU, both the interferon alfa and insulin biosimilar candidates failed to demonstrate analytical comparability and the abatacept candidate failed to show PK bioequivalence with EU- and US-sourced reference product. Therefore, these cases demonstrated that issues in showing biosimilarity were identified either at the analytical or clinical PK level, prior to entering the comparative efficacy trial. The interferon example also illustrates that a successful comparative efficacy study cannot compensate for gaps in the analytical data.

4 A Path Forward

High similarity of the physicochemical and functional properties as well as the equivalent distribution of the product in the human body provide sufficient evidence to obviate the need for comparative efficacy studies in most cases. Our retrospective evaluation of clinical studies supporting biosimilar development programs in the EU and US revealed that the efficacy endpoints in comparative efficacy studies added no value to the successful biosimilar development programs. For two development programs, efficacy trials revealed differences in immunogenicity due to process-related impurities but not differences per se in the efficacy of the molecule. Therefore, these data question the scientific value of undertaking comparative clinical studies powered for efficacy endpoints to support biosimilar development. This analysis also demonstrates the need for a considered review of the current biosimilar development paradigm. We propose a biosimilar clinical development program that may exclude a comparative efficacy and/or PD study as illustrated

in Fig. 1. It includes a stepwise evaluation of the physicochemical and functional CQAs.

4.1 Evaluation of the Physicochemical Critical Quality Attributes (CQAs)

The guidelines in both the EU and US are clear on the need to apply sensitive, state-of-the-art analytical methods to characterize and where appropriate, quantify all clinically relevant physiochemical attributes of a biological product. As part of this effort, identification and selection of the CQAs that may have potential impact on product quality and resulting clinical performance is required. These include primary and higher order structure, product-related variants (size, charge, glycans, aggregation, etc.), process-related impurities (host cell proteins, DNA, leachables, endotoxin, etc.) and other obligatory CQAs of the presentation (pH, appearance, protein concentration, etc.). The identification and selection of CQAs and the linkages to product safety and efficacy follows a systematic risk-based approach that became regulatory expectation for the biopharmaceutical industry with the implementation of the ICH guidelines Q8, Q9, and Q10 more than a decade ago [46].

4.2 Evaluation of the Functional CQAs

The functions of a protein are defined by its structure, that is, physicochemical attributes. Therefore, evaluation of the functional properties by a panel of sensitive bioassays establishes all relevant in vitro biological activities are virtually indistinguishable between biosimilar and reference product and also confirms that the physicochemical CQAs are properly characterized and controlled. Knowledge of clinical relevance of a specific functional attribute is helpful in biosimilar development and evaluation (e.g. to set the appropriate criticality and risk score); however, it is not required. In other words, if the link of a certain protein function to safety and efficacy is unknown, this attribute automatically gets a high criticality score, less latitude for differences is allowed, and the biosimilar needs to match the quality range of the reference product.

4.3 Comparing CQAs Between Biosimilar Candidate and Reference Product

It is highly unlikely that minor differences in physicochemical and functional attributes will be detected in a less sensitive comparative clinical study powered for efficacy endpoints. Consequently, the consensus opinion of both the US-FDA and EMA is that clinical trials cannot be used to justify differences in physicochemical attributes. The logical consequence is to closely match the reference product at the physicochemical and functional level, while the variability

