PERSPECTIVE

feature

Q2 ## Streamlined approval of biosimilars: moving on from the confirmatory efficacy trial

Q3 **Marie-Christine Bielsky,** marie.bielsky@mhra.gov.uk**, Anne Cook, Andrea Wallington, Andrew Exley, Shahin Kauser, Justin L. Hay, Leonard Both and David Brown**

Licensing of biosimilars is essential to promote patient access to 21st-century biological medicines. Regulatory approval of biosimilars is based on the totality of evidence from a head-to-head comparison with reference products (RPs). A clinical efficacy trial is usually required, but this is increasingly questioned. Based on a thorough review of biosimilar applications in the European Union (EU), we conclude that in-depth knowledge of the reference product, allied with high-performing analytical tools, largely predicts clinical comparability, subject to confirmation by a comparative pharmacokinetic (PK) trial. We provide a blueprint for a biosimilar pathway that reduces the need for clinical efficacy trials in exceptional cases, together with qualifying criteria and requirements for streamlined assessment to expedite wider access to affordable biological medicines.

Features • PERSPECTIVE

Introduction

The first European regulatory guidelines on biosimilars were issued after 2004 following an amendment to the EU Directive 2001/83/EC relating to medicinal products for human use, which introduced the new concept of similar biological products [1]. These guidelines identified the type and amount of nonclinical and clinical data to be provided in addition to the pharmaceutical and bioequivalence studies required for generics. This biosimilar approach was dependent on the state-of-the-art analytical procedures and manufacturing processes used at that time, as well as on clinical and regulatory experiences [2]. Since 2005, these guidelines have been regularly updated; for example, the systematic requirement for a comparative

animal repeat-dose toxicity study has been waived in view of its limited contribution to the evaluation of biosimilarity.

Regulatory approval of biosimilars is currently based on the totality of evidence from a head-to-head comparison with a RP, using a stepwise, albeit often overlapping, approach including analytical studies, both structural and functional; nonclinical in vivo studies, exceptionally required; followed by human studies investigating PK and pharmacodynamics (PD) where possible, as well as efficacy, safety, and immunogenicity [3–5]. Critically, the discriminatory power of these studies decreases in this order and current EU guidance enables a waiver of comparative clinical efficacy studies if a PD parameter is an accepted surrogate marker for efficacy, such as

for insulin [6], low-molecular-weight heparin (LMWH) [7], and granulocyte-colony stimulating factor (G-CSF) [8]. However, given that, typically, no surrogate PD marker exists, particularly for the more complex biosimilars, an efficacy trial is required in an appropriately sensitive 'clinical model' (in terms of target population, dose, and treatment effect) able to detect clinically relevant differences should they exist. Importantly, clinical data cannot be used to justify substantial differences in analytical characteristics.

In 2020, it is increasingly questioned whether clinical efficacy trials of biosimilars are still necessary [9–11]. Revisiting the requirement for efficacy trials appears timely because significant technological advances have been made over the past 10 years and considerable scientific and

regulatory experience has been accrued since the first biosimilar was licensed in 2006.

Limitations of comparative efficacy trials in the assessment of biosimilars

Efficacy trials are typically equivalence trials and the margins need to be statistically and clinically justified. However, if margins were chosen to truly rule out all differences that could be of clinical importance, large numbers of patients (up to a few thousand) would be required. The situation becomes even more problematic when the treatment effect is small, which is often the case in oncology [12]. In the extreme situation of orphan or paediatric drugs, underpowered comparative trials will be conducted, which are unlikely to reach definite conclusions. The sensitivity of the 'clinical model' is questioned further when the RP is part of a combination therapy or when the therapeutic dose of the RP is on the plateau of the dose–response curve.

Although safety and immunogenicity data are also collected during efficacy trials, these trials are not powered for these endpoints and numerical differences in the rates of occurrence of adverse drug reactions (ADRs) and antidrug antibodies (ADAs) between the biosimilar and RP are expected to occur by chance. In particular, evaluation of the immune response in efficacy trials is confounded by multiple patient- and disease-related factors.

Furthermore, even if a clinical trial could be conducted in a limitless number of patients, clinical differences between the biosimilar and RP (in efficacy, safety, or immunogenicity) would be unlikely once analytical and PK comparability has been shown.

Overall, these efficacy trials are neither an effective discriminating tool for the comparison between biosimilar and RP, nor an efficient use of limited resources. Moreover, even when differences were observed in clinical trials, they did not, on their own, preclude the approval of the biosimilar, as reflected in our review.

Experience from EU applications for complex biosimilars

We conducted a detailed review of the applications for all 20 different complex biosimilar products related to six RPs [five monoclonal antibodies (infliximab, rituximab, adalimumab, bevacizumab, and trastuzumab) and a fusion protein, etanercept], which were evaluated and approved in the EU by the end of 2019. We consider that our selection of complex molecules (as opposed to simpler proteins) for which no PD marker exists, provides the most problematic case studies with which to test our

proposed framework. Nevertheless, we acknowledge that first-generation biosimilars were occasionally approved despite differences reported in the efficacy trials (e.g., somatropin and epoetin) after appropriate justification or additional studies; these differences largely reflect issues resolvable with current quality measures.

