



Evolution of the EU Biosimilar Framework: Past and Future

Elena Wolff-Holz¹ · Klara Tiitso² · Camille Vleminckx² · Martina Weise³

Published online: 20 September 2019
© The Author(s) 2019

Abstract

The approval of biosimilars in the EU follows a comprehensive scientific assessment based on stringent regulatory standards. While the initial approach to biosimilars was understandably cautious and conservative in that uncharted territory to protect patients' safety, the analytical and scientific progress and accumulated experience with biosimilars continues to reshape regulatory requirements, generally leading to a reduced burden of clinical trials. This trend is expected to continue, for example, by increasingly employing pharmacodynamic endpoints and biomarkers, but much work remains to make this happen, especially for complex molecules with several or unknown mechanisms of action. We reviewed the available guidance and European Public Assessment Reports (EPARs) of biosimilars approved in the EU via the centralised procedure. This review focuses on the nature and extent of clinical confirmation of biosimilarity considered necessary in addition to analytical and functional data. Cases with conflicting results from different parts of the comparability exercise are discussed, with the aim of identifying whether certain elements of the comparability exercise are more important than others in determining biosimilarity. Taken together, analytical and functional comparison is the foundation of any biosimilar development. In addition, pharmacokinetic similarity is an indispensable prerequisite for any biosimilar approval, so careful planning on behalf of the applicant is mandated to avoid potential failure of such studies, for example, because of large interindividual variability, underpowered trial designs or other methodological causes. Comparative pharmacokinetic studies are a basic requirement for biosimilar development and are usually more sensitive than clinical efficacy trials when detecting potential product-related differences. This may explain why a demonstration of equivalent efficacy could not overrule a finding of dissimilar pharmacokinetic profiles in two cases of biosimilar pegfilgrastim. However, the outcome of efficacy trials depends not only on drug exposure but also on proper pharmacological action of the biological substance *in vivo*. Therefore, the objectives of both types of studies differ. Efficacy trials should usually be designed as equivalence trials to ensure that the efficacy of the biosimilar is neither decreased nor increased compared with the reference product. However, some remaining uncertainty regarding potentially increased efficacy of the biosimilar may be acceptable in exceptional cases, provided that the data from other parts of the comparability exercise clearly support a conclusion of biosimilarity and safety is assured. In contrast, uncertainties regarding potentially inferior efficacy of the biosimilar may not be acceptable at all. We conclude that the EU biosimilar regulatory framework is robust and able to adapt to advancing knowledge and experience and to strike a balance between regulatory standards, patient safety and feasibility of biosimilar development.

✉ Elena Wolff-Holz
Elena.Wolff-Holz@pei.de

Klara Tiitso
Klara.Tiitso@ema.europa.eu

Camille Vleminckx
Camille.Vleminckx@ema.europa.eu

Martina Weise
Martina.Weise@bfarm.de

¹ Paul-Ehrlich-Institut, Federal Institute for Vaccines and Biomedicines, PEI Str 51-58, 63207 Langen, Germany

² European Medicines Agency (EMA), Domenico Scarlattilaan 6, 1083 HS Amsterdam, The Netherlands

³ Federal Institute for Drugs and Medical Devices, Kurt-Georg-Kiesinger-Allee 3, 53175 Bonn, Germany

Key Points

Case studies are presented based on a review of European Public Assessment Reports from July 2012 to July 2019 to illustrate complex situations when assessing the totality of data during marketing authorization procedures, i.e. how to weigh conflicting results from different parts of the comparability exercise.

The pharmacokinetic study is emerging as a major gatekeeper in the clinical biosimilarity exercise, a hurdle that needs to be overcome and all results fully justified before further clinical data can be deemed acceptable.

The experience acquired with each biological product class, together with advances in terms of characterization techniques and progress in understanding pharmacology and disease pathogenesis, have paved the way towards reduction of clinical data requirements and are expected to continue doing so.

1 Introduction

Biosimilars available in the EU and other International Conference on Harmonisation (ICH) regions are highly regulated medicinal products of proven pharmaceutical quality that have undergone a comprehensive developmental programme and an extensive regulatory scientific assessment process before their authorisation.

Since the introduction of the legal pathway for licensing of biosimilars in 2004 and up to 17 July 2019, a total of 91 marketing authorisation applications (MAAs) for biosimilar products have been submitted to the European Medicines Agency (EMA). This includes duplicates, as applicants can in certain cases obtain more than one marketing authorisation for the same medicinal product under different brand names. Of the 91 MAAs submitted, 61 biosimilars have received a positive opinion from the EMA's Committee for Medicinal Products for Human Use (CHMP) and subsequently been authorised by the European Commission (EC), 11 MAAs are currently under review and 19 were either withdrawn during the review process or received a negative opinion from the CHMP. Seven biosimilars were withdrawn after approval by the respective marketing authorisation holders for commercial reasons, none because of safety issues. The 54 currently valid biosimilar marketing authorisations in Europe cover 15 distinct reference medicinal products. Of note, the number of biosimilar MAAs that have either been rejected or withdrawn by the company during the CHMP review process could be an indication of the stringent entry barrier for applicants to obtain approval and reach the market. Information regarding products under

review, CHMP opinions and details of approved, rejected and withdrawn products, including European Public Assessment Reports (EPARs), is available on the EMA website [1].

The foundation of demonstrating biosimilarity is the in-depth comparison of the biosimilar and reference (originator) product at the analytical and functional level, followed by confirmation of biosimilarity by clinical data. The so-called comparability exercise involves a comprehensive comparison of structural/physicochemical and functional attributes using multiple lots of the proposed biosimilar product and the reference product and, usually, final confirmation of similar efficacy and safety, including immunogenicity, in a randomized, double-blind, controlled clinical trial in an approved indication. The same principles of comparability are applied as for approved biological products that undergo manufacturing process changes [2–4]. In most cases of changes to the manufacturing process, analytical and functional comparability testing is considered sufficient, but—in some cases—it may also be necessary to generate clinical data. ICH guideline Q5E, which addresses comparability of biotechnological/biological products subject to changes in their manufacturing process, reflects that such data may be required to address any residual uncertainties [2]. As a biosimilar development could be seen as an exceptional or extreme case of a manufacturing process change, including different manufacturing sites, manufacturing processes and starting materials (e.g. cell banks), it is not surprising that clinical data are usually required.

It should be noted that, once the marketing authorisation has been granted, the same requirements apply to biosimilars as to any other biological product. All biological products should be appropriately controlled with tight specifications to avoid any major shifts or drifts that may impact on the clinical performance. Process changes should be controlled by demonstrating comparability in line with ICH Q5E. Variability of attributes may occur but will need to be corrected to previous levels if deemed too high by regulators. Specifications (both the type and the range) can evolve with advanced knowledge, but at any time a robust system is in place to control potential variability of biologicals. There is no regulatory requirement to repeat the demonstration of biosimilarity against the reference product once the marketing authorisation has been granted, for example, in the context of a change in the manufacturing process [5].