Fig. 1 Decision tree for tailoring biosimilar development. *CQA* critical quality attribute, *PK* pharmacokinetic, *PD* pharmacodynamic
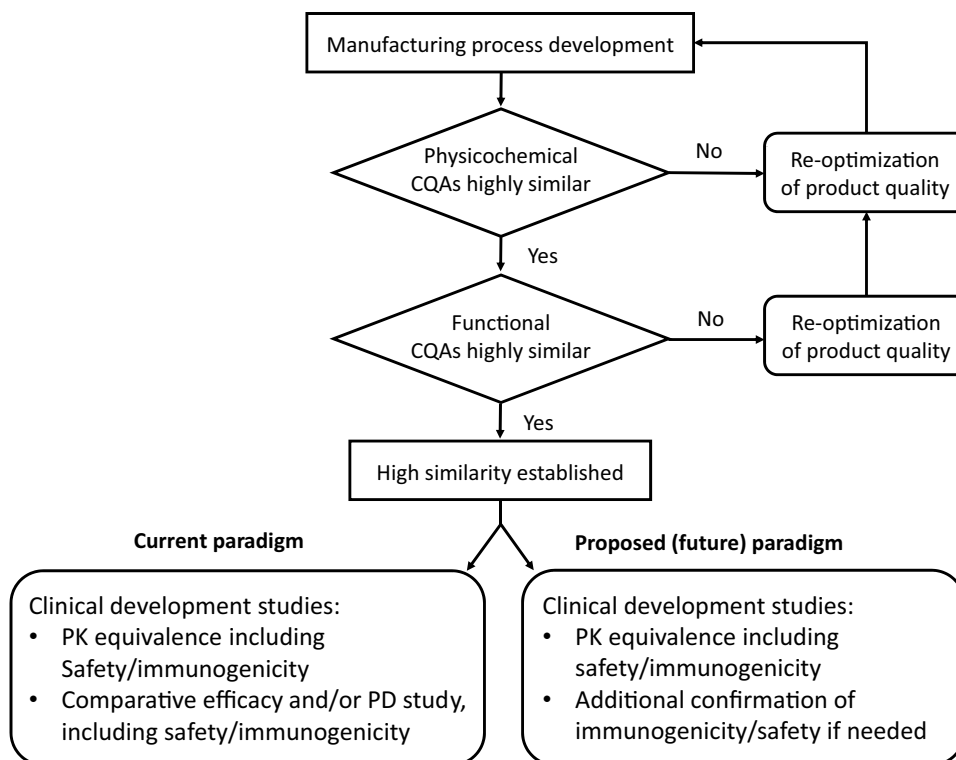
of the reference product, as already accepted by the regulators, sets the goalposts for the biosimilar candidate. Constant attributes, such as the amino acid sequence, need to be identical. For CQAs that vary between batches, the biosimilar needs to match the batch-to-batch variability of the reference product unless it can be justified that observed differences are clinically meaningless. The US-FDA recently published a quality range approach to compare these quality attributes, which builds on today's regulatory science and how it has been applied to control manufacturing processes, including manufacturing changes of biologicals in general [47].

Whether differences in physicochemical CQAs are potentially clinically relevant can be assessed by measuring their impact on different functions using a panel of functional assays. If differences in physicochemical and consequently functional CQAs cannot be justified, it is incumbent to re-optimize the manufacturing process to ensure that the physicochemical attributes and functional activities of the biosimilar are better aligned with the reference product. A complete bioassay toolbox is therefore a key enabler for applying the proposed clinical development paradigm. The toolbox requires multiple assays, ideally cell-based, to cover all relevant functions of a molecule with accurate and precise quantitative read-outs, and agreement with the regulators on the bioassay designs including their validation.

Using mAbs as examples, after decades of research, a well-equipped bioassay toolbox is now available that measures the functional outcomes of a mAb binding to its target

antigen, either soluble or cell bound, as well as to Fc receptors that may involve overlapping surface regions of the mAb [48, 49]. Measuring a multitude of functions mitigates situations where a specific receptor/target interaction of biosimilar and reference product might be different and influence efficacy.

4.4 Decision Tree for Tailoring Clinical Biosimilar Development

As shown in Fig. 1, the decision tree highlights the current and proposed clinical biosimilar development paradigm. The cornerstone of demonstrating biosimilarity is the demonstration of physicochemical and functional similarity. Furthermore, clinical PK equivalence is an important confirmation of the comparable combined effect of protein, device, and formulation, especially because device and formulation may differ. Accordingly, PK studies likely remain a typical requirement in biosimilar development. An immunogenicity risk assessment that considers product-and patient-related risk factors and includes data derived from comparative PK studies will inform on the need and extent of an additional safety study. The latter would not be necessary for products where the immunogenicity risk is deemed low and can be addressed with PK studies and/or appropriate impurity data, such as filgrastim or insulin [50–52].

It might appear radical to grant a waiver for both efficacy and PD trials in the new paradigm. Although a PD

biomarker study can be a useful confirmation of biosimilarity, it represents an earlier measure of a clinical physiological response that is required for efficacy and therefore serves the same purpose as a comparative efficacy trial. If sufficient analytical and PK data are available to allow a robust conclusion of biosimilarity without an efficacy trial, this equally applies to PD biomarker studies.

Our retrospective analysis revealed two cases where process impurities were poorly controlled. Although comparison of process-related impurities is not formally part of a biosimilar development program, it is incumbent on all biological product manufacturers to limit and control process-related impurities in the drug substance. A critical learning from our retrospective analysis in this regard is that with today's quality standards and sensitive assays, the repetition of these two cases is unlikely. In support, no such case has been observed since 2010 in the many biosimilars that have been developed in the past decade.