The tabulated summary of the quality and clinical results is based on the information provided in the European Public Assessment Reports available on the European Medicines Agency (EMA) website (Table 1). The main findings are discussed hereafter.

Quality results
The guiding principle is that the variability of the RP provides a reference range that is unlikely to be of clinical relevance. Sufficient batches of RP should be analysed to reflect the variability of the RP over a period of time.

Chemical differences
The most frequent notable differences observed between the biosimilar candidate and RP were an increase in high-molecular-weight (HMW) species, particles, or aggregates, which could increase immunogenicity; increased deamidation in the complementarity-determining regions (CDR), which can decrease antibody specificity and affinity, with an impact on biological activity; an increase in methionine oxidation, which can decrease binding to the neonatal Fc receptor (FcRn) and, hence, increase drug clearance; and differences in C-terminal lysine truncation, which are not considered to have any impact, because C-terminal lysine is removed within 1–2 h in serum. Where differences are relatively minor, these can be assessed as unlikely to have any clinical impact (e.g., differences in HMW species when overall levels are below 1%, or values just outside the range of RP batch data). Other differences, for example in the level of acidic, main, or basic species, should be evaluated further by additional testing to determine whether these impact the biological activity. Differences in charged species are often attributed to differences in the C-terminal lysine levels.

Glycosylation differences
The most frequent notable differences observed between the biosimilar candidate and RP were increases in afucosylated glycans (including high mannose glycans), which might increase affinity for Fc-gamma receptor (FcγR) IIIa and, hence, increase antibody-dependent cell-mediated cytotoxicity (ADCC); increases in sialylated

glycans, which could decrease affinity for FcγRIIIa and, hence, decrease ADCC; increases in galactosylated glycans, which might increase binding to complement component C1q and, hence, increase complement-dependent cytotoxicity (CDC); increases in high mannose glycans, which can increase drug clearance via a mannose receptor-mediated mechanism; and the elimination of nonhuman glycan structures (e.g., alfa 1,3-galactose and N-glycolylneuraminic acid) when the biosimilar is produced in Chinese hamster ovary (CHO) mammalian cell lines rather than in SP2/0 murine cell lines, which can decrease immunogenicity.

There are multiple examples of biosimilars with differences in afucosylation and FcγR binding. For some infliximab biosimilars, ADCC was not affected (Flixabi and Zessly); for Remsima/Inflectra, ADCC was decreased in the most sensitive in vitro models and additional tests were required to further confirm that the differences were not relevant in physiological conditions. For etanercept biosimilars, differences in ADCC were not considered clinically relevant because ADCC is not a mechanism of action (MOA) for this molecule, but the potential impact of increased ADCC on safety (e.g., cytokine release syndrome) was also addressed. For trastuzumab biosimilars, changes in the level of afucosylation with lower ADCC activity occurred transiently for multiple batches of the RP (Herceptin with expiry date from August 2018 to ~June 2019) [13]. When such batches were used in the efficacy trial, differences in favour of the biosimilar were reported (Ontruzant and Kanjinti), which did not preclude a final conclusion of biosimilarity [14]. Differences in the level of high mannose glycans were observed in several biosimilars, but the pivotal PK trials enabled the conclusion that they did not affect exposure to the biosimilar for bevacizumab (Mvasi and Zirabev), adalimumab (Halimatoz and Idacio), trastuzumab (Ogivri) or rituximab (Rixathon).

PK trial results
Pivotal PK trials were typically conducted in healthy volunteers (HV), although this was not possible for rituximab because of long-lasting B cell depletion at the therapeutic dose. The standard design was a parallel 3-arm trial (between 30 and 100 HV s per arm) comparing the biosimilar candidate to both EU and US RPs. For three biosimilar candidates, a first PK trial failed to show PK comparability before a second trial was successful (etanercept Erelzi and adalimumab Cyltezo and Halimatoz). No product-related cause was identified in these cases and the initial failure was mainly attributed to underestimated

TABLE 1

Summary of the results of the comparability exercise for all complex biosimilar products approved in the EU until December 2019[a,b,c,d]