This article aims to give an overview of current requirements of the biosimilar comparability exercise, particularly the nature of clinical studies required for each medicinal product class. Case studies are presented based on a review of EPARs since 2012 to illustrate the concept of “totality of the evidence” [6], which is an integral part of the biosimilar concept [7–9]. These case studies reflect the complex situations when assessing the totality of the data: How to interpret and weigh conflicting results from different aspects of

the comparability exercise? Which performance characteristic is more important: a successful clinical trial, functional data or quality attributes? Can clinical data overrule a failed PK trial? Can a formally failed clinical trial be acceptable at all? Based on the knowledge gained, the article also tries to provide an outlook on potential changes in future requirements for biosimilar developments.

2 Clinical Data Requirements for Biosimilars

Generally, one confirmatory randomized controlled trial would be conducted to confirm biosimilarity in terms of efficacy, safety and immunogenicity [5, 10, 11]. An exception could be foreseen for biosimilar development in orphan indications, where a randomized clinical trial may not be feasible.

The guidelines state that the aim of clinical data is to address slight differences observed at previous steps and to confirm comparable clinical performance of the biosimilar and the reference product. However, clinical data cannot be used to justify substantial differences in quality attributes [5, 12]. Analytical and functional *in vitro* assays are generally more specific and sensitive than studies in humans when detecting potential differences between the biosimilar and the reference product and are therefore seen as paramount for the overall biosimilar comparability exercise. However, the clinical relevance of differences in quality attributes, if detected, may not be known, requiring further *in vitro* or *in vivo* studies for clarification.

The nature and extent of clinical work necessary for a biosimilar comparability exercise depends on several factors, including the complexity and immunogenic risk of the molecule, whether the drug has several or unknown mechanisms of action and whether an accepted pharmacodynamic (PD) endpoint is available to assess efficacy (Table 1) [12].

Comparative pharmacokinetic (PK) studies are a basic requirement for development of a biosimilar. In the presence of suitable PD endpoints and a clear mechanism of action, a PK/PD study may be sufficient clinical work for marketing approval [13, 14].

In recent years, a dedicated comparative efficacy trial was no longer deemed necessary for several product categories (insulin, low-molecular-mass heparins and (peg)filgrastim), for which pivotal evidence for similarity may be derived from physicochemical, functional, pharmacokinetic and pharmacodynamic comparisons [15–17]. Furthermore, it has been reasoned that, provided the biosimilar and the reference product exhibit comparable quality and pharmacology profiles, adverse reactions that are related to exaggerated pharmacological effects can be expected to occur at similar frequencies and severity. If, in addition, the impurity profile and the nature of excipients of these biosimilars with

an acknowledged low immunogenic risk do not give rise to concerns, a safety and/or immunogenicity study may be waived [15–17].

However, for complex, multifunctional biologicals, comparative efficacy and safety clinical trials in patients are still viewed as a necessary component of the biosimilar development [12]. A number of aspects are still seen as a barrier to easing clinical requirements, such as the size of the molecule, diverse moieties with different functions (e.g. Fab/Fc-parts), multiple mechanisms of action, impact of glycosylation pattern and potential for immunogenicity and potentially life-threatening adverse effects (AEs). There should always be a sound scientific rationale for conducting clinical studies and, where the impact of minor changes at the analytical and functional level is clearly understood and several adequate orthogonal analytical and functional assays exist, clinical efficacy trials may be commensurately reduced in the future.

2.1 Efficacy Endpoints

While clinical trials of originator products aim to demonstrate patient benefit, they are intended to exclude clinically relevant product-specific differences in biosimilar developments. Clinical endpoints used in biosimilar comparability studies should therefore ideally measure unconfounded pharmacological effects and be sensitive in detecting potential clinically relevant differences between the biosimilar candidate and its reference medicinal product. Hard clinical endpoints such as overall survival are rather insensitive in this respect and are often influenced by disease- and patient-related factors. It is thus not surprising that endpoints in biosimilarity studies often differ from primary endpoints used in the pivotal efficacy studies for the reference products [12]. This could be either because of scientific advances, i.e. validation of new endpoints (pathological complete response [pCR] in breast cancer [18–20]), or a shift towards endpoints that are viewed as more sensitive and thus more discriminatory (Disease Activity Score [DAS] 28 vs. American College of Rheumatology [ACR]-20 in rheumatoid arthritis [RA] [21] and objective response rate [ORR] in solid tumours and lymphoma [22, 23]).

PD endpoints that largely explain the clinical effect of the biologic are the preferred endpoints to establish similar efficacy. For smaller proteins with a less complex structure, examples of currently accepted PD parameters include absolute neutrophil count (ANC) for granulocyte-colony stimulating factor (G-CSF), blood glucose concentrations in clamp studies for insulins, magnetic resonance imaging-related endpoints for interferon- β and, more recently, serum calcium levels for teriparatide [14]. For low-molecular-weight heparins that are a mixture of glycan structures and are considered biologicals in the EU, PD endpoints such as anti-Factor X and anti-Factor II activity are acceptable for proof of efficacy.

Table 1 Clinical study recommendations for biosimilars of different product classes. Source: European Public Assessment Reports and guidelines

Substance	Structure of protein	Molecular mass (kDa)	Current clinical study recommendations
Teriparatide	Single-chain, non-glycosylated	3.7	PK/PD study in HV No efficacy/safety trial
Low-molecular-weight heparin	Polysaccharide chain	4.5	PK/PD study in HV No efficacy trial Safety/immunogenicity study may be waived
Insulin	Two-chains, non-glycosylated	6	PK/PD (clamp) study in HV or patients with T1DM No efficacy trial Safety/immunogenicity study may be waived
(Peg)filgrastim	Single-chain, non-glycosylated (pegylated)	20	PK/PD study in HV No efficacy trial Safety/immunogenicity study may be waived
Interferon- β	Single-chain, non-glycosylated	20	PK/PD study in HV Efficacy/safety trial in patients with MS using MRI-related efficacy endpoint
Somatropin	Single-chain, non-glycosylated	22	PK/PD study in HV Efficacy/safety trial in children with growth hormone deficiency using height velocity as efficacy endpoint
Follitropin- α	Two-chains, glycosylated	23	PK/PD study in HV Efficacy/safety trial in patients undergoing superovulation for ART using 'number of oocytes retrieved' as efficacy endpoint
Epoetin	Four helices, extensively glycosylated	34	PK/PD study in HV Efficacy/safety trial in patients with renal anaemia using haemoglobin as efficacy endpoint
Etanercept	VEGF-Fc complex 4-chain fusion protein, glycosylated	130	PK study in HV Efficacy/safety trial in patients with RA using ACR20 or DAS28 as primary efficacy endpoint
Infliximab	Complex 4 chains, glycosylated	165	PK study in HV
Bevacizumab		149	Efficacy/safety trial in patients in a therapeutic indication approved for the reference medicine, which is sensitive to detect potential differences
Adalimumab		150	
Trastuzumab		145	
Rituximab		144	PK study in one therapeutic area plus efficacy/safety trial (plus PK data) in the other therapeutic area

Bold copy indicates instances when trials are not required or may be waived

ACR20 improvement of 20% according to criteria of the American College of Rheumatology, ART assisted reproductive treatment, DAS28 Disease Activity Score based on evaluation of 28 joints, HV healthy volunteers, MRI magnetic resonance imaging, MS multiple sclerosis, PD pharmacodynamic, PK pharmacokinetics, RA rheumatoid arthritis, T1DM type 1 diabetes mellitus, VEGF vascular endothelial growth factor

For larger molecules with a complex mechanism of action such as monoclonal antibodies (mAbs), comparative studies in patients using conventional clinical efficacy endpoints are usually still required, but sensitive PD endpoints are also increasingly being discussed in the scientific community, e.g., bone mineral density (BMD) together with serum C-terminal crosslinks (CTX), a bone resorption marker, as co-primary efficacy endpoints for denosumab, a mAb used to treat and prevent osteoporosis. Viewed alone, BMD has a very low dynamic range and very high inter-subject variability, making it a less sensitive marker, and the PD marker 'baseline-normalised serum CTX' covers treatment effects in patients with osteoporosis [24] but lacks data related

to skeletal effects in patients with cancer [25]. Therefore, studying both PD parameters provides complementary information covering all mechanisms of action and indications.