It is challenging to apply retrospective knowledge to future decision-making processes. Accordingly, the new paradigm might be first applied prospectively to therapeutic protein types that are covered in this retrospective view, such as cytokines, growth factors, IgG1 monoclonal antibodies/fusion proteins, and related molecules. The application to other proteins depends on the available product knowledge, the ability to characterize all relevant CQAs, and the understanding of the targets and receptors that can interact with the protein under physiological conditions. The decision whether to conduct comparative efficacy and PD studies should be discussed at an early program stage and confirmed with regulators upon review of the results of the physicochemical and functional comparison. Regulators' assessments also benefit from their long experience with manufacturing process changes of biologics in general and reference products in particular [53].

5 Conclusions

The regulatory environment and scientific understanding of biosimilar medicines has advanced since initial establishment of biosimilar regulatory pathways over a decade ago. In the EU and US, robust regulatory standards exist that ensure approval of biosimilars that are as safe and efficacious as the reference product. With the vast experience and knowledge accrued through evaluation of over 42 unique biosimilar candidates, it is time to revisit the existing development paradigm and ensure that a tailored clinical approach is adopted that is supported by the science. Currently, a comparative efficacy study is routinely expected by the EMA and US-FDA to confirm biosimilarity, especially if no suitable PD markers exist that can be assessed in a PK/PD study. Since suitable PD markers are not available

for most protein therapeutics, particularly mAbs, the option to waive a comparative efficacy study has been limited in practice. The retrospective review of approved biosimilars in the EU and US revealed that in 100% (38 out of 38) of biosimilar development programs, the comparative efficacy studies confirmed efficacy of the biosimilar candidate. For 95% (36 out of 38) of biosimilar development programs, the comparative efficacy studies added no value to the scientific review process to approve a biosimilar. In the remaining 5% of the development programs (2 out of 38), the comparative efficacy studies showed that despite meeting efficacy outcomes, the biosimilar candidate exhibited clinical differences in immunogenicity that required changes to the manufacturing process and additional clinical studies prior to eventual biosimilar approval. Considering today's state-of-the-art assays and control strategies, a repetition of these two cases is unlikely, and has not been observed since 2010. These data confirm that state-of-the-art analytical methods that include physicochemical and functional assays, along with PK studies, best inform efficacy outcomes with no routine need for comparative efficacy trials. This approach may apply to complex proteins like monoclonal antibodies and will not jeopardize the regulatory standards in either the EU or US.

Compliance with Ethical Standards

Conflict of interest MS, GR, KW, BJ, KR, BC, MT, and PB are employed by companies that are developing and commercializing biosimilar medicines. JMJ is working for trade associations representing generic and biosimilar medicines manufacturers.

Ethical approval No patient data was accessed; therefore, no ethical review board review was required.