Biosimilar	Quality differences						Results of clinical trials						Added value	
	Formulation	Analytical	Receptor binding	Biological activities	HMW/ particles/ aggregates	Comment	PK equivalence: number and type of subjects (source of RP)		Efficacy equivalence: indication and combined treatment; number of patients		Safety similarity	Immunogenicity similarity	PK trial	Efficacy trial
Enbrel (etanercept)														
Benepali	Y	Y	Y (FcγR)	Y (ADCC)	Y	ADCC not MOA	N = 138 HV (EU/US RP)	Y	RA + MTX; N = 596	Y	Similar	Similar (low)	Formulation, immunogenicity	N
Erelzi	Y	Y	N	Y (ADCC and CDC)	Y	ADCC/CDC not MOA	N = 108 HV	N	Psoriasis; N = 531	Y	Similar	No ADA detected with biosimilar	Formulation, immunogenicity, clearance	N
							N = 108 HV (EU RP)	Y						
Remicade (infliximab)														
Remsima, Inflectra	N	Y	Y (FcγR)	Y (ADCC)	Y	Additional functional testing: no difference	AS; N = 250 (EU RP)	Y	RA + MTX; N = 606	Y	Similar	Similar	Immunogenicity	N
Flixabi	N	Y	Y (FcγR, FcRn, C1q)	N	Y		N = 159 HV (EU/US RP)	Y	RA + MTX; N = 584	Y	Similar	ADA + rate ↑ in both trials	Clearance, immunogenicity	N
Zessly	Y	Y	Y (FcγR, FcRn)	N	N		N = 151 HV (EU/US RP)	Y	RA + MTX; N = 650	Y	Similar	Similar	Formulation, clearance, immunogenicity	N
Humira (adalimumab)														
Amgevita	Y	Y	(–)	N	N		N = 203 HV (EU/US RP)	Y	Psoriasis; N = 350	Y	Similar	ADA + rate ↓ in PK trial, similar in efficacy trial	Formulation, local tolerance, immunogenicity	N
Imraldi	Y	Y	N	Y (CDC)	Y	Similarity justified	N = 189 HV (EU/US RP)	Y	RA + MTX; N = 544	Y	Similar	Similar	Formulation, local tolerance, immunogenicity	N
Cyltezo	Y	(–)	Y (FcγR)	Y (reverse signalling, CDC, ADCP)	Y	(no details) Similarity justified (no details)	N = 193 HV	N	RA +MTX; N = 645	Y	Similar	Similar	Formulation, local tolerance, immunogenicity	N
							N = 324 HV (EU/US RP)	Y						
Halimatoz, Hefiya, Hyrimoz	Y	Y	Y (mTNFα, C1q)	Y (ADCC)	Y	Specification limits to control ADCC	N = 219 HV	N	Psoriasis; N = 465	Y[e]	Similar	ADA+ rate↓ in PK and efficacy trials	Formulation, clearance, local tolerance, immunogenicity	N
							N = 318 HV (EU/US RP)	Y						
Hulio	Y	Y	N	N	Y		N = 180 HV (EU/US RP)	Y	RA + MTX; N = 728	Y	Similar	Similar	Formulation, local tolerance	Y[f]
Kromeya, Idacio	Y	Y	Y (FcγR)	Y (ADCC)	N	Additional functional testing: no difference	N = 237 HV (EU/US RP)	Y	Psoriasis; N = 443	Y[g]	Similar	Similar	Formulation, clearance, local tolerance	N
Avastin (bevacizumab) Avastin (bevacizumab)														
Mvasi	N	Y	Y (FcγR)	N	Y		N = 202 HV (EU/US RP)	Y	Metastatic NSCLC + chemo; N = 642	Y	Similar	Similar	Clearance	N
Zirabev	Y	Y	Y (FcγR)	Y (growth inhibition)	Y	Comparable potency	N = 102 HV (EU/US RP)	Y	Metastatic NSCLC + chemo; N = 719	Y	Similar	Similar	Clearance	N
Herceptin (trastuzumab) Herceptin (trastuzumab)														

TABLE 1 (*Continued*)

Biosimilar	Quality differences						Results of clinical trials						Added value	
	Formulation	Analytical	Receptor binding	Biological activities	HMW/particles/aggregates	Comment	PK equivalence: number and type of subjects (source of RP)		Efficacy equivalence: indication and combined treatment; number of patients		Safety similarity	Immunogenicity similarity	PK trial	Efficacy trial
Ontruzant	N	Y	Y (FcγR)	Y (ADCC)	N	Comparable ADCC except for some EU Herceptin batches with lower ADCC	N = 108 HV (EU/US RP)	Y	HER2 + EBC or locally advanced BC + chemo; N = 875	N[h]	Similar	Similar (low)	Clearance, IRR	N
Herzuma	Y	Y	Y (FcγR)	N	Y		N = 70 HV (US RP)	Y	HER2 + EBC + chemo; N = 549	Y	Similar	Similar (low)	Clearance, formulation, IRR, immunogenicity	N
Kanjinti	N	Y	Y (FcγR, C1q)	Y (ADCC)	Y	Comparable ADCC except for cluster of Herceptin batches with lower ADCC; CDC not MOA	N = 157 HV (EU/US RP)	Y	HER2 + EBC + chemo; N = 725	N[i]	Similar	Similar (low)	Clearance, IRR, immunogenicity	N
Trazimera	Y	Y	N	N	N		N = 105 HV (EU/US RP)	Y	HER2 + MBC + chemo; N = 707	Y	Similar	Similar (low)	Clearance, formulation, IRR	N
Ogivri	Y	Y	N	N	N		N = 132 HV (EU/US RP)	Y	HER2 + MBC + chemo; N = 500	Y	Similar	Similar (low)	PK clearance, formulation, IRR	N
MabThera (rituximab)														
Truxima, Blitzima, Ritemvia	N	(–)	N	N	(–)		RA + MTX; N = 154 (pivotal; EU RP)	Y	RA +MTX; N = 372	Y	Similar	Similar	IRR, immunogenicity	N
							N = 189 (supportive, EU/US RP)		AFL +chemo; N = 140	Y[j]				
Rixathon, Riximyo	N	Y	N	N	N		RA + MTX, N = 173 (EU RP)	Y	AFL + chemo; N = 629	Y[k]	Similar	ADA+ rate ↓ in RA, similar (very low) in AFL	Clearance, immunogenicity	N