For natalizumab, a mAb indicated for the treatment of multiple sclerosis [26], α 4-integrin receptor saturation could be an acceptable PD marker since it has been established that α 4-integrin receptor binding by natalizumab is directly linked to clinical outcomes [27].

For eculizumab, used to treat paroxysmal nocturnal haemoglobinuria, atypical haemolytic uraemic syndrome and myasthenia gravis, a potential PD marker to study biosimilarity could be serum lactate dehydrogenase levels because of the sustained reduction observed in

intravascular haemolysis for the treatment period, which results in a reduced need for red blood cell transfusions and less fatigue [28].

2.2 Comparability Margins

According to EU guidance, equivalence trials are the standard requirement to ensure that efficacy of the biosimilar is neither decreased nor increased compared with the reference product. Non-inferiority trials may be acceptable if justified on the basis of a strong scientific rationale and taking into consideration the characteristics of the reference product. For example, it may be accepted where the possibility of significant and clinically relevant increases in efficacy can be excluded on scientific and mechanistic grounds [12]. A theoretical example could be an active substance with a very high response rate where higher efficacy would be hard to achieve. However, in practice, non-inferiority trials as pivotal evidence of comparability have not yet been accepted.

Comparability margins are established based on the knowledge of the effect size of the reference medicine and on clinical judgement. They should represent the largest difference in efficacy that would not matter in clinical practice; treatment differences within this range would then be acceptable because they have no clinical relevance [29].

The acceptable equivalence margins depend on patient population, endpoints, backbone therapy and estimated treatment effect, and slight differences may occur depending on the selection of publicly available reference studies. One example is illustrated by two different trastuzumab biosimilars, Ontruzant [19] and Herzuma [18], both of which were studied in early-stage breast cancer but used slightly different primary endpoints: breast pCR (bpCR) versus total pCR (tpCR). In addition, different reference studies were used to derive the comparability margins [18, 19], leading to slightly different equivalence ranges for planning purposes (± 13 vs. $\pm 15\%$) in otherwise largely similar comparability trials.

Another example is a confirmatory clinical trial with biosimilar bevacizumab in non-small-cell lung cancer (NSCLC), as the anticipated treatment effect of the originator may be calculated using outcomes of different reference studies [22]. Further, the frequency of epidermal growth factor receptor mutations is higher ($\geq 40\%$ versus $< 10\%$), and the response rates greater in Asian than Caucasian populations, as reported in the pivotal study of the originator [30–32].

Such population differences should be accounted for when planning comparative trials and may necessitate revision of margins, stratification and/or subset analyses when considering global development.

3 Totality of the Evidence Approach: Case Studies

In this part, case studies are presented to illustrate how components of a comparability exercise interrelate and whether any uncertainties around biosimilarity can be reduced by other data from the comparability exercise.

The review focuses on biosimilars for which differences were observed in different parts of the comparability exercise and on the justification for why these differences did not preclude regulatory approval. The cases are summarised in Table 2 and further discussed in the following sections.

3.1 Analytical Comparability

3.1.1 Infliximab

For one biosimilar, infliximab [33], extensive analytical tests showed physicochemical and structural comparability except for a small difference in the proportion of afucosylated forms. This glycoform is associated with the binding affinity of the molecule to the Fc γ RIIIa receptor expressed on various immune cells. The biosimilar and the reference infliximab demonstrated comparable binding to complement receptor and all types of Fc-receptors except for Fc γ RIIIa/b, where an approximately 20% reduction of binding was seen. This was accompanied by a 20% lower antibody-dependent cell-mediated cytotoxicity (ADCC) activity measured for the biosimilar in a highly sensitive assay using Jurkat cells overexpressing tumour necrosis factor receptors and natural killer cells. However, this difference disappeared under more physiological conditions, e.g. when serum was added or peripheral blood mononuclear cells were used as effector cells, suggesting that the observed difference in Fc γ RIIIa binding was not relevant in vivo [34]. These differences were confirmed not to be clinically relevant in a large PK study in patients with ankylosing spondylitis and a large phase III equivalence trial in patients with RA.

All indications of the reference product were approved in the EU and in the USA but initially not in Canada, where extrapolation from autoimmune arthritis to inflammatory bowel disease was not accepted, as the ADCC mechanism of action was viewed as more important in the latter. Health Canada's decision was reversed in 2015 based on additional functional data and the totality of evidence [35].

3.2 Comparability Regarding Pharmacokinetics and/or Pharmacodynamics

3.2.1 Adalimumab

Two examples of an initially failed and subsequently successful PK study conducted with adalimumab biosimilars

Table 2 Summary of differences or omissions in comparability exercise not precluding conclusion on biosimilarity

Reference	Quality	Functional	PK	PD	Clinical (E/S)	Immunogenicity	Regulatory outcome in EU
Remsima (infliximab)	Decreased % of afucosylated species	Decreased ADCC activity	✓	Supportive	✓	✓	Approved
Cyltezo (adalimumab)	✓	✓	Initial PK study failed	Supportive	✓	✓	Approved
Hyrimoz (adalimumab)	✓	✓	Initial PK study failed	Supportive	✓	✓	Approved
Terrosa (teriparatide)	✓	✓	90% CI of primary endpoints not covering unity	✓ (PD analysis included in PK study)	NA	NA	Approved
Efgratin (pegfilgrastim)	✓	✓	PK study failed	✓	✓	✓	Withdrawn
Ziextenzo (pegfilgrastim)	✓	✓	Initial PK study failed	✓ (PD analysis included in PK study)	✓	✓	Approved
Grastofil (filgrastim)	✓	✓	90% CI of some endpoints not covering unity	✓	NA	NA	Approved
Semglee (insulin glargine)	✓	✓	✓	Formally failed PD (AUCGIR 0–30 h and GIR max) due to several patients with very low PD response	✓	✓	Approved
Ontruzant (trastuzumab)	✓	✓	✓	NA	Failure to meet upper equivalence limit for 1st EP of clinical S/E trial	✓	Approved
Kanjinti (trastuzumab)	✓	✓	✓	NA	Failure to meet upper equivalence limit for 1st EP of clinical S/E trial	✓	Approved
Rixathon (rituximab)	✓	✓	✓	Supportive	Pivotal trial: 1st EP: ORR was met, uncertainty regarding 2nd EPs PFS, OS	✓	Approved
Truxima (rituximab)	✓	✓	✓	Supportive	Pivotal trial: 1st EP: ACR20 was met, uncertainty regarding 2nd EPs PFS, OS	✓	Approved
Flixabi (infliximab)	✓	✓	✓	Supportive	✓	Slightly increased ADA	Approved
Benepali (etanercept)	✓	✓	✓	✓	✓	Reduced immunogenicity	Approved