References

1. EU Directive 2004/27/EC. https://ec.europa.eu/health/files/eudralex/vol-1/dir_2004_27/dir_2004_27_en.pdf. Accessed Nov 2019.
2. Schiestl M, Zabransky M, Sörgel F. Ten years of biosimilars in Europe: development and evolution of the regulatory pathways. Drug Design Dev Ther. 2017;11:1509–15.
3. Wolff-Holz E, Tiitso K, Vleminckx, Weise M. Evolution of the EU biosimilar framework: past and future. 2019: https://doi.org/10.1007/s40259-019-00377-y.
4. Cohen HP, Blauvelt A, Rifkin RM, Danese S, Gokhale SB, Woollett G. Switching reference medicines to biosimilars: a systematic literature review of clinical outcomes. Drugs. 2018;78(4):463–78.
5. Concept paper on the development of a committee for proprietary medicinal products guideline on comparability of biotechnology-derived products. CPMP/BWP/1113/98, 24 June 1998. https://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003966.pdf. Accessed Feb 2020.
6. FDA Guidance on Demonstration of comparability of human biological products, including therapeutic biotechnology-derived products. April 1996. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstration-comparability-human-biological-products-including-therapeutic-biotechnology-derived. Accessed Feb 2020.
7. ICH Q5E Guideline on comparability of biotechnological/biological products subject to changes in their manufacturing process. 2004. https://database.ich.org/sites/default/files/Q5E_Guideline.pdf. Accessed Feb 2020.
8. Vezér B, Buzás Z, Sebeszta M, Zrubka Z. Authorized manufacturing changes for therapeutic monoclonal antibodies (mAbs) in European Public Assessment Report (EPAR) documents. Curr Med Res Opin. 2016;32(5):829–34.
9. Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: nonclinical and clinical issues EMEA/CHMP/BMWP/42832/2005 Rev. 1. (2014). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-similar-biological-medicinal-products-containing-biotechnology-derived-proteins-active_en-2.pdf. Accessed Nov 2019.
10. Li J, Florian J, Campbell E, Schrieber SJ, Bai JPF, Weaver L, et al. Advancing biosimilar development using pharmacodynamic biomarkers in clinical pharmacology studies. Clin Pharm Ther. 2019. https://doi.org/10.1002/cpt.1653.
11. EPAR Ratiograstim EMEA/502481/2008. https://www.ema.europa.eu/en/documents/assessment-report/ratiograstim-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
12. EPAR Zarzio EMEA/CHMP/651339/2008. https://www.ema.europa.eu/en/documents/assessment-report/zarzio-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
13. FDA Zarxio review, FDA application no 125553. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2015/125553Orig1s000TOC.cfm. Accessed Nov 2019.
14. EPAR Udenyca EMA/552721/2018. https://www.ema.europa.eu/en/documents/assessment-report/udenyca-epar-public-assessment-report_.pdf. Accessed Nov 2019.
15. FDA Udenyca review, FDA application no 761039. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/761039Orig1s000TOC.cfm. Accessed Nov 2019.
16. EPAR Pelmeg EMA/703393/2018. https://www.ema.europa.eu/en/documents/assessment-report/pelmeg-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
17. Frapaise FX. The end of phase 3 clinical trials in biosimilars development? Biodrugs. 2018;32:319–24.
18. Webster CJ, Wong A, Woollett GR. An efficient development paradigm for biosimilars. Biodrugs. 2019. https://doi.org/10.1007/s40259-019-00371-4.
19. EPAR Cyltezo EMA/CHMP/750187/2017. https://www.ema.europa.eu/en/documents/assessment-report/cyltezo-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
20. EPAR Hyrimoz EMA/CHMP/404076/2018. https://www.ema.europa.eu/en/documents/assessment-report/hyrimoz-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
21. EPAR Ziextenzo EMA/CHMP/706001/2018. https://www.ema.europa.eu/en/documents/assessment-report/ziextenzo-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
22. EPAR Erelzi EMA/CHMP/302222/2017. https://www.ema.europa.eu/en/documents/assessment-report/erelzi-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
23. VonRichter O, Lemke L, Haliduola H, Balfour A, Zehnpfennig B, Skerjanec A, Jauch-Lembach J. Differences in immunogenicity associated with non-product related variability: insights from two pharmacokinetic studies using GP2017, an adalimumab biosimilar. Exp Opin Biol Ther. 2019;10:1057–64. https://doi.org/10.1080/14712598.2019.1603959.
24. EPAR Ontruzant EMA/CHMP/9855/2018. https://www.ema.europa.eu/en/documents/assessment-report/ontruzant-epar-public-assessment-report_en-0.pdf. Accessed Nov 2019.
25. EPAR Kanjinti EMA/CHMP/261937/2018. https://www.ema.europa.eu/en/documents/assessment-report/kanjinti-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
26. FDA Kanjinti review application no 761073. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/761073Orig1s000TOC.cfm. Accessed Nov 2019.
27. FDA Ontruzant review application no 761100. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/761100Orig1s000TOC.cfm. Accessed Nov 2019.
28. Kim S, Song J, Park S, Ham S, Paek K et al. Drifts in ADCC-related quality attributes of Herceptin®: impact on development of a trastuzumab biosimilar. mAbs. 2017;9:704–14.
29. Lee JH, Paek K, Moon JH, Ham S, Song J, Kim S. Biological characterization of SB3, a trastuzumab biosimilar, and the influence of changes in reference product characteristics on the similarity assessment. BioDrugs. 2019;33:411–22.
30. Schiestl M, Stangler T, Torella C, Čepeljnik T, Toll H, Grau G. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat Biotechnol 2011;29(4):310–2.
31. EPAR Pelgraz EMA/595848. https://www.ema.europa.eu/en/documents/assessment-report/pelgraz-epar-public-assessment-report_en.pdf. Accessed Nov 2019.
32. FDA licensed biosimilars. https://www.fda.gov/drugs/biosimilars/biosimilar-product-information. Accessed Nov 2019.
33. EPAR Fulphila EMA/724003/2018. https://www.ema.europa.eu/en/documents/assessment-report/fulphila-public-assessment-report_en.pdf. Accessed Nov 2019.
34. EPAR Omnitrope EMA 24 April 2006. https://www.ema.europa.eu/en/documents/scientific-discussion/omnitrope-epar-scientific-discussion_en.pdf. Accessed Nov 2019.
35. Seidl A, Hainzl O, Richter M, Fischer R, Böhm S, Deutel B, et al. Tungsten-induced denaturation and aggregation of epoetin. Pharm Res. 2012;29(6):1454–67.
36. Rubic-Schneider T, Kuwana M, Christen B, Aßenmacher M, Hainzl O, Zimmermann F, et al. T-cell assays confirm immunogenicity of tungsten-induced erythropoietin aggregates associated with pure red cell aplasia. Blood Adv. 2017;1:367–79.
37. EPAR Binocrit EMA/590338/201. https://www.ema.europa.eu/en/documents/procedural-steps-after/binocrit-epar-procedural-steps-taken-scientific-information-after-authorisation_en.pdf. Accessed Nov 2019.