[a]Abbreviations: AFL, advanced follicular lymphoma; AS, ankylosing spondylitis; chemo, chemotherapy; EBC, early breast cancer; EU RP, EU-sourced reference product; IRR, infusion-related reaction; MBC, metastatic breast cancer; mTNFα, membrane tumour necrosis factor α; MTX, methotrexate; NSCLC, nonsmall cell lung cancer; US RP, US-sourced reference product.
[b]Y, yes; N, no; (–), no details.
[c]↓ or ↑: increased or decreased for the biosimilar compared with the RP.
[d]N = total number of subjects randomised in the trial.
[e]Equivalence in the US group; lower for the biosimilar in the small EU group.
[f]Immunogenicity because of unreliable assays in the PK trial.
[g]Much higher efficacy response than expected.
[h]Upper limit of equivalence margin exceeded.
[i]Non-inferiority trial.
jEquivalence for overall response rate (primary endpoint) but lower progression-free survival for the biosimilar.

intersubject variability in parallel arm design trials (e.g., sex, body weight, and injection site). Confirming PK comparability was especially relevant when the formulation was different or when quality differences had been detected (e. g., differences in the level of high mannose or in FcRn binding). The pivotal PK trials also provided information on safety (infusion-related reactions and local tolerance of subcutaneous injections) and immunogenicity, for example when small differences in protein aggregation had been detected.

Efficacy trial results

Efficacy trials were usually conducted in a single indication and enrolled between 350 and 900 patients in total. They showed similar efficacy within the predefined equivalence margins except in the two cases of trastuzumab biosimilars discussed earlier. Efficacy issues were also raised for three other biosimilars. First, infliximab Flixabi showed a higher numerical ADA incidence rate, which was associated with slightly lower efficacy in patients with rheumatoid arthritis (RA); long-term data from a preplanned 78-week extension of the pivotal trial confirmed that this difference was not clinically meaningful [14]. Rituximab Rixathon showed a lower progression-free survival despite similar objective response rates, but data were still immature [14]. Lastly, the efficacy trial for adalimumab Halimatoz, where 80% of the patients were from the USA, showed similar efficacy compared with US RP but the few EU centres suggested lower efficacy for the biosimilar versus the EU RP. No conclusion could be drawn because of differences in the baseline characteristics of US and EU patients, together with the small EU sample size, and the observation was most likely a chance finding. None of these differences precluded a conclusion of biosimilarity.

Our review did not identify any instance where efficacy trials added crucial information to establish biosimilarity. Moreover, the differences in efficacy findings identified in a few cases were finally discarded based on the remaining evidence (i.e., analytical testing and clinical PK). Finally, we have not found any publicly available results of an efficacy trial being the critical reason for not submitting a marketing authorisation application for a biosimilar candidate. We also reviewed the EU applications that received a negative Committee for Medicinal Products for Human Use (CHMP) opinion and for which a withdrawal/refusal assessment report is available on the EMA website. Three insulin products (from the same applicant) and three different pegfilgrastim products

presented with manufacturing issues and/or comparability issues in their PK/PD pivotal trial; these issues were subsequently resolved for two pegfilgrastim biosimilars (Fulphila and Ziextenzo). For two other biosimilar candidates, interferon-alfa-2a (Alpheon) and rituximab (from Mabion), major manufacturing and analytical comparability issues were raised. In both cases, the efficacy trials met their primary and secondary endpoints, but some differences were pinpointed in ancillary or subgroup analyses; these were highlighted because of the quality issues but would not have precluded an approval on their own [15,16].

Therefore, we contend that, in most cases, an efficacy trial is no longer required to establish biosimilarity once comparability has been demonstrated through the analytical and clinical PK studies. We present hereafter the rationale for this new paradigm and a working blueprint, which details the qualifying criteria and documentation required for the marketing authorisation application of a biosimilar without an efficacy trial.

Rationale for allowing biosimilar applications without comparative efficacy trials

In-depth clinical knowledge of the RP

The efficacy of the RP can usually be derived from its MOA, acknowledging that complex biologicals often exhibit several MOAs. Likewise, its safety profile can be mainly predicted from exaggerated and adverse pharmacological effects at the target. Other common ADRs are injection-related reactions, some being mediated by an ADA response, which differs widely across products in terms of incidence, type, chronology, and effects. Extensive post-licensing experience in clinical practice further delineates the crucial aspects of efficacy, safety, and immunogenicity of the RP that emerged during clinical development.

High-performing analytical tools

Comparative analytical studies follow a scientific evidence-based approach using state-of-the-art methods capable of detecting small differences in the physicochemical characteristics, target, or receptor binding and various bioactivities between the RP and biosimilar candidate [17]. There is an extensive literature on the impact of molecular differences on the binding and functional properties of biologicals, and further on their immunogenic potential and PK profile [18–22]. An increasing understanding of the MOA(s) of the RP in the target diseases highlights which physicochemical and functional

properties are more likely to relate to the clinical outcomes. The guiding principle is that the variability of the RP provides a reference range that is unlikely to be of clinical relevance. The continual improvement in analytical methods has led to a greater understanding of how to control crucial factors in the production of complex biologicals, resulting in biosimilar candidates that are closely aligned with the RP.