The table lists products for which a particular component of the comparability exercise raised a red flag during the assessment. Review of products for which an opinion was given between July 2012 and July 2019. References in all instances are the EPAR of the products

ACR20 improvement of 20% according to criteria of the American College of Rheumatology, ADA antidrug antibodies, ADCC antibody-dependent cell-mediated cytotoxicity, CI confidence interval, E efficacy, EP efficacy endpoint, GIR glucose infusion rate, NA not applicable, ORR objective response rate, PD pharmacodynamics, OS overall survival, PFS progression-free survival, PK pharmacokinetics, S safety, ✓ indicates results supporting biosimilarity

in healthy volunteers have been published [36, 37]. In both instances, it was argued that the differences in glycan structures known to affect PK (high mannose content) were too small to explain the initially observed PK differences, as

only high mannose content of at least 20% would have the potential to alter the systemic exposure because of increased receptor-mediated elimination [38].

In one of these cases, the initial failed PK study [37] was performed with a clinical trial formulation exhibiting differences in the buffering system compared with the commercial formulation. Post-hoc analyses correcting, for example, for body weight and protein content as covariates were performed but were unable to provide a satisfactory rationale for the observed differences. PK similarity was demonstrated in a second, improved PK study using the formulation intended for commercialization, a larger subject sample size considering the high PK variability, a standardized injection site and the predefined covariates body weight and age.

In the other case, an extensive root cause investigation was performed, including batch selection, investigational medicinal product (IMP) storage and transport, IMP preparation, IMP administration, PK sampling, PK sample shipping and testing, impact of body weight and antidrug antibody (ADA) development and other subject characteristics [36]. However, no root cause driving the negative outcome of the PK study could be identified by the applicant.

Therefore, a second PK study was conducted with an adapted study design aiming at reducing inter-subject variability (body mass index [BMI] restriction, inclusion of only male subjects, increased sample size). IMP handling and dosing was also simplified by using prefilled syringes so they did not require IMP compounding as in the previous study. With this improved design, PK similarity could be shown.

Generally life-cycle management on the part of the originator, such as with the recent formulation changes to originator adalimumab may pose a further challenge to biosimilar developers. Differences in formulation strengths will require analytical comparability testing in line with ICH Q5E. Depending on the results, the need for (non-)clinical comparability studies will be identified but at any rate a clinical PK study is highly recommended.

In conclusion, a failed comparative PK study requires a thorough root cause analysis. The insights, if any, should lead to improved design of a subsequent study. To decrease the notably high PK variability of many biologicals, test conditions should be standardized as much as possible.

3.2.2 Teriparatide

Teriparatide is a relatively simple molecule, as it is a 34 amino acid monomer and contains no glycosylation or other post-translational modifications.

To claim similarity of a biosimilar teriparatide to its reference product, a thorough physicochemical, structural and biological characterization as well as impurity profiling was performed [14]. The clinical development programme consisted of one single-dose comparative PK study in 54 healthy women.

The predefined equivalence range of 80–125% for the relative bioavailabilities were met, but 100%, i.e. unity, was not included in the 90% confidence interval (CI), indicating statistically significant differences, which is acceptable with appropriate justification [39]. Mean exposure and peak serum concentrations were around 8% lower with the biosimilar than with the reference product, but the clinical impact of the observed difference was considered to be negligible given the available literature on the reference product regarding the impact of body weight and administration site on exposure. No relevant differences were identified by further analysing delivered volumes, active substance content, structure of the active substance and the PK assay.

In terms of the PD properties of teriparatide, the applicant measured serum calcium at several timepoints during the comparative PK study, as teriparatide is known to cause transient increases in calcium after each dose. An equivalence margin was not prespecified, but statistical analyses of all serum calcium concentrations showed close similarity between test and reference.

The close analytical and functional similarity, together with the similar PD profiles and the absence of a relevant difference in PK supported CHMP's conclusion of biosimilarity.

3.2.3 Pegfilgrastim

Pegfilgrastim is associated with notably high PK variability [17]. Therefore, every effort should be undertaken to minimize this variability in comparative PK studies. Since the dose-exposure relationship of pegfilgrastim is greatly disproportional, i.e. in healthy subjects, a tenfold increase in dose was shown to lead to an approximately 75-fold increase in exposure [40], and correction for protein content using linear models is inappropriate, attention should also be paid to administering exactly the same dose of test and reference product.

Interestingly, the high PK variability is not paralleled by high PD variability. On the contrary, the exposure-response relationship appears rather flat, having led to highly similar PD responses (i.e. ANCs) even in cases of high PK variability or failed PK similarity, thus rendering PD endpoints less sensitive than PK endpoints to detect potential differences between two pegfilgrastim-containing products.

Two of six marketing authorization applications for biosimilar pegfilgrastims originally had failed PK trials using traditional comparability margins of 80–125% [41, 42].

Despite highly similar PD responses and data from phase III trials showing similar clinical performance, CHMP did not accept that predefined PK similarity margins were not met, since PK studies are generally considered to be more sensitive to detect product-related differences than PD endpoints or clinical trials in patients. Both products were

withdrawn during the initial review process but were later resubmitted with results from new PK studies to support biosimilarity. However, justification is always needed as to why evidence of similar PK profiles from new studies could outweigh results of failed PK comparability in previous studies. In one of the cases just described, the following reasons were accepted by CHMP. The initial study was a parallel-group study that was underpowered based on a grossly underestimated inter-subject variability. The second PK study employed a crossover design which per se reduces variability and was sufficiently powered to show PK similarity. In addition, sample heterogeneity was further reduced in the second study by formulating stricter inclusion requirements for ANC and BMI, both of which are known to be key drivers of pegfilgrastim PKs. Thus, insights gained from the failed first study were used to design a more appropriate and successful second study [43]. In the other case, the CHMP was concerned about the validity of results from the PK studies, and the applicant chose to withdraw the second application [44].

For the other four approved pegfilgrastim biosimilars, similar PK profiles were shown in crossover studies. The new draft revision of the respective guideline [17] suggests widening of the acceptance range may be possible for the main PK parameters of pegfilgrastim since the PK variability is high and the PK–PD relationship is rather flat. In contrast, narrow comparability margins of 90–111% are suggested for the main PD parameter, ANC. Of note, based on the consideration that the endpoint ‘duration of neutropenia’ in clinical trials in patients receiving chemotherapy is rather insensitive with regard to detecting product-related differences of two (peg)filgrastim-containing medicinal products, and a suitable PD endpoint is available and measurable in healthy subjects, the requirement for a clinical trial has been lifted in the new draft revision of the guideline [17].

3.2.4 Insulin Glargine

For one biosimilar insulin glargine [45], the analysis of the main PD endpoint ‘glucose infusion rate’ (GIR) in the insulin clamp study presented by the applicant yielded results within the predefined comparability range. However, study subjects with very low glucose requirements (approximately <5% of mean values) were excluded, which was not pre-specified in the statistical analysis plan. When including glucose profiles from all subjects in the analysis, the predefined acceptance range was not met. Therefore, PD similarity was formally not shown.