38. USP-NF, General chapter ⟨1132⟩ residual host cell protein measurements in biopharmaceuticals, Official as of Dec 2015. https://www.usp.org. Accessed Nov 2019.

39. European Pharmacopoeia, General chapter 2.6.34 host-cell protein assays, effective April 2017. https://www.edqm.eu. Accessed Nov 2019.

40. Liu W, Swift R, Torraca G, Nashed-Samuel Y, Wen ZQ, Jiang Y, et al. Root cause analysis of tungsten-induced protein aggregation in pre-filled syringes. PDA J Pharm Sci Tech. 2010;64:11–9.

41. Büttel IC, Chamberlain P, Chowers Y, Ehmann F, Greinacher A, Jefferis R, et al. Taking immunogenicity assessment of therapeutic proteins to the next level. Biologicals. 2011;39:100–9.

42. Singh SK. Impact of product-related factors on immunogenicity of biotherapeutics. J Pharm Sci. 2011;100(2):354–87.

43. EPAR Alpheon EMA 2006. https://www.ema.europa.eu/en/documents/assessment-report/alpheon-epar-refusal-public-assessment-report_en.pdf. Accessed Nov 2019.

44. EPAR Solumarv EMA/596513/2015. https://www.ema.europa.eu/en/documents/assessment-report/solumarv-epar-public-assessment-report_en.pdf. Accessed Nov 2019.

45. Press release, Momenta and Mylan Report Initial Results from Phase 1 Clinical Trial for M834, a Proposed Biosimilar of ORENCIA® (abatacept). Nov 2017. https://ir.momentapharma.com/static-files/5c04118f-2e2f-4993-acc9-3d49bf6060ac. Accessed Nov 2019.

46. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. Nat Biotechnol. 2009;27:26–34.

47. FDA draft guidance on Development of therapeutic protein biosimilars: comparative analytical assessment and other quality-related considerations. May 2019. https://www.fda.gov/media/125484/download. Accessed Feb 2020.

48. Cymera F, Becka H, Rohde A, Reusch D. Therapeutic monoclonal antibody N-glycosylation—structure, function and therapeutic potential. Biologicals. 2018;52:1–11.

49. Prior S, Hufton SE, Fox B, Dougall T, Rigsby P, Bristow A. International standards for monoclonal antibodies to support pre- and post-marketing product consistency: evaluation of a candidate international standard for the bioactivities of rituximab. MAbs. 2018;10(1):129–42.

50. EMA guideline on similar biological medicinal products containing recombinant granulocyte-colony stimulating factor (rG-CSF). 2018. EMEA/CHMP/BMWP/31329/2005 Rev 1. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-similar-biological-medicinal-products-containing-recombinant-granulocyte-colony_en.pdf. Accessed Nov 2019.

51. EMA guideline on non-clinical and clinical development of similar biological medicinal products containing recombinant human insulin and insulin analogues. 2015, EMEA/CHMP/BMWP/32775/2005_Rev. 1. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-non-clinical-clinical-development-similar-biological-medicinal-products-containing_en-0.pdf. Accessed Nov 2019.

52. FDA Draft guidance on Clinical immunogenicity considerations for biosimilar and interchangeable insulin products. Nov 2019. https://www.fda.gov/media/133014/download. Accessed March 2020.

53. Kurki P, Van Aerts L, Wolff-Holz E, Giezen T, Skibeli V, Weise M. Interchangeability of biosimilars: a European perspective. BioDrugs. 2017;31:83–91.