Therefore, if the analytical profile of the candidate biosimilar is highly similar to that of the RP, there is no scientific reason for its PK, efficacy, safety, and immunogenicity to be different. If the binding and functional assays cover all potential MOAs, an efficacy trial in patients is not expected to show differences undetected by extensive analytical testing. Likewise, differences in chemical and post-translational modifications that might impact immunogenicity, especially those affecting protein aggregation, are detected through analytical tests [17,23,24], which are more sensitive than the evaluation of the human immune response because of multiple confounding patient- and disease-related factors.

In practice, some analytical differences are always observed given the power of the current tools. Differences outside the reference range determined for the RP cannot be accepted for analytical characteristics if these are relevant to clinical efficacy and safety outcomes. Exceptionally, an in vivo nonclinical study might be required to evaluate the potential impact of some differences (e.g., a novel excipient in the formulation). Overall, differences might be accepted only if there is enough scientific evidence to conclude that they do not impact efficacy and safety.

Confirmatory PK trial

Although PK comparability can be predominantly concluded from analytical comparability, a pivotal comparative PK trial is still considered necessary. It allows for the comparison of the bioavailability of the formulated products, which cannot be predicted only from the analytical comparison because the formulations of the RP and biosimilar are not required to be identical [3,4]. This is particularly important in the case of a novel excipient or a different administration device. Even if the product is administered intravenously, differences in distribution and elimination might be related to physicochemical differences; therefore, this comparison is helpful to confirm clinically that these differences are not relevant. Finally, it also provides some safety and immunogenicity information from exposure in humans.

Features • PERSPECTIVE

Robust pharmacovigilance

We are confident that pharmacovigilance systems and risk management planning are sufficiently robust to detect safety signals in post-marketing use. However, more than a decade of clinical experience indicates that a new safety signal solely identified with a biosimilar is extremely unlikely unless there are additional factors unrelated to the active molecule itself (e.g., different excipient or device). Accurate traceability of the biosimilars is essential to effective pharmacovigilance and has improved by insisting on clear identification of brand name and batches [25]. As part of additional pharmacovigilance activities, participation in established disease registries has been required for some of the complex biosimilars (infliximab, etanercept, and adalimumab) when already requested for the RP in the EU. These registries have provided reassurance about the incidence of rare events and about the safety of switching between RP and biosimilars [26,27].

Requirements for an application without an efficacy trial
Prerequisites for the reference product

Biosimilar development starts with in-depth knowledge of the RP. Its main and potential secondary MOAs in the target diseases need to be known and demonstrable. As a result, the physicochemical and functional properties that are more likely to relate to the clinical outcomes will be identified, which will allow definition of critical quality attributes (CQA); that is, the quality attributes that would need to be controlled or maintained within appropriate limits or ranges to ensure that clinical effects of the product are not impacted. If the main MOA of a complex RP is unknown, it is likely that an efficacy trial will be required.

The PK, efficacy, safety, and immunogenicity profiles of the RP need to be well characterised and the impact of ADAs on PK (increased or decreased clearance), efficacy (neutralising ADAs), and safety (injection-related reactions) need to be known. Population PK or PK–PD models available in the literature should be taken into consideration [28]. If there are notable uncertainties around these clinical aspects, the need for a comparative efficacy/safety trial is likely.

Quality documentation
Extensive quality data generated for the RP

A robust data quality package is fundamental to the claim that the biosimilar candidate is highly similar to the RP. The quality target product profile (QTTP) for development of the biosimilar should be generated by analysis of multiple different batches of RP sourced from the appropriate market, over a period of months to years to reflect the manufacturing variability of the RP. The number of RP batches regarded as sufficient depends on the CQA tested and the data generated, (e.g., the determination of biological activity will require many batches, whereas fewer batches are needed for confirmation of the higher order structure). The RP batches should be stored under recommended (label) conditions and tested within their shelf life. Age at the time of testing (relative to expiry date) should be considered. CQAs should be defined at an analytical and in vitro functional level. Analytical methods need to be sensitive, qualified, and sufficiently discriminatory to detect possible differences. Robust data require the application of orthogonal methods. Functional assays should be relevant to the MOA in all therapeutic indications [e.g. ADCC, CDC, antibody-dependent cellular phagocytosis (ADCP), binding to Fc receptors, apoptosis, etc.] and detailed in the submission.

Extensive quality data generated for the biosimilar candidate

Comparative analytical studies should follow a scientific evidence-based approach deploying state-of-the-art methods with the sensitivity to detect small differences in the physicochemical characteristics, target or receptor binding, and various bioactivities between the RP and biosimilar candidate [17]. As a guiding principle, the range of variability of the RP should provide a reference range that is unlikely to be of clinical relevance.