Nevertheless, the following findings and arguments led to the conclusion of no relevant product-related PD differences:

Firstly, variable response to insulin is known from the literature, and cases of low glucose requirements were seen equally with test and reference product.

Secondly, sensitivity analyses using different cut-offs for exclusion of ‘low glucose requirements’ yielded point estimates consistently around one, suggesting intra-individual variability rather than product-related differences being the likely cause for not meeting the primary PD objective of the clamp study.

Thirdly, the primary analysis was based on log-transformed values, which is problematic for values close to zero and increases variance. An analysis of non-transformed data including all study subjects resulted in 95% CIs meeting predefined equivalence margins.

Finally, the guideline on biosimilar insulin and insulin analogues [15] provides the possibility of considering PD endpoints as secondary (thus allowing descriptive analyses) if analytical and functional similarity can be concluded from comprehensive characterisation and comparison of test and reference product and if the PD data reasonably support PK results.

All these requirements were considered fulfilled. Therefore, biosimilarity was concluded based on the totality of data and scientific considerations, and the biosimilar was finally approved.

3.3 Comparability Regarding Clinical Efficacy Outcomes

3.3.1 Trastuzumab

To date, five trastuzumab products have been approved in Europe, three from the pivotal comparability study in early breast cancer (EBC) and two in metastatic breast cancer with different endpoints and equivalence margins depending on the indication, backbone therapy and reference studies. Both clinical development paths are viewed as feasible and both are sensitive models and allow for meaningful results and extrapolation to other lines of therapy, a view shared by regulators in other jurisdictions.

For two trastuzumab biosimilars, the phase III study in patients with human epidermal receptor-2 (HER2)-positive EBC/locally advanced breast cancer did not formally meet the upper bound of the predefined equivalence margins for the primary endpoint (pCR), confirming non-inferiority but not formally excluding the possibility of superior efficacy [19, 46]. Overall, structural and functional similarity was shown in a comprehensive head-to-head comparison. However, the applicants noted a slightly reduced ADCC activity of some more recent batches of the reference product Herceptin used in the clinical trial, an observation also described in the literature [47]. The overall contribution of ADCC activity versus antiproliferative effects through inhibition of ligand-independent HER2 signalling to the therapeutic benefit of trastuzumab is unknown. The observed difference in pCR was considered at least partly confounded by the

small shift in ADCC activity in a number of the Herceptin batches used in the pivotal trial. Overall, it was considered doubtful that a shift as small as the one observed would have any significant impact in terms of clinical outcomes, although numerically it is thought to have contributed to a more extreme location of the point estimate and upper bound of the CI, shifting the latter beyond the prespecified equivalence margin.

In both instances, no clinically meaningful differences in the safety profile were found, notably no differences in cardiac toxicity as measured by left ventricular ejection fraction and incidence of symptomatic heart failure, which could be of concern with truly increased efficacy. Furthermore, the small difference was viewed as not being clinically significant given the small magnitude and the unlikely impact of small changes in pCR on clinically relevant time-dependent clinical endpoints.

From a PK perspective, comparability between the biosimilars and the reference product was demonstrated in healthy volunteers since the ratios (90% CI) of geometric means for both primary PK endpoints were within the acceptability range of 80–125%. Similarly, minimum concentrations measured at steady state (C trough) during the phase III trial supported similarity between treatments.

Given the totality of data submitted, similarity between the two trastuzumab biosimilars and the reference product was considered sufficiently established.

3.3.2 Rituximab

During the development of a rituximab biosimilar [23], shifts in some quality attributes (e.g. charge variants, glycan structures, ADCC) were noted for the EU reference product and US Rituxan [48]. As both versions of the reference products were on the market simultaneously, both were considered safe, effective and appropriate for use in comparability studies.

Similar efficacy and safety of the biosimilar and the reference product was concluded based on the comparative clinical trial in patients with previously untreated, advanced-stage follicular lymphoma who received rituximab-cyclophosphamide, vincristine, prednisone (CVP) combination treatment. The difference in ORR was only -0.40% (95% CI -5.94 to 5.14) and the entire 95% CI for the difference in ORR between the two treatments was within the prespecified equivalence margin of $\pm 12\%$, allowing conclusion of equivalence.

However, at the time of data cut-off, more patients in the biosimilar group than in the reference product group had progressed or died, with an adjusted hazard ratio estimate (biosimilar/reference product) for progression-free survival (PFS) of 1.33 (90% CI 0.98–1.80). For CHMP, these figures suggested immature PFS data as 30.3% and 25.4% of

patients had < 6 months and 6–12 months of PFS follow-up, respectively, and considering that the majority of PFS events occurred during the maintenance phase of the trial. In conclusion, the difference in PFS was considered to be due to patient heterogeneity or random data variation rather than a real treatment effect.

In comparison, for another biosimilar rituximab [21] the pivotal efficacy trial was an equivalence trial conducted in patients with RA, and a second, supportive study was performed in patients with advanced follicular lymphoma covering eight treatment cycles to demonstrate similar PK and non-inferior efficacy of the biosimilar to the reference product. The primary endpoint was met in both trials; however, in the supportive trial, again, the secondary endpoints PFS and overall survival were inconclusive at the time of analysis which, based on the totality of the evidence, did not preclude approval in the EU but led to additional requests by the US FDA. Therefore, another prospectively designed comparative clinical trial was conducted with rituximab monotherapy in patients with low tumour burden follicular lymphoma. The primary endpoint was ORR at 7 months, and the FDA-requested 90% CI for the difference in ORR between the two treatments fell within the prespecified margins, thus leading to FDA approval and confirming the positive scientific opinion reached by CHMP before [49].

Taken together, clinical comparability between a biosimilar candidate and rituximab originator can be demonstrated based on a PK trial in healthy subjects together with an efficacy trial performed in an autoimmune or oncological indication and a PK bridging study in the other indication. In the future, CHMP may revisit the value of the currently requested additional PK bridging data.

3.4 Comparability Regarding Immunogenicity

3.4.1 Infliximab

For one biosimilar infliximab [50], the primary efficacy analysis of the comparative phase III study in patients with RA demonstrated equivalent ACR20 response rates at week 30 with biosimilar and reference product. However, ADA rates measured with a highly sensitive assay were about 5–12% higher in the biosimilar cohort at the individual time points of determination (with about 50% of patients in the biosimilar cohort determined ADA positive at any time in the trial).

Some CHMP members argued that, although the predefined equivalence margins for the efficacy endpoint were met, the point estimate was consistently below one, indicating that efficacy of the biosimilar may be somewhat lower than that of the reference product. They further argued that it was not possible to exclude with reasonable certainty that this was the result of the higher incidence of ADAs,

considering that ADAs were shown to reduce efficacy. There was also concern about extrapolation to other indications as patients with RA are treated concomitantly with immunosuppressive therapy, which is not the case for other infliximab-licensed indications. Therefore, the difference in ADA development could be even greater in the other indications.

However, the majority view was that the numerical differences in ADA rates had no meaningful impact on any of the efficacy parameters analysed, as the primary endpoints fell within the predefined comparability margins. Additional data showed that a similar percentage of patients in both treatment arms required increased doses of study drug irrespective of ADA status, which provided further evidence that ADAs did not have a relevant impact on efficacy. Further, the numerically higher incidence of ADAs in patients treated with the biosimilar was not associated with an unfavourable safety profile compared with the reference product. Moreover, despite concomitant treatment with methotrexate, ADA development is reportedly highest in patients with RA, indicating that immunogenicity would not be of greater concern if the biosimilar were to be used in the other licensed indications of the reference product

From an analytical, functional and PK perspective, convincing similarity of the biosimilar and the reference product was shown, further supporting the conclusion of biosimilarity and leading to approval.

The company submitted 78-week follow-up data post-authorisation which showed slightly higher ADAs for patients treated with Flixabi, but without any clinical impact, i.e. no statistically significant or clinically meaningful difference between treatment groups [51].

3.4.2 Etanercept

For a biosimilar etanercept [52], analytical, functional and PK data supported biosimilarity. The pivotal efficacy study was conducted in patients with RA and provided robust evidence of equivalent efficacy between the biosimilar and the reference product based on ACR20 response at week 24, supported by most secondary efficacy parameters and sensitivity analyses.

However, in terms of immunogenicity, there was a significant difference in overall ADA formation at week 24. While only three biosimilar-treated patients tested positive for ADAs at some point of the study, 39 patients tested positive in the reference product group, one of which also tested positive for neutralizing antibodies. The clinical impact of the difference in ADAs seemed negligible and the difference largely vanished after 8 weeks of treatment. In addition, the applied electrochemiluminescence assay suffered from a low drug tolerance, rendering the finding of reduced ADA incidence with the biosimilar uncertain.

In summary, the ADA data suggested reduced immunogenicity of the biosimilar, but this would not preclude a conclusion of biosimilarity [12].

4 Discussion

Since the implementation of the legal framework for biosimilars in the EU in 2004, extensive experience has been gained with the development, approval and use of increasingly complex biosimilars. Guidance documents laying down the requirements for the development and licensing of biosimilars are 'living documents', meaning they are revisited on a regular basis and updated as needed. While the initial approach to the then novel field of biosimilars was understandably cautious and conservative to protect patients' safety, the analytical and scientific progress and the generally improved understanding of the biosimilar concept continue to reshape the regulatory requirements for developing and licensing of biosimilars.

We were particularly interested in evaluating the role and nature of clinical confirmation required in addition to analytical and functional *in vitro* data and the totality of the evidence approach, especially in cases with seemingly contradictory results from different levels of the comparability exercise.

It is undisputed that a comprehensive analytical and functional comparison of the biosimilar candidate and the reference product are the mainstay for establishing biosimilarity. Major advances in the analytical sciences allow in-depth characterization of increasingly complex biologicals. If differences in quality attributes are detected, their impact on functional aspects of the molecule must be examined, and it must be judged whether they are clinically meaningful. There is an abundance of quality attributes, and it is often not entirely clear which features contribute to certain clinical effects and to what extent. Therefore, it is important to improve the understanding of the impact of a given attribute on certain performance categories (i.e. PK, PD, efficacy, safety or immunogenicity profiles), which could further reduce the extent or nature of requirements for clinical data.

It has been suggested that regulators define which structural/physicochemical and functional attributes of the reference product are considered critical, how that information should be used in the development of an analytical similarity assessment plan and whether or which statistical approaches could be recommended. To this extent, the EMA published a reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development [53], and a stakeholder meeting was held after the end of the consultation period in May 2018. Many questions are still open, for example, the number of reference product batches required for establishing the quality target range for

the biosimilar and for the subsequent comparative analyses, the types and variability of the analytical methods, the level and variability of the quality attributes and the relevance of the quality attributes for the performance category. However, these factors are highly dependent on the specific biosimilar development. Therefore, European regulators maintain a science-based case-by-case approach by looking at the totality of evidence rather than defining rigid cut points. However, statistical approaches may be considered supportive in establishing analytical similarity.

Our review of the role of PK studies reveals that they can be considered a sine qua non condition for any biosimilar approval. It is conceivable that demonstration of similar exposure to the respective biological substance from the biosimilar and the reference product is seen as a basic and critical requirement. The notion that PK studies are generally more sensitive in detecting potential product-related differences than clinical trials may explain why a finding of similar efficacy was not able to overrule the differences in PK. This may be perceived as overly strict, but the outcome of efficacy trials depends not only on drug exposure but also on proper pharmacological action of the biological substance *in vivo*. Therefore, the objectives of both types of studies differ.

It is recognized that specific acceptance ranges for PK parameters have not been established for biosimilars, and the appropriateness of using the traditional bioequivalence margins for small-molecular generics is unclear. However, it is up to the applicant to justify the comparability margins employed, and a post-hoc widening of the margins when the predefined margins have not been met is generally not acceptable from a regulatory viewpoint. In addition, CHMP acknowledges that PK acceptance ranges may not be uniform for biosimilars and may depend, among other things, on the dose-exposure-response relationship as suggested by the draft revision of the guideline for biosimilar recombinant G-CSF products [17].

As detailed in a recently published question and answer document [54], a failed PK study is acceptable as part of a marketing authorization application if the results from the root cause analysis are adequately reflected in the planning and conduct of a subsequent sensitive PK study which then shows similar PK profiles. It is crucial that developers of biosimilar medicines provide scientific justification as to why the failed PK comparability in the previous study can be disregarded. In all instances, careful analyses of quality components known to affect PK must lead to reassurance that no significant differences in quality attributes are observed that could lead to true differences in PK.

So far, for all approved biosimilar products, the pivotal PK study had to meet the prespecified equivalence margins, which were typically the traditional margins for generics. However, in some instances, unity was not included in the 90% CI, indicating significant differences in bioavailability.

Since it is unclear whether the acceptance range for generics can also be applied to biosimilars, such results had to be further explained and justified by the applicant in the context of evidence for similarity coming from other comparative studies/assays within the development programme.

Comparative clinical efficacy trials are still needed for most biosimilar developments, especially for biologicals with a complex mechanism of action such as mAbs, but sensitive endpoints should be employed focussing on pharmacological action of the biological rather than patient benefit. Therefore, PD endpoints, if relevant for efficacy, are generally preferred over hard outcome endpoints. However, such PD endpoints are not always available or may not be readily or sensitively measurable. Considering the burden of performing large, costly and time-consuming clinical trials that are also often insensitive because of the confounding effects of disease- or patient-related factors or concomitant therapies, alternative approaches should be explored and may be suggested to the regulators during scientific advice procedures.

The conclusion on whether biosimilarity has been established is generally based on the totality of the evidence. As shown for two trastuzumab biosimilars [19, 46], it may—in exceptional cases—be acceptable if the 95% CI of the primary efficacy endpoint does not formally meet the upper equivalence margin. First, it should be acknowledged that the potential for greater efficacy is less of a concern for some products than for others. Second, this does not necessarily mean that the biosimilar is indeed more effective than the reference product, just the uncertainty is increased above the usually accepted statistical level. In the cases referred to, the differences observed could at least partly be attributed to factors not related to the biosimilar, and the true difference was considered likely to fall within the equivalence range and to be of no clinical relevance. In addition, analytical and functional similarity was demonstrated and reassurance provided that the potentially increased efficacy of the biosimilars was not associated with any increased risks for patients.