All relevant batches of the biosimilar candidate, including process performance qualification, stability, and clinical study batches, should be analysed and compared with the RP. Ideally, batches of the biosimilar candidate would undergo analysis of biological activity (e.g., cell-based assays) at the same time as RP batches. Otherwise, an established and characterised reference standard, calibrated against international reference standards wherever possible, should be used to minimise day-to-day assay variability [29,30].

It is a key principle that the biosimilar candidate data should be within the analytical range established for the RP. Wide 'similarity ranges' based on inappropriate statistical methods should not be used. It is also expected that any differences between the biosimilar candidate and RP ranges will be further characterised and justified as unlikely to impact on clinical efficacy and safety, based on specific experiments and available literature [17–24]. For example, reduced binding affinity to FcγRIIIb (by SPR) could be acceptable if ex vivo experiments using isolated neutrophils demonstrated comparable binding. Likewise, an increase or decrease in levels of C-terminal lysine could be acceptable because it has been shown to be removed in serum within 1–2 h [31]. By contrast, any differences that might have an impact (e.g., increased levels of aggregation) should be reduced by modifications in the manufacturing process, rather than being justified.

The potential impact of a different formulation should be discussed in terms of product stability or clinical characteristics. The manufacturing process should be well controlled. Release and stability specification limits should be based on RP and biosimilar batch data, to prevent future 'drift' or 'shift' in CQAs after demonstration of biosimilarity. The specification limits should be stringent at the time of approval of the biosimilar, because, afterwards, each product will have their own lifecycle [3,4].

Key requirements for the quality assessment are summarised in Box 1.

Clinical documentation
An appropriately designed PK study with safety and immunogenicity evaluation

It is imperative that the comparative PK study is robust and demonstrates equivalence of the primary PK parameters. As such, certain criteria need to be considered and, where relevant, adhered to (Box 2). PD markers are lacking for most products other than those covered by specific guidelines (insulin, G-CSF, and LMWH). If available, PD parameters can be measured during the PK trial and descriptive results should be presented to support a conclusion of biosimilarity. Although this is valuable additional information, PD is usually less discriminating than PK and, therefore, not considered an essential requirement.

Justification for comparable efficacy

Although precise correlations between clinical efficacy and pharmacological effects are usually lacking, the efficacy of the RP can usually be derived from its MOAs. Therefore, a justification should be provided that comparable efficacy can be derived from comparable binding properties and functional characteristics (Box 1). Any observed differences must be justified as not clinically relevant, based on specific experiments and available literature.

Justification for comparable safety and immunogenicity

The safety profile of the RP is largely predicted from on-target side effects (i.e., exaggerated and

BOX 1

Key requirements for the quality assessment

In-depth knowledge of the RP
The main MOA is known and is demonstrable
CQAs are known
Sufficient batches of the RP are analysed to reflect the variability of the RP over a period of time, within the stated shelf-life
Analytical methods are sensitive, qualified, and sufficiently discriminatory, with orthogonal methods used wherever possible
Range of variability is defined at analytical and in vitro functional levels
Functional assays are relevant for the MOA in all indications
Attributes of the biosimilar
All relevant batches representative of the commercial scale demonstrate variability within the analytical range established for the RP. Wide 'similarity ranges' based on inappropriate statistical methods should not be used
Any differences between the RP and biosimilar are characterised and justified as unlikely to impact clinical efficacy and safety
Additional orthogonal analysis, including in vitro functional assays (e.g., cell-based assays) are used to evaluate differences, as necessary
Differences that might have a clinical impact should be addressed by modifications in the manufacturing process
The manufacturing process is well controlled. Release and stability specification limits are stringent and based on RP and biosimilar batch data

BOX 2

Key requirements for the PK assessment

Robust clinical study design and statistical analysis plan
Subject selection: the most sensitive and homogeneous population (healthy subjects versus patient populations) to detect clinically meaningful differences in PK between RP and biosimilar
Parallel versus cross-over design should be justified
Selection of most sensitive dose(s) to detect clinically meaningful differences
Linear (nonspecific) clearance and nonlinear (target mediated) clearance should be addressed, such as selection of dose and need to assess partial areas under the curve (AUCs), etc.
The equivalence margins must be prespecified; 80.00–125.00% is generally acceptable but narrower limits might sometimes be required
Standardisation of investigational medicinal product (IMP) handling and administration as well as PK sampling, sample handling, and analysis
In a parallel group trial, covariates to be used in the analysis (e.g. body weight or subject sex) should be predefined
If appropriate population PK or PK–PD models are available for the RP in the literature, modelling and simulation should be considered for optimising study design (dose, subject selection criteria, and sample size)
Collection of safety and immunogenicity data with prespecified subgroup analysis of ADA-negative and ADA-positive subjects where high incidence of ADA formation is expected
Batch selection and protein content correction
Product intended for commercialisation should be used
Protein content adjustments should be prespecified
Interpretation of results
Extrapolated AUC > 20% in >20% of observations requires discussion of the validity of the study
In case of a failed PK study (i.e., 90% confidence intervals for the primary PK parameters are not contained completely within the prespecified acceptance limits), root cause analysis should be provided, with conclusions adequately reflected in the planning and conduct of a subsequent PK study. If no root cause is identified and another study is conducted and is positive, the initial study should not be ignored when reaching conclusions on PK similarity

Extensive clinical experience with the RP informs a risk-based assessment of the immunogenicity of the biosimilar, potential rates of binding or neutralising ADAs, and their clinical relevance. The pivotal PK trial should also demonstrate comparable rates of injection site and infusion-related reactions, as well as broadly comparable rates and characteristics of ADA (Box 3).

BOX 3

Key requirements for the immunogenicity assessment

RP-derived risk-based immunogenicity assessment
ADA rates and impact PK, efficacy, and safety in original studies defined
Data from formal postauthorisation studies and real-world clinical experience on ADA rates
Consensus view on clinical relevance of ADAs and potential mitigation measures (e.g., drug monitoring in selected cases)
Biosimilar attributes
Quality characteristics justify extrapolation to comparable safety and immunogenicity
Critical drug product release characteristics defined: protein aggregates, impurities, and stability
Data collected in the PK trial (ADA rates, titres, binding/neutralising, and impact on PK) support comparability
Exceptional circumstances excluded
Potential for neutralising Abs against critical endogenous factors (e.g., epoetin), requiring clinical studies in patients

BOX 4

Key requirements for pharmacovigilance

RP-derived risk management plan
Safety concerns, including any additional concerns, where needed
Any additional pharmacovigilance measures, including participation in ongoing disease registries for the RP
Additional risk minimisation measures (e.g., educational materials)
Biosimilar attributes to aid product traceability
Use of distinguishable brand name
Peel off labels detailing brand name and batch number to be placed in patients' notes after administration
In the future, use of 2D barcodes to enhance traceability
Other
Reinforced training for healthcare professionals and patients on ADR reporting by brand name and batch number

adverse pharmacological effects at the target, including on normal tissues). Other common ADRs are injection-related reactions, which are triggered by various mechanisms, some being mediated by ADAs. The quality attributes and formulation of the biosimilar should form the basis for justification that the safety and immunogenicity are comparable to the RP.

Risk management plan

The risk management plan, an EU requirement, should reflect that of the RP (Box 4). The biosimilar candidate should be included in ongoing disease registries, preferably those already in place for the RP, to enable collection of real-world information to support detection of potential safety signals related to the RP and its biosimilars according to their use in clinical practice. A patient registry restricted to the follow-up of a cohort of patients treated with the biosimilar would not normally be required because this information in isolation (with no comparator, using biosimilar-specific methodology) is not considered optimal. Any additional risk minimisation measures requested for the RP should also be implemented for the biosimilar, such as educational materials or patient alert cards.

Recommendations

This evidence-based approach should be applicable to most biosimilar candidates and would save time and drug development resources. The applicant should be able to justify the absence of a clinical comparative efficacy trial by submitting a robust quality package and a comparative clinical PK study supported by well-documented arguments. Potential exceptions include biosimilars where the main MOA of the RP is not known or where it is difficult to predict the impact of analytical differences that have not been resolved by adaptations to the manufacturing process. Therefore, a comparative head-to-head efficacy trial might still be required on a case-by-case basis.

Exceptionally, additional clinical safety data might be required where safety uncertainties cannot be resolved without patient exposure prelicensing. For example, epoetins can induce neutralising ADA that cross-react with endogenous erythropoietin and lead to pure red cell aplasia. This rare serious ADR was reported with an increased incidence after a manufacturing change in the RP, where the possible root cause was leaching from uncoated rubber stoppers into the syringes of organic compounds that acted as adjuvants. The same reaction was later reported for a biosimilar candidate and attributed to contamination by tungsten during the manufacture of the syringes causing protein denaturation and aggregation [32]. Hence, for serious ADRs with unpredictable root causes, exposure of a patient cohort to the biosimilar candidate is considered the most appropriate approach to resolve any residual uncertainty around immunogenicity and safety.

Finally, regardless of the biosimilar regulatory pathway, convincing evidence of biosimilarity should be available before marketing authorisation. Postapproval trials or real-world data cannot be proposed by applicants to resolve remaining efficacy or safety uncertainties because patients should not be exposed to additional risks when they can be treated with an effective and safe RP.

Concluding remarks

Our review indicates that comparative efficacy trials are neither an effective discriminating tool for the comparison of the biosimilar with the RP, nor an efficient use of limited resources. This conclusion is based on the scientific and regulatory experience accrued since the first biosimilar was licensed in the EU in 2006, significant technological advances in the past 10 years, a review of the more complex EU biosimilar applications, and the overwhelmingly positive clinical experience with biosimilars to date.

We consider that extensive comparative analytical studies, together with an abbreviated clinical package consisting of a PK trial with simple PD (if available), safety and immunogenicity evaluation, is sufficient to assess biosimilarity in most cases. We have identified the key prerequisites for an RP that might qualify for this pathway and the key requirements for comparing the RP and biosimilar candidate using a discriminatory range of structural and functional studies, together with crucial PK parameters. This regulatory approach reduces the need for clinical efficacy/safety trials of biosimilars to exceptional cases and provides a streamlined pathway for expediting approval of biosimilars, thereby promoting wider patient access to these crucial therapies.

Disclaimer

This publication reflects the views of the individual authors and should not be understood to represent official views or policies of the UK Medicines and Healthcare products Regulatory Agency (MHRA). Andrew Exley is now employed by Adept Biologica Consulting Limited.

References

1 European Parliament and Council of the European Union (2001) Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. *Off. J. Eur. Union* 311, 67–128

2 EMA (2005) *Guideline on Similar Biological Medicinal Products.* EMA

3 EMA (2014) *Guideline on Similar Biological Medicinal Product.* EMA

4 EMA (2014) *Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Quality Issues.* EMA

5 EMA (2014) *Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Non-Clinical and Clinical Issues.* EMA

6 EMA (2015) *Guideline on Non-Clinical and Clinical Development of Similar Biological Medicinal Products Containing Recombinant Human Insulin and Insulin Analogues.* EMA

7 EMA (2016) *Guideline on Non-Clinical and Clinical Development of Similar Biological Medicinal Products Containing Low-Molecular-Weight-Heparins.* EMA

8 EMA (2006) *Guidance on Similar Medicinal Products Containing Recombinant Granulocyte-Colony Stimulating Factor.* EMA

9 Frapaise, F.X. (2018) The end of Phase 3 clinical trials in biosimilars development? *BioDrugs* 32, 319–324

10 Schiestl, M. *et al.* (2020) The path towards a tailored clinical biosimilar development. *BioDrugs* 34, 297–306

11 Webster, C.J. *et al.* (2019) An Efficient development paradigm for biosimilars. *BioDrugs* ZZ, XXX–YYY

12 Schellekens, H. *et al.* (2016) Safety and efficacy of biosimilars in oncology. *Lancet Oncol.* 17, e502–e509

13 Kim, S. *et al.* (2017) Drifts in ADCC-related quality attributes of Herceptin(R): Impact on development of a trastuzumab biosimilar. *MAbs* 9, 704–714

14 Wolff-Holz, E. *et al.* (2019) Evolution of the EU Biosimilar Framework: past and future. *BioDrugs* ZZ, XXX–YYY

15 EMA (2006) *Refusal Assessment Report for Alpheon.* EMA

16 EMA (2019) *Withdrawal Assessment Report Rituximab Mabion.* EMA

17 Duivelshof, B.L. *et al.* (2019) Glycosylation of biosimilars: recent advances in analytical characterization and clinical implications. *Anal. Chim. Acta* 1089, 1–18

18 Beck, A. and Liu, H. (2019) Macro- and micro-heterogeneity of natural and recombinant IgG antibodies. *Antibodies* 8 XXX–YYY

19 Goetze, A.M. *et al.* (2011) High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans. *Glycobiology* 21, 949–959

20 Higel, F. *et al.* (2016) N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* 100, 94–100

21 Liu, L. (2015) Antibody glycosylation and its impact on the pharmacokinetics and pharmacodynamics of monoclonal antibodies and Fc-fusion proteins. *J. Pharm. Sci.* 104, 1866–1884

22 Reusch, D. and Tejada, M.L. (2015) Fc glycans of therapeutic antibodies as critical quality attributes. *Glycobiology* 25, 1325–1334

23 Doevendans, E. and Schellekens, H. (2019) Immunogenicity of innovative and biosimilar monoclonal antibodies. *Antibodies* 8, XXX–YYY

24 Jefferis, R. (2016) Posttranslational modifications and the immunogenicity of biotherapeutics. *J. Immunol. Res.* 2016, 5358272

25 Vermeer, N.S. *et al.* (2019) Identifiability of biologicals in adverse drug reaction reports received from European clinical practice. *Clin. Pharmacol. Ther.* 105, 962–969

Q6

Q7

Q8

Q5

Features • PERSPECTIVE

26 Glintborg, B. et al. (2019) To switch or not to switch: results of a nationwide guideline of mandatory switching from originator to biosimilar etanercept. One-year treatment outcomes in 2061 patients with inflammatory arthritis from the DANBIO registry. *Ann. Rheum. Dis.* 78, 192–200

27 Glintborg, B. et al. (2017) A nationwide non-medical switch from originator infliximab to biosimilar CT-P13 in 802 patients with inflammatory arthritis: 1-year clinical outcomes from the DANBIO registry. *Ann. Rheum. Dis.* 76, 1426–1431

28 Wang, Y.C. et al. (2019) Role of modeling and simulation in the development of novel and biosimilar therapeutic proteins. *J. Pharm. Sci.* 108, 73–77

29 Metcalfe, C. et al. (2019) The first World Health Organization International Standard for infliximab products: a step towards maintaining harmonized biological activity. *MAbs* 11, 13–25

30 Prior, S. et al. (2018) 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* 10, 129–142

31 Cai, B. et al. (2011) C-terminal lysine processing of human immunoglobulin G2 heavy chain *in vivo*. *Biotechnol. Bioeng.* 108, 404–412

32 Seidl, A. et al. (2012) Tungsten-induced denaturation and aggregation of epoetin alfa during primary packaging as a cause of immunogenicity. *Pharm. Res.* 29, 1454–1467

Please cite this article in press as: Bielsky, M.C. et al.