In contrast, increased uncertainty regarding inferior efficacy of the biosimilar candidate (i.e. when the 95% CI of the primary efficacy endpoint extends below the lower margin of the equivalence range) may not be acceptable, even if other parts of the comparability exercise would be consistent with biosimilarity, because treating a patient with a potentially inferior product would be unacceptable. In this context, proposals to accept non-inferiority trials or trials with asymmetric equivalence margins, i.e. a lower margin for efficacy and a higher margin for safety, have been entertained but so far not accepted.

For immunogenicity, it is stated [12] that increased immunogenicity compared with the reference product may not be acceptable, whereas decreased immunogenicity (e.g. because of increased purity) would not preclude approval as

a biosimilar medicine. Whether differences in ADA development are product or patient related or just a play of chance can often not be decided, but a thorough root cause analysis and investigation of potential clinical relevance is necessary in any case.

This review provides evidence that regulatory assessment is based on a scientific evaluation of the totality of evidence obtained in the comparability exercise and will take into consideration that any remaining uncertainty can only be towards a potential benefit for the patient. This approach has allowed safe and effective biosimilars to reach the EU market as illustrated by pre- and post-marketing data accumulated since the introduction of the first biosimilar medicine in the EU [55–58]. Of note, while there is no formal regulatory designation in Europe, biosimilars are also being switched for their reference products in Europe, and no interruption in therapeutic outcomes have been seen to date [55, 56].

Although the burden of proof for establishing biosimilarity could be seen as high compared with the requirements for demonstration of comparability after a manufacturing change, it should be acknowledged that the starting point differs (i.e. differences throughout the manufacturing process). The need for (non)clinical studies in a biosimilar development is foreseen in Directive 2001/83/EC because differences in raw materials and manufacturing processes may potentially affect efficacy or safety of a biological substance. Considering the complexity of biological medicines and the limited clinical and regulatory experience with biosimilars at the time, it is not unexpected that a more cautious and conservative approach was initially followed for all product classes. However, applicants can present sound proposals that would allow demonstration of biosimilarity despite deviating from the guidelines.

Non-clinical and clinical requirements have undoubtedly evolved since the initial implementation of the biosimilar framework in the EU with the introduction of a risk-based approach for non-clinical studies [59] and with the gradual waiving of clinical studies as already reflected in some revised guidelines [15–17]. The experience acquired with each biological product class, together with advances in terms of characterization techniques and progress in understanding pharmacology and disease pathogenesis, have paved the way towards reduction of clinical data requirements and are expected to continue doing so.

When discussing the need to conduct clinical efficacy and safety trial(s) for comparability purposes, the sensitivity of the studies to detect product-related differences and the availability of alternative tools such as PD endpoints and biomarkers to assess them is an important consideration as discussed. The need to generate immunogenicity data is another aspect that deserves discussion and should take the

immunogenic risk, including clinical consequences that are known for the reference product, into account.

Generally, clinical trials for biosimilars, like for any medicinal products, are subject to scientific and ethical review, and unnecessary duplication of clinical trials should be avoided. Towards this end, the overarching guideline has also been revised to allow the use of a version of the reference product that is not approved in the European Economic Area but is shown to be a relevant comparator for the clinical comparability studies. The justification may rely on analytical bridging data only [5, 9]. This is in line with the objectives of the clinical trials regulation EU no. 536/2014, which is expected to come into force in 2019 [60].

Overall, the revision of several biosimilar guidelines shows that the EU regulatory framework is moving stepwise in the direction of reducing the clinical requirements for biosimilars, ultimately aiming at timely access to safe and efficacious medicine in Europe while helping avoid unnecessary repetition of trials.

5 Conclusions

With increasing experience and analytical capabilities, and with progressive knowledge of structure–function relationships and disease-specific mechanisms of actions of therapeutic proteins, a continuing reduction in clinical data requirements for biosimilar developments can be foreseen.

The PK study is emerging as a major gatekeeper in the clinical biosimilarity exercise. Demonstration of similar rate and extent of exposure to the respective biological substance from biosimilar and reference product turns out to be a *sine qua non* condition, a hurdle that needs to be overcome and all results fully justified before further clinical data can be deemed acceptable.

The EU biosimilar framework is robust and able to strike a balance between scientific progress, regulatory standards, patient safety and feasibility of biosimilar developments and taking the totality of the evidence into account.

Acknowledgements The authors thank the following colleagues for their valuable comments: Enrica Alteri, Liese Barbier, Antonella Baron, Paul Blake, Laurent Brassart, Rosa Gonzalez-Quevedo, Juan Garcia, Ana Hidalgo-Simon, Sara-Mihoko Kamo, Peter Richardson and Armin Ritzhaupt.

Compliance with Ethical Standards

Funding No sources of funding were used to conduct this study or prepare this manuscript.

Conflict of interest Elena Wolff-Holz, Klara Tiitso, Camille Vleminkx and Martina Weise have no conflicts of interest that are directly relevant to the content of this article.

Disclaimer The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the regulatory agencies with which the authors are affiliated.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial 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

- European Medicines Agency—website. (n.d.). <http://www.ema.europa.eu>. Accessed May 2019.
- ICH Q5E Biotechnological/biological products subject to changes in their manufacturing process: comparability of biotechnological/biological products: CPMP/ICH/5721/03. (n.d.).
- Boero S, Morabito A, Banelli B, Cardinali B, Dozin B, Lunardi G, et al. Analysis of in vitro ADCC and clinical response to trastuzumab: possible relevance of FcγRIIIA/FcγRIIA gene polymorphisms and HER-2 expression levels on breast cancer cell lines. *J Transl Med.* 2015;13:324.
- Vezér B, Buzas 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.
- Guideline on similar biological medicinal products CHMP/437/04 Rev. 1. (2014). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Demonstrating Biosimilarity to a Reference Product, Guidance for Industry. (n.d.). <http://www.fda.gov>. Accessed May 2019.
- McCarmish, Woolett. The state of the art in the development of biosimilars. *Clin Pharmacol Ther.* 2012;91(3):405–7.
- EMA: Tailored scientific advice pilot. (n.d.). http://www.ema.europa.eu/ema/index.jsp?url=pages/regulation/general/genera_l_content_000049.jsp&mid=WC0b01ac05800229b9.
- Biosimilars in the EU, Information guide for healthcare professionals, prepared jointly by the European Medicines Agency and the European Commission, European Medicines Agency. (2017). https://ec.europa.eu: https://ec.europa.eu/growth/content/information-patients-what-i-need-know-about-biosimilar-medicines-0_en. Accessed May 2019.
- DIRECTIVE 2001/83/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. (2001). Official Journal of the European Communities.
- EUDRALEX. (n.d.). Eudralex: European Commission Directive 2001/83/EC Art. 10(4) and Part II of the Annex I of Directive 2001/83/EC, as amended. EudraLex, Volume 1, Pharmaceutical legislation medicinal products for human use. http://ec.europa.eu: http://ec.europa.eu/health/files/eudralex/vol-1/dir_2001_83_cons2009/2001_83_cons2009_en.pdf. Accessed 20 Jul 2011.
- 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). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Filgrastim Hexal EMEA/CHMP/651324/2008. (2009). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Terrosa EMA/84371/2017. (2016). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Guideline on non-clinical and clinical development of similar biological medicinal products containing recombinant human insulin and insulin analogues EMEA/CHMP/BMWP/32775/2005_Rev. 1. (2015). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Guideline on non-clinical and clinical development of similar biological medicinal products containing low molecular-weight-heparins; EMEA/CHMP/BMWP/118264/2007 Rev. 1. (2016). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Guideline on similar biological medicinal products containing recombinant granulocyte-colony stimulating factor (rG-CSF), EMEA/CHMP/BMWP/31329/2005 Rev 1. (2018). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Herzuma EMA/44005/2018. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Ontruzant EMA/CHMP/9855/2018. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Guideline on the evaluation of anticancer medicinal products in man; EMA/CHMP/205/95 Rev.5. (2017). <http://www.ema.europa.eu>. Accessed May 2019.
- EPAR Truxima EMA/CHMP/75695/2017. (2016). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Mvasi EMA/798844/2017. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Rixathon EMA/303207/2017. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Prolia EMA/21672/2010. (2010). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Xgeva EMEA/H/C/002173. (2011). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Tysabri. (2007). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Derfuss T, Kovarik J, Kappos L, Savelieva M, Chhabra R, Thakur A, et al. α4-integrin receptor desaturation and disease activity return after natalizumab cessation. *Neurol Neuroimmunol Neuroinflamm.* 2017. <https://doi.org/10.1212/NXI.0000000000388>.
- EPAR Soliris. (2007). <http://www.ema.europa.eu>. Accessed May 2019.
- Biosimilars in the EU, Information for healthcare professionals, 2017, p. 19. (2017). <https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/biosimilar-medicines-marketing-authorisation>. Accessed May 2019.
- Sandler A, Gray R, Perry M, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med.* 2006;355(24):2542–50.
- Niho S, Kunitoh H, Nokihara H, Horai T, Ichinose Y, Hida T, et al. Randomized phase II study of first-line carboplatin-paclitaxel with or without bevacizumab in Japanese patients with advanced non-squamous non-small-cell lung cancer. *Lung Cancer.* 2012;76(3):362–7. <https://doi.org/10.1016/j.lungcan.2011.12.005> (Epub 2012 Jan 14).
- Zhou C, Wu Y-L, Chen G, Liu X, Zhu Y, Lu S, et al. Beyond: a randomized, double-blind, placebo-controlled, multicenter, phase III study of first-line carboplatin/paclitaxel plus bevacizumab or placebo in Chinese patients with advanced or recurrent nonsquamous non-small-cell lung cancer. *J Clin Oncol.* 2015;33(19):2197–204.
- EPAR Remsima EMA/CHMP/589317/2013. (2013). <http://www.ema.europa.eu>. Accessed Dec 2018.
- Weise M, Kurki P, Wolff-Holz E, Bielsky M, Schneider C. Biosimilars: the science of extrapolation. *Blood.* 2014;124:3191–6.
- Weise M, Wolff-Holz E. Opportunities and challenges of extrapolation for biosimilars. *Z Gastroenterol.* 2016;54(11):1211–6.
- EPAR Hyrimoz EMA/CHMP/404076/2018. (2018). <http://www.ema.europa.eu>. Accessed Dec 2018.
- EPAR Cyltezo EMA/CHMP/750187/2017. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.

38. Goetze AM, Diana Y, Zhang Z, Shah B, Lee E, Bondarenko PV, et al. High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans. *Glycobiology*. 2011;21(7):949–59.
39. Clinical pharmacology and pharmacokinetics: questions and answers. (n.d.). <http://www.ema.europa.eu>. Accessed Dec 2018.
40. Roskos L, Lum F, Lockbaum F, Schwab G, Yang B. Pharmacokinetic/pharmacodynamic modeling of pegfilgrastim in healthy subjects. *Clin Pharmacol*. 2006;46:747–57.
41. EPAR Ziextenzo (pegfilgrastim), EMA/47326/2017; Withdrawal of the marketing authorisation application. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.
42. EPAR Efgratin. EMA/18691/2017; Withdrawal assessment report. (2016). <http://www.ema.europa.eu>. Accessed Dec 2018.
43. EPAR Ziextenzo EMA/CHMP/706001/2018. (2018). <http://www.ema.europa.eu>. Accessed Dec 2018.
44. Efgratin: Withdrawal of the marketing authorisation application EMA/912780/2019. (2019). <http://www.ema.europa.eu>. Accessed May 2019.
45. EPAR Semglee EMA/119474/2018. (2018). <http://www.ema.europa.eu>. Accessed Aug 2019.
46. EPAR Kanjinti EMA/CHMP/261937/2018. (2018). <http://www.ema.europa.eu>. Accessed Dec 2018.
47. Kim S, Song J, Park S, Ham S, Paek K, Kang M, et al. Drifts in ADCC-related quality attributes of Herceptin[®]: impact on development of a trastuzumab biosimilar. *MAbs*. 2017;9(4):704–14.
48. Stangler T, Torella C, Cepeljnik T, Toll H, Grau R. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. *Nat Biotechnol*. 2011;29(4):310–2.
49. FDA Briefing document, Oncologic Drug Advisory Committee Meeting, BLA 761088, CT-P10, a proposed Biosimilar. (2018). <http://www.fda.gov>: <https://www.fda.gov/AdvisoryCommittees/Calendar/ucm620416.htm>. Accessed Dec 2018.
50. EPAR Flixabi EMA/CHMP/272283/2016. (2016). <http://www.ema.europa.eu>. Accessed Dec 2018.
51. EPAR Flixabi: procedural steps taken and scientific information after the authorization. <http://www.ema.europa.eu>. Accessed Aug 2019.
52. EPAR Benepali EMA/CHMP/819219/2015. (2015). <http://www.ema.europa.eu>. Accessed Dec 2018.
53. Statistical methodology for the comparative assessment of quality attributes in drug development EMA/CHMP/138502/2017 Draft reflection paper. (2017). <http://www.ema.europa.eu>. Accessed Dec 2018.
54. CHMP request on biosimilarity to PK Working Party, EMA/CHMP/230419/2017 (2018).
55. 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.
56. Kurki P, van Aerts L, Wolff-Holz E, et al. Interchangeability of biosimilars: a European perspective. *BioDrugs*. 2017;31:83–91.
57. Vermeer NS, Straus SM, Mantel-Teeuwisse AK, Domergue F, Egberts TC, Leufkens HG, De Bruin ML. Traceability of biopharmaceuticals in spontaneous reporting systems: a cross-sectional study in the FDA Adverse Event Reporting System (FAERS) and EudraVigilance databases. *Drug Saf*. 2013;36(8):617–25.
58. Vermeer NS, Giezen TJ, Zastavnik S, Wolff-Holz E, Hidalgo-Simon A. Identifiability of biologicals in adverse drug reaction reports received from European clinical practice. *Clin Pharmacol Ther*. 2018;105(4):962–9.
59. van Aerts LA, De Smet K, Reichmann G, van der Laan JW, Schneider CK. Biosimilars entering the clinic without animal studies. A paradigm shift in the European Union. *MAbs*. 2014;6(5):1155–62.
60. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC.