Biosimilars: Harmonizing the Approval Guidelines

Sarfaraz K. Niazi

College of Pharmacy, University of Illinois, Chicago, IL 60612, USA; niazi@niazi.com; Tel.: +1-312-297-0000

Abstract: Biosimilar approval guidelines need rationalization and harmonization to remove the inconsistencies and misconceptions to enable faster, safer, and more cost-effective biosimilars. This paper proposes a platform for a model guideline based on the scientific evaluation of the regulatory filings of the 130+ products approved in the US, UK, and EU and hundreds more in the WHO member countries. Extensive literature survey of clinical data published and reported, including Clinicaltrials.gov, a review of all current guidelines in the US, UK and EU, and WHO, and detailed discussions with the FDA have confirmed that removing the animal and clinical efficacy testing and fixing other minor approaches will enable the creation of a harmonized guideline that will best suit an ICH designation.

Keywords: biosimilars; WHO; MHRA; BPCIA; FDA; EMA; ICH; approval guidelines; analytical assessment; animal testing; efficacy testing; immunogenicity testing; biological drugs

1. Introduction

Biosimilars are recombinant DNA products that join DNA from different species and subsequently insert the hybrid DNA into a host cell, often a bacterium or mammalian cell, to express the target protein; this molecular chimera was first created by researchers from UC San Francisco and Stanford in 1972 [1,2]. Stanley Cohen of Stanford and Herbert Boyer of UCSF received the US patent in 1980. Boyer co-founded Genentech, Inc. in 1976. The Cohen-Boyer patents will eventually have more than 500 licensees to biotechnology and pharmaceutical companies and earn Stanford and UCSF more than USD 250 million in royalties. These patents have now expired [3,4].

On 26 July 1974, ten researchers, including six future Nobel Laureates (James Watson, Paul Berg, Stanley Cohen, David Baltimore, Ronald Davis, and Daniel Nathans), wrote a letter to the most reputable journal Science [5] to urge the NIH to regulate the use of recombinant DNA technology and urged scientists to halt recombinant DNA experiments until they better understood whether the technique is safe. These concerns eventually led to the 1975 Asilomar Conference [6], where one hundred prominent scientists gathered to discuss the safety of manipulating DNA from different species. The meeting resulted in a set of NIH guidelines in 1976 that has been revised several times since then [7].

On 7 July 2022, the author published his letter in the same Science magazine that had carried the earlier warning, suggesting a ban on the animal testing of copies of approved recombinant DNA products, the biosimilars [8]. The US FDA is telling us not to test even new recombinant products in animals, except for carcinogenicity assessment [9]. Efforts are underway to forbid animal testing at the US Congressional level to remove the possibility of animal testing being used to justify analytical dissimilarity.

Biosimilars include monoclonal antibodies, cytokines, growth factors, enzymes, immunomodulators, and thrombolytics, proteins extracted from animals or microorganisms, including recombinant versions of these products (except clotting factors), and other non-vaccine therapeutic immunotherapies. Billions of patients receiving biosimilars have shown therapeutic equivalence [10–21]. None of these products have shown adverse events more than the reference product [22–28], including immunogenicity responses. It is estimated...
that the cumulative exposure to EU-approved biosimilars was more than two billion patient treatment days in 2020 [29], with no adverse event reporting or withdrawal from the market due to safety reasons and no biosimilar-specific adverse effects have been added to the product information [29,30]. Such an impeccable record of safety and efficacy that is better than the record for chemical drugs needs serious consideration about the regulatory guidelines to assure that we are not wasting resources and committing unethical practices.

The fast growth of these products has brought over 250 FDA-approved peptides and therapeutic protein products [31]. As the patents of these products expire, the need for cost-effective copies, the biosimilars, is accelerated.

The EMA introduced the first biosimilar guideline and approved the first product in 2006 [32]. The EU currently lists 95 centrally approved biosimilars [33]. The FDA brought its guidelines in 2009 [34] and has 37 products approved [35]. Both agencies have made public the details of the regulatory submissions of biosimilars. As of April 2022, 86 European Public Assessment Reports (EPAR) were available [36]. The FDA provides access to these data through its website portal on AccessData [37]. No other regulatory agency makes this information accessible. The WHO is not a regulatory agency; it only provides scientific advice to its 194 member countries.

As the end of the Brexit transition period approached last year, the MHRA released draft guidance for consultation that was finalized on 14 May 2022 [38]. This guideline will likely change how biosimilars are approved and concur with most of the suggestions made in this article.

Now with 17 years of the use of biosimilars, the safety and efficacy of biosimilars have been fully validated, with no recall; the same holds for all clinical trials conducted, including testing after switching and alternating, as suggested in the US. Thus far, no clinical efficacy study has failed to meet the acceptance criteria.

Biosimilars would have been treated like generic products if it was possible to declare them chemically equivalent. Thus, the backbone of the approval of biosimilars is their analytical assessment in a side-by-side comparison with the reference product. Recent advances in analytical sciences now allow more stringent evaluation, making all other tests less sensitive in identifying clinically relevant functional attributes.

After a biosimilar product meets the analytical similarity criteria, it is generally put through animal testing, despite the knowledge that the peculiar mechanism of action of biological products involves mainly receptor binding, which is not possible because of the lack of these receptors in most animal species.

After animal testing comes to the pharmacokinetic and pharmacodynamic profiling as an extension of the analytical assessment—essentially, to test how the body “sees” the molecule and how the molecule “sees” the body. These studies also establish bioequivalence and other pharmacokinetic parameters that may be able to tell the differences in the rate and extent of receptor binding. Recently, the author suggested novel protocols with narrow inclusion criteria and a two-dose, cross-over study to reduce the study cost. The pharmacodynamic comparisons are also made in the same study, well as immunogenicity responses.

The last stage of testing involves clinical safety and efficacy testing. Except for a small number, all biosimilar products approved conducted extensive clinical efficacy testing; none of these studies failed due to poor sensitivity, as observed in the case of animal testing. Therefore, it is proposed to disallow such studies for these studies for the same reason as suggested for animal testing—disallow approval of biosimilars based on fallacious clinical efficacy testing.

A recent report from Mckinsey & Co. states [39] that “the biosimilars industry needs to reduce its costs, particularly in drug development, to preserve its sustainability. A typical biosimilar today costs between USD 100 million and USD 300 million to develop, with clinical trials accounting for more than half of the budget. If biosimilars are to deliver the promise of making biological drugs accessible, these cost barriers must come down exponentially.
This paper presents a scientific and rational plan to harmonize the approval guidelines by removing irrelevant testing and resolving regulatory misconceptions to allow global acceptance of approved biosimilars.

2. Regulatory Background

Harmonizing biosimilar approval guidelines requires a clear understanding of the differences in the existing guidelines and their roots of inclusion. There is only one example of harmonization, albeit partial—the International Conference for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), which was established in 1990. In 2015, the organization and operations were renewed (this was called “ICH reform”) as a legal entity under Swiss law, and its name was changed to the International “Council” for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) [40]. While the ICH guidelines are now widely accepted, they are limited to specific technical issues such as testing methods and qualification of manufacturing operations. The ICH guidelines do not recommend any approval process even among the three original members—US, EU, and Japan. After Brexit, the MHRA applied for full membership to ICH since it was engaged as a member of the EU in the past; this membership was approved on 16 June 2022.

It is the goal of this paper to suggest a similar harmonization to the ICH, without the limitations found in the WHO guidelines that too suggest a harmonized approach among its 194 countries but come short in scientific merit [41,42]. A harmonized guideline will be based only on scientific yet rational suggestions to enable a cost-effective development and cost reduction without compromising safety and efficacy. This harmonization will also allow global acceptance of registration, allowing companies to sell their products at substantially reduced prices because of the wider distribution possibility.

2.1. The US Scene

There are over 100 biosimilar programs enrolled with the FDA [43]. To expedite the approval process, the FDA has taken several significant steps.

The FDA has created two new guidelines, the extension of the Q&A presentations [44] and the third revised draft guidance [45] titled “New and Revised Draft Q&As on Biosimilar Development and the BPCI Act”. The details refer to fulfilling pediatric assessment or PREA requirements, post-approval filing, and asserting that the 351(k) cannot have a different route or dosage form. However, the strength issue was delayed, adding new indications and orphan exclusivity. The FDA also updated The Purple Book FAQ section [46].

FDA has also published new fact sheets to provide additional educational materials on biosimilar and interchangeable products and the biosimilar regulatory review and approval process.

The BPCIA states [47] that the “Secretary may determine, in the Secretary’s discretion, that an element described in clause (i) (I) [the biosimilar testing] is unnecessary in an application submitted under this subsection”. The FDA has subtly implemented this change in its new biosimilar guidance. However, unlike the EMA, MHRA, or WHO, the FDA is bound by the BPCIA legislation that states that “an application submitted under this subsection shall include information demonstrating that the biological product is biosimilar to a reference product based upon data derived from analytical studies, animal studies, and clinical studies”. The new education material includes the phrase “in addition to analytical studies, other studies that may be needed”, not shall be, as stated in the BPCIA.

The FDA posts details of its approval of biosimilar products. However, a biosimilar developer may object and secure under the Freedom of Information Act [48].

2.2. The EMA Scene

In 2001, much of the EU’s directive-based legislation concerning the regulation of medicines was codified as Directive 2001/83/EC. The EMA has issued concept papers, draft guidance, and public scientific workshops. The EMA’s Committee for Medicinal
Products for Human Use (CHMP) has also issued product class-specific guidance such as recombinant erythropoietin, granulocyte-colony stimulating factor, recombinant human soluble insulin, low-molecular-weight heparins, somatropin, and recombinant interferon alfa [49]. However, EMA has announced that they intend not to issue more specific biosimilar guidelines but instead prefer to give tailored advice on a case-by-case basis. This change in the EMA policy came from the FDA, and such guidelines can misdirect the development of biosimilars.

Patents are not a significant issue in the EMA filing; the litigation is left to the claiming parties. The patent laws in the EU are also different. The exclusivity for biological drugs is ten years in the EU and 12 years in the US, giving the EU filings at least a two-year head start. However, the ten years of exclusivity for patents and other exclusivity rights can last longer than ten years after market approval. In the EU, process patents are rarely awarded, reducing the significant barrier experienced by US filings, where the patent dance involves the product and a multitude of process patents. The differences in the patent laws between the US and the EU significantly impact the speed and scope of introducing biosimilars. This topic is of interest to determine whether a harmonized guideline should include the intellectual property issue, as elaborated later in this paper [50].

EMA guidelines and the decision-making of the EMA in approving biosimilars have evolved significantly; the EMA is now promoting removing animal testing, though it is not yet been made clear. In addition, like the FDA, EMA has recently begun approving biosimilars without requiring clinical efficacy testing.

2.3. The WHO Scene

The World Health Organization (WHO) is not a regulatory authority, but it is mandated to support regulatory authorities in its 194 Member States. The WHO guidelines on evaluating biosimilars [51] provide suggestions to National Regulatory Agencies (NRAs) principles for approving biosimilars.

In 2019, the WHO Expert Committee on Biological Standardization (ECBS) considered that a more tailored and potentially reduced clinical data package might be acceptable in cases where the available scientific evidence supported this. In addition, the committee endorsed the review of current scientific evidence to consider updating the Guidelines to provide more flexibility and clarity. Thus, the WHO reviewed scientific evidence and experience to identify issues/cases for further reducing non-clinical and clinical data. The progress was reported to the committee in 2020 (72nd and 73rd report) [52]. It has resulted in additional suggestions on evaluating biosimilar monoclonal antibodies (mAbs) and an expanded Q&A document.

In April 2022, the WHO published [53] a revised guideline based on the 22 comments received. While the newest guideline and suggestions made by the WHO represent the views of global regulatory agencies, it still falls short of establishing a rational scientific platform. Listed below are some of the notable shortcomings in the WHO understanding that should not be made part of a harmonized guideline:

- The WHO states that “the clinical data should be generated using the biosimilar product derived from the final manufacturing process, reflecting the product for which authorization is being sought. Any deviation from this recommendation must be justified, and additional data may be required. For changes in the manufacturing process, relevant guidelines like the ICHQ5E should be followed”. However, the ICH comparability guideline applies only to the changes in the manufacturing of a biotechnology product that has already been approved and thus requires testing the product before and after the change, not with the reference product. To avoid confusion, the FDA has made a strong point by labeling these studies as “analytical assessments,” not even analytical comparisons.
- In its earlier guidelines, the WHO had indicated no need for any statistical modeling of the critical quality attribute comparisons. The recent draft suggests using statistical modeling but warns about the risks of employing statistical tests on limited samples
(false-positive and false-negative conclusions). This reluctance of the WHO to propose solid statistical modeling has resulted in many agencies requiring only 3–4 lots [54] for testing. It is well understood that a larger number of lots are required before the statistical modeling can be initiated. The WHO also states that the most frequently applied overall similarity criteria require that a certain percentage of the biosimilar batches (usually between 90% and 100%) fall within the similarity range. Given that in an equivalence range, 90% of biosimilar lots must fall within three standard deviations for the reference product. This means that only one lot out of ten can fall outside the range, but if there are less than ten lots tested, the analysis becomes moot.

- For efficacy studies, the WHO allows using a non-inferiority model discouraged by the FDA and EMA as inappropriate to consider higher efficacy leading to higher safety issues.
- The WHO suggests that the chosen reference product must have been marketed for a “suitable period” with proven quality, safety, and efficacy to serve the reference product. No suitable period is defined, and the advice has led to distrust in the safety of biological drugs approved under stringent regulatory compliance. While there is a 12-year restriction in the US and ten years in the EU, the WHO member agencies do not have to comply with this restriction. The WHO statement has caused great damage to the adoption of biosimilars in developing countries, and it must be removed.
- The WHO suggests that a biosimilar developer may use one source of reference product for analytical testing and another for clinical testing. This argument is illogical; all testing should be performed using the same reference product derived from the same manufacturing source and bearing the same approval designation.
- The WHO maintains its position, despite many criticisms, that regional agencies can decide the labeling and prescribing information. This is not only improper, but it is also unethical, giving the regulatory agencies to modify the safety and efficacy disclosures. The FDA has provided details of how the prescribing information should be developed; this should be followed by the WHO.
- According to the WHO, if a comparison reveals differences in product-related substances and impurities between the biosimilar and the reference product, the impact of the differences on the clinical performance of the drug product (including its biological activity) should be evaluated. This is the most misguiding advice. A reference product had been thoroughly tested with its impurity profile for safety and efficacy. Unmatched impurities cannot be validated by any means, including animal testing or clinical efficacy testing. The quantity of matched impurities can vary, within certain limits, as they can only bring a change in efficacy that is not likely to be significant. This argument extends to process-related impurities as well. The process-related impurities can be adjusted; thus, there is little rationale for qualifying an impurity not present in the reference product. If a biosimilar product production can remove these uncertainties, it should, regardless of their assumed risk.
- In the past, the WHO had given little importance to accelerated or stress condition testing; this is now changed to follow the rationale that these testing are meant to be part of the analytical assessment.
- The WHO statement, “It is up to the manufacturer to justify the relevance of the established similarity ranges and criteria”, is inappropriate. These determinations should be based on scientific principles, not individual agency preferences. This advice from the WHO has resulted in the NRAs adopting irrational test limits without justification.
- The WHO statement, “Nevertheless, any quality attributes not fulfilling the established similarity criteria should be considered a potential signal for non-similarity and assessed for possible impact on clinical safety and efficacy”, invites developers to seek waivers based on animal or clinical testing. Here, the WHO goes back to the assumption that differences in analytical similarity can be justified through any non-clinical or clinical study.
WHO states, “Based on the totality of quality and nonclinical in vitro data available and the extent to which there is residual uncertainty about the similarity of a biosimilar and its reference product, it is at the discretion of the involved NRA to waive or not to waive a requirement for additional nonclinical in vivo animal studies”. This statement is misleading as it has caused many agencies to develop extensive animal testing, such as the Indian CDSCO [54], which suggests using several times the human dose to establish safety. The WHO further states, “To address the residual uncertainties, the use of conventional animal species and specific animal models (for example, transgenic animals or transplant models) may be considered”. This is not sound scientific advice, leaving an impression that it may be possible to resolve differences in analytical similarity using tests without relevance.

WHO suggests that local tolerance studies are not required unless excipients are introduced for which there is little or no experience with the intended clinical route of application. Biosimilars can have formulations different from the reference product, and a tolerance study is required to evaluate the formulation. If a formulation includes ingredients that have not been used before, this creates significant risk and cannot be resolved.

According to the WHO, “Clinical studies should be designed to demonstrate confirmative evidence of the similar clinical performance of the biosimilar and the reference product, and therefore need to use testing strategies that are sufficiently sensitive to detect any clinically relevant differences between the products”. The testing strategies are always the same, either a response on a clinical marker. Both are the least sensitive to tell the difference compared to analytical assessment and clinical pharmacology testing.

Using reference products remains unclear with issues such as using a foreign reference product instead of a domestic product if a suitable reference product is not licensed locally. In this case, the NRA may accept a reference product that has been licensed in another jurisdiction.

If required by the legislation in place, the comparability of the local and foreign-sourced versions of the product should be demonstrated by analytical “bridging” studies and, where needed, complemented by additional PK/PD data. Here the WHO allows precedence of any local regulations to overcome scientific arguments.

The WHO statement, “It may also be prudent not to waive the efficacy and safety study when the reference product has common or unpredictable serious adverse effects that cannot be merely explained by exaggerated pharmacological action”, is based on the wrong assumption that efficacy study can overcome any unusual effects.

The WHO also allows the NRAs to develop their prescribing information, which leads to abuse of biosimilar products. There must be a unified approach to creating the label.

2.4. The MHRA Scene

Since exiting from Brexit, the MHRA has revised its guidelines independent of the EMA or ICH guidelines. As the end of the Brexit transition period approached last year, the MHRA released draft guidance for consultation that was finalized on 14 May 2022 [38]. This guideline has taken a more clear and more definite approach to the issues of animal testing and clinical efficacy evaluation.

2.5. The ROW Scene

Most countries follow the guidelines discussed above; however, the WHO members are more inclined to follow them unless they are more affluent, such as Saudi Arabia, where the FDA/EMA guidelines apply. Many countries treat biosimilars like generic chemical products with no clinical testing. In most cases, clinical testing is suggested for a fixed number of patients. Recently, the concept of biological API (active pharmaceutical ingredient) has risen, where companies import the drug substance and finish it locally.
3. Definitions

3.1. Terminology

The first step in harmonizing the regulatory guidelines requires uniformity in the terminology used in the context of biosimilar products. The terminology used in describing biosimilars and the testing requirements can make a difference in the scope of testing. For example, the term “biogeneric” was coined in 2004, but it was refused by the FDA on legal grounds adopting “biosimilars” instead [55]. The term “biosimilar” means biologically similar products, not otherwise, as it would be concluded if we call them similar biologicals. Table 1 shows the current designations.

Table 1. Designation of off-patent biologics across the globe as of 2022.

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biosimilars</td>
<td>WHO, Canada, China, Egypt, Ghana, Indonesia, Iran, Jordan, Malaysia, Korea, Singapore, Thailand, USA, and Zambia</td>
</tr>
<tr>
<td>Follow-on Biologics</td>
<td>Japan, Brazil</td>
</tr>
<tr>
<td>Similar Biologies</td>
<td>India, Peru</td>
</tr>
<tr>
<td>Similar Biologic Medicinal Product (SBMP)</td>
<td>EU, Ukraine</td>
</tr>
<tr>
<td>Multisource Known Biological Products</td>
<td>Cuba</td>
</tr>
<tr>
<td>Bioanalogue</td>
<td>Russia</td>
</tr>
</tbody>
</table>

Another term that needs attention is “comparability”, frequently used in place of “similarity” when comparing a biosimilar candidate’s structural or biological attributes with its reference product. The confusion starts with the ICH Guideline ICHQ5E [56] “Comparability of Biotechnological/Biological Products”, which is intended to qualify the changes in the manufacturing process of an approved biotechnology product, where the “comparability” is established between the products and before and after the change, not a reference product. Statements such as “for changes in the manufacturing process, relevant guidelines like the ICHQ5E should be followed” [52] by the WHO are misleading. It has caused developers to compare the development lots with their commercial product lots. The similarity assessment can only be made with the reference product; the “comparability” comes into the picture after a product has been approved. The FDA has made it more explicit by replacing “testing” with “assessment”.

3.2. Reference Product

What constitutes a reference product should be uniform—a product approved using a full dossier in one of the developed country’s regulatory agencies. This concept is not new; when generic drugs came into existence, the WHO concluded that “Comparator products should be purchased from a well-regulated market within the jurisdiction of a stringent regulatory authority (SRA). For WHO Medicines prequalification, an SRA is considered the regulatory authority of a country officially participating in the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and a member of ICH before 23 October 2015, namely: the US Food and Drug Administration, the European Commission and the Ministry of Health, Labor and Welfare of Japan also represented by the Pharmaceuticals and Medical Devices Agency. Therefore, the consensus that the reference biological product must be approved by one of the initial members of the ICH should be acceptable.

The WHO statement, “if the reference product is not authorized locally, the NRA may allow the use of a product licensed by an experienced NRA that follows the WHO or corresponding regulatory standards”, creates disharmony. The WHO further suggests bridging studies if the reference product is licensed locally but sourced from another jurisdiction.
There is a legal glitch in the BPCIA (FDA) that requires a reference product to be "licensed", and biologicals are "licensed" only in the US. There are many cases where the same product is licensed in the US and authorized in the EU wherein “essentially” the same registration dossier is submitted for registration. The EU will not require a bridging study; the FDA has recently required a PK bridging study in such situations. The bridging studies, especially clinical PK/PD studies, have been criticized since they complicate the global development of biosimilars. The bridging studies should not be conducted if the reference comparator has been approved in any ICH jurisdiction and there is evidence in the public domain that the reference product has been approved in both jurisdictions upon some of the same phase III clinical data” [57]. It is doubtful that the EU- and US-sourced reference products have meaningful differences [58].

Health Canada permits using a foreign-sourced reference product licensed in an ICH country when it does not have a locally sourced product [59].

A foreign-sourced reference product can be used in clinical studies. This is the case for the EU and US, where a biosimilar must always refer to a local reference product for legal reasons. Still, clinical studies can be performed with a non-European Economic Area (EEA)/non-US version of the reference product, provided this has been authorized by a regulatory authority with similar scientific and regulatory standards. In this case, the FDA guidelines require analytical and PK/PD “bridging” studies by default. In contrast, the EMA and Health Canada guidelines require analytical bridging, but PK/PD bridging only if analytical bridging alone is insufficient.

The MHRA UK states that the reference product is sourced from the EU with evidence that the RP is licensed in the EU via the centralized, decentralized, or mutual recognition procedures, providing confirmation that these are the same as the Great Britain RP [54].

3.3. Materials and Standards

In-house Primary Reference Material: An appropriately characterized material prepared by the manufacturer from a representative lot(s) for biological assay and physicochemical testing of subsequent lots and against which in-house working reference material is calibrated.

In-house Working Reference Material: A material prepared similarly to the primary reference material established solely to assess and control subsequent lots for the individual attribute in question. It is always calibrated against the in-house primary reference material.

Publicly available reference standards (e.g., Ph. Eur.) cannot be used as the reference product for the demonstration of biosimilarity. However, using these standards is vital in method qualification and standardization.

4. Expression System

The Current US and EU guidelines allow the use of any expression system; however, if the expression system is not the same, at least to the extent it is described in the prescribing information (type, not necessarily the subtype), there can be many systematic issues with product quality that will require additional testing. In addition, in some instances, any "residual uncertainty" cannot be removed. The most widely used non-human host cell lines for recombinant expression are CHO, NS0, Sp2/0, HEK293, and PER.C6, BHK21, E. coli, and S. cerevisiae.

The developers are also advised to select an expression system that is more steady than productive; recently, very high-yielding cell lines have been developed, but when cell systems are pushed to produce, they also end up producing variants. Since the cost of goods of recombinant products is based on the carbon input, it may not significantly alter the cost. For example, the WHO calculates that the cost of production of monoclonal antibodies ranges from USD 95–200 per gram [60] given the current market price of 100–1000 × the COGS; developers are advised to base their selection on cell lines that will allow for faster approval if they produce a consistent product. Furthermore, important are the
considerations in the contamination of cell lines with antibiotics that should be avoided in their design.

5. Formulation

Biosimilars are allowed to have a formulation different from the reference product formulation. However, unless prevented by intellectual property, a formulation with the same or fewer inactive ingredients is preferred, notwithstanding any minor differences in the composition of ingredients. If a different formulation is used, it may contain new excipients that may not have been tested for the specific drug substance. If it is not used for the formulation of biological products of the same classification; all excipients must be free of animal materials; The formulation of the biosimilar should be selected considering state-of-the-art technology and does not need to be identical to that of the reference product. Regardless of the formulation selected, the suitability of the proposed formulation with regards to stability, compatibility (i.e., interaction with excipients, diluents, and packaging materials), integrity, activity, and strength of the active substance should be demonstrated. Suppose a different formulation and container/closure system to the reference product is selected (including any material in contact with the medicinal product). In that case, its potential impact on the efficacy and safety of the biosimilar should be appropriately justified.

Biosimilar epoetin α was the first epoetin α product that demonstrated the risk of neutralizing antibodies cross-reacting with endogenous erythropoietin, which has caused pure red cell aplasia in patients treated with the reference product [61]. This led to the discontinuation of the development of the product for subcutaneous administration until the underlying problem (which was not related to the quality of the active substance itself but tungsten leaching from the needle of the syringe) was eliminated. Thus, the licensed biosimilar epoetin α products have not shown excess immunogenicity compared to the reference product [62–65].

A novel excipient not used in the recombinant protein formulations should be avoided. The use of excipient(s) in the proposed biosimilar product not used in the RP is not encouraged from a biosimilarity perspective. However, changes that may benefit patients (for example, reducing injection pain or stinging) are encouraged and should be carefully considered. Where different excipient(s) are used, there could be instances where this would be the first time that this route had used the excipient; a discussion should be presented addressing the safety of that excipient by the route intended.

In most instances, the excipient(s) will be used by the route intended at similar amounts to other products. If so, a discussion to establish this can be sufficient. However, suppose a novel excipient or a novel route for an excipient is used in the proposed biosimilar product. In that case, this should be justified and includes the possibility that results from new safety studies are presented, if appropriate. As studies intended to characterize the safety of the excipient, compliance with GLP is expected.

6. Release Specification

The first step in developing a biosimilar product is establishing the release specification drawn from the analysis of the reference product. Characterization of a biological product includes the determination of its physicochemical properties, biological activity, immunochemical properties, purity, and impurities using qualified testing methods. Testing a larger number of reference product lots is favorable to biosimilar developers, as it enables the justification of ranges of specifications that are more rational. The test lots can come from the lots used throughout the development process. However, at least one lot tested must be the one used for the first clinical trial, the PK/PD study.

The manufacturing process of the reference product evolves through its lifecycle, which may lead to detectable differences in some quality attributes. Such events could occur during the development of a biosimilar product. They may result in development according to a QTPP, which is no longer fully representative of the reference product
available on the market. The ranges identified before and after the observed shift in quality profile could normally be used to support the biosimilar comparability exercise at the quality level, as either range represents the reference product. Quality attribute values outside or between the range(s) determined for a quality attribute of the reference product should be appropriately justified concerning their potential impact on safety and efficacy.

Many legacy attributes are independently established, such as sterility, invisible particles (a controversial issue with biosimilars to consider them as aggregates), protein content, potency, and physical properties specific to the biosimilar product; however, these remain controversial. For example, the commonly acceptable for having no more than 3% impurity and no single impurity more than 1% should be acceptable unless these ranges are higher, in which case, they must be justified. In addition, the impurities must not include any impurity not found in the reference product. Any concessions in this regard are the remnants of the understanding of the chemical products, where the immunological consideration is unimportant. Therefore, attempts to justify the safety of unmatched impurity are futile; it is better to remove the unmatched impurity.

7. Analytical Assessment

Analytical assessment is the strongest element of establishing biosimilarity. With newer analytical technologies, it is now a more vigorous exercise. Though the critical quality attributes are well-established, and so are the tests necessary, the developers have shown great discord in the choice of tests. For example, companies have submitted different numbers of analytical studies for adalimumab—25 by Pfizer and 71 by Boehringer—to achieve the same goal [66].

There is a disconnect between what constitutes orthogonal testing and what is duplicate testing. An orthogonal test is required if a validated or suitable test can provide aberrant results. For example, an HPLC method to measure protein content can be an orthogonal test to UV absorbance testing, but not another spectrometric test or another HPLC method. The validated methods are required for release testing but not for analytical assessment, and it is for this reason that side-by-side testing is needed and all testing at the same time.

The burden of analytical assessment can be significantly reduced if it is limited to quality parameters other than those included in the release specification. Every analytical assessment report of approved biosimilars uses release specification parameters in analytical assessment. For example, protein content or potency tests are release specification attributes.

The WHO does not consider a need for comparative stability profiling of the biosimilar candidate with the reference product. Because of differences in formulation, a biosimilar will have its lifecycle. Any process change post-approval should follow the ICH Q5E guideline. However, the stability of the biosimilar product should be determined according to ICH Q5C. The applicant should demonstrate that the desired product (including product-related substances) present in the finished product of the biosimilar is similar to that of the reference product. In contrast, process-related impurities may differ between the originator and biosimilar products, although these should be minimized. It is preferable to rely on purification processes to remove these impurities.

Product-related attributes are generally not modified by changing such parameters as upstream conditions; these are mostly driven by the expression system. Product-related attributes can and should be optimized; for example, one way to remove an unmatched impurity is to lose the yield and cut off the peak instead of justifying the impurity. The EU and UK fully agree with this suggestion [38].

The process-related attributes are not tested for analytical assessment; they are part of the release specification established by testing multiple lots of the reference product. However, any legacy attribute such as protein content or potency need not be tested in both instances.

Testing requires reference materials, and there have been many misconceptions about the role of pharmacopeias. Product release is based on using in-house reference materials,
not on standards and reference materials (e.g., from Ph. Eur., WHO) that can only be used for method qualification and standardization.

An interesting example of disputes relates to the release of insulin products. The United States Pharmacopeia (USP) mandates an animal-based assay in rabbits in its Chapter “<121> Insulin Assays” (USP <121>) for the potency evaluation (biological activity) of insulin and insulin analogs. As the biodentity test is mandatory in the US, it is included in the quality specification for insulin drug substances for the US market. Since physicochemical assays such as HPLC assays used to determine the content of human insulin and insulin analogs are much more precise and accurate than the rabbit blood sugar test, most of the Pharmacopeias (e.g., in Europe, Japan, and India) decided to forgo the testing in living animals. Consequently, the EMA recommends that marketing authorization holders use the chromatography method for insulin, while the FDA insists on the rabbit test [67,68].

The pharmacopeias general monographs include tests for sterility, endotoxins, microbial limits, volume in the container, uniformity of dosage units, and acceptable particulate matter. However, if provided in a monograph, the specification is not acceptable to FDA and EMA; a side-by-testing must be established.

The EMA provides more comprehensive guidance divided into immunogenicity testing, quality issues, clinical and non-clinical testing, pharmacokinetic modeling, and guidance on changing the manufacturing process of recombinant drugs. In addition, the product-specific guidelines of the EMA are of great value for biosimilar developers [69].

The European and British Pharmacopoeias have developed monographs of several critical biological products defining quality attributes to establish release specifications. The USP has stated that it will not develop monographs for a biologic unless there is stakeholder consensus supporting its creation, including the support of the FDA [70]. The FDA has discouraged the USP from creating biologics monographs to ensure that innovator biologics makers do not use the monograph process to block biosimilar competition by incorporating patented characteristics of their product that are not relevant to safety, purity, or potency, thereby further impacting competition [71]. The FDA also stayed away from creating product-specific monographs, unlike the EMA.

**Statistical Modeling**

Comparing quality attributes is key in evaluating biosimilars and manufacturing process changes. Different statistical approaches are required, but there is no regulatory consensus on a quantitative and scientifically justified definition and an underlying hypothesis of statistically equivalent quality. Therefore, the comparisons must be made using methods to calculate the operating characteristics for false acceptance and rejection rates of a claim for statistically equivalent quality. These error rates should be as low as possible to allow a meaningful application of a statistical approach in regulatory decision-making.

Statistical data modeling, whenever comparative testing is conducted, is highly controversial. An earlier FDA guideline, “Statistical Approaches to Evaluate Analytical Similarity”, which recommended a rigorous statistical approach for establishing similarity, was withdrawn [72] and replaced with a new guideline [73] in response to many objections, including a citizen petition [74]. The new guideline removed the controversial tier one assessment of quality attributes.

Historically, the WHO had maintained that there is no need for any statistical exercises to compare the data; the most recent WHO guideline states: To mitigate the risks inherent in employing statistical tests on limited samples (false-positive and false-negative conclusions), a comprehensive control strategy must be established for the biosimilar to ensure consistent manufacturing” [51,52]. However, while the guideline supports the quality range approach, it fails to suggest a minimum number of lots needed, as does the FDA guideline [75,76].

The EMA, which had been silent on statistical methods, has described these in detail in its newest guideline that describes the critical approaches for testing biosimilars [76]. While it recommends the interval range approach, it fails to mention the number of lots required and leaves it up to the developer. A statistical testing requires a minimum number
of lots to be of any value. Only the FDA guideline suggests using at least 6–10 lots, apart from offering to conform to the suitability of the number of lots required.

A basic understanding of data teaches us that the application of any model is based on the nature of data; if it is normally distributed, then many tests can be applied. The first consideration when applying statistical tests is the essential flexibility of the requirement of “high similarity” to allow for differences if they are clinically meaningless. Statistics may facilitate the detection of differences, e.g., in data distributions or ranges. Still, the determination of whether these differences are clinically relevant is a scientific question that cannot be addressed by a statistical approach alone. Here is a list of various approaches to compare the two products used [76].

- **Visual display.** This is most suitable where spectra are produced; most important is the peak locations. This is one of the most important tests as it applies to the critical comparison of primary, secondary, and tertiary structures.
- **MinMax:** A MinMax range is defined by a sample’s lowest and highest values. The MinMax test is accepted if the MinMax range of the test sample is within the MinMax range of the reference sample (minTest > minRef and maxTest < maxRef). The MinMax is a conservative approach with a low false acceptance rate but a high false rejection rate.
- **3Sigma:** the 3Sigma range is calculated for the reference sample as $(\mu_{\text{ref}} - 3\sigma_{\text{ref}}, \mu_{\text{ref}} + 3\sigma_{\text{ref}})$. The 3Sigma test is accepted if the MinMax range of the test sample is within the 3Sigma range. The 3Sigma approach provides a more practical compromise of error rates, further improving with a larger sample size.
- **Tolerance interval (TI):** The tolerance interval is calculated for the reference sample as $(\mu - k*\sigma_{\text{ref}}, \mu + k*\sigma_{\text{ref}})$. The k-factor is calculated two-sided with a confidence level of 0.9 and a proportion of the population covered by the tolerance interval of $p = 0.99$. The tolerance interval test is accepted if the MinMax range of the test sample is within the tolerance interval calculated for the reference sample. Tolerance interval testing is only usable if the sample size is sufficiently large.
- **Equivalence testing of means (EQT):** A two one-sided t-tests’ (TOST) procedure is used to test for equivalency of the means of the reference product and the test product. The equivalence margin is $\delta = 1.5 SD_{\text{ref}}$ (standard deviation of the reference product sample). The Type I error probability is controlled at level $\alpha = 0.05$ for a conclusion of equivalence. The test is accepted if the $1 - 2\alpha = 90\%$ confidence interval for the difference in the means is within $(-\delta, +\delta)$. Equivalence testing has a high false rejection rate and, with increasing sample size, a considerable false acceptance rate.

The method of defining the acceptance ranges of critical quality attributes is well described in EMA and FDA quality guidelines. In particular, the FDA biosimilar guidelines thoroughly explain the risk assessment of quality attributes, while EMA guidelines refer to other guiding documents. In general, both FDA, Health Canada, and EMA highlight the importance of state-of-the-art orthogonal analytical methods and in vitro functional/potency tests in the characterization of biosimilars. The quality section of the WHO guidelines needs to be updated to align with the current expectations for analytical characterization and demonstration of biosimilarity.

The quality attributes where statistical modeling is applicable include purity profile, aggregate profile, and function assay profile; glycosylation is better compared with equivalence test; the equivalent testing of mean is least likely to be of any value as used in every regulatory submission.

While the purpose of analytical assessment is to show the difference, it is highly unlikely that a significant difference can be justified; the test limits proposed in all the above tests are arbitrary; it is for this reason, analytical differences are not allowed since, in some cases, a minor change can produce an adverse response. The analytical assessment extends to clinical pharmacology testing, discussed below.
8. Non-Clinical Testing

Non-clinical in vitro testing has brought many new possibilities for characterization, and these are considered part of the analytical assessment.

Non-clinical in vivo studies remain unresolved; the FDA BPCIA states, “demonstrating similarity based on evidence from pre-clinical studies (including toxicity assessment) is required”. In several submissions, the FDA has refused to consider multiple animal pharmacology and toxicology studies; however, the FDA remains mute on waivers of these studies, even though the FDA has suggested that animal testing is not needed for new biological drugs [9]. One reason for this dichotomy is the language used in the legislation describing how biosimilars should be approved.

According to EMA, “Non-clinical studies should be performed stepwise before initiating clinical studies. If necessary, in vitro studies should be conducted first, followed by in vivo. These studies should be designed to detect differences between the similar biological product and the reference product” [77]. However, the EMA has recently supported the in vivo non-clinical testing waiver. It is up to the developers to seek the waiver. However, what is needed is a ban on these useless studies a suggested by the author [9,78].

The WHO states, “Non-clinical evaluation of a new biotherapeutic normally encompasses a broad spectrum of PD, PK and toxicological studies to validate the efficacy and safety of an SBP and should be comparative. Non-clinical studies should be performed in a stepwise manner before initiating clinical studies. If necessary, in vitro studies should be conducted first, followed by in vivo. These studies should be designed to detect differences between the similar biological product and the reference product”.

In deciding the rationale for non-clinical testing, the WHO states, “Based on the totality of quality and nonclinical in vitro data available and the extent to which there is residual uncertainty about the similarity of a biosimilar and its RP, it is at the discretion of the involved NRA to waive or not to waive a requirement for additional nonclinical in vivo animal studies”. The WHO may consider using conventional animal species and specific animal models (for example, transgenic animals or transplant models) to address the residual uncertainties.

The EMA guideline provides the concept of step-wise progression in non-clinical testing and reduction of animal work following the 3R principles according to Directive 2010/63/EU. Health Canada states that in vivo toxicological studies are generally not needed. It is acknowledged that biotechnology-derived proteins may mediate in vivo effects that cannot be fully elucidated by in vitro studies. Therefore, non-clinical evaluation of in vivo studies may be necessary to provide complementary information, provided that a relevant in vivo model about species or design is available [79].

According to the revised UK guideline: “No in vivo studies from animals are requested as these are not relevant for showing comparability between a biosimilar candidate and its RP: this includes pharmacodynamic studies, kinetic studies, and toxicity studies” [80].

According to the EU and UK, factors to be considered when the need for in vivo non-clinical studies is evaluated include, but are not restricted to:

- Presence of potentially relevant quality attributes that have not been detected in the reference product (e.g., new post-translational modification structures).
- Presence of potentially relevant quantitative differences in quality attributes between the biosimilar and the reference product.
- Relevant differences in formulation, e.g., excipients not widely used for biotechnology-derived proteins.

If product-inherent factors that impact PK and biodistribution, such as extensive glycosylation, cannot sufficiently be characterized on quality and in vitro level, in vivo studies may be necessary. If there is a need for additional in vivo information, the availability of a relevant animal species or other relevant models (e.g., transgenic animals, transplant models) should be considered. If a relevant in vivo animal model is not available, the applicant may choose to proceed to human studies considering principles to mitigate any potential risk.
Testing in animals is an old routine for new drugs to avoid toxicity to humans. It works well for chemical drugs because the reactive chemical groups can interact with multiple tissues to produce a toxic response. However, biological drugs may not always show a pharmacologic response in animal species; thus, the toxicity is an extension of the pharmacological response for biological drugs. The primary mechanism of action of biological drugs involves receptor binding. A pharmacological or toxicological response is not expected if an animal species does not carry these receptors.

Another reason animal toxicology data are less relevant is how the testing is conducted. Generally, animal testing protocols require administering a higher dose to induce a toxic response; however, within this dose range, the responses are not expected to be linear, making it impossible to differentiate between compared products that are supposed to be the same. Nevertheless, despite this knowledge and expertise, animal testing is extensively conducted for biosimilars, evidenced by the recent FDA and EMA filings.

Another controversial issue in animal studies is the use of non-human primates, the only species that may have relevant receptors; it is frequently recommended to conduct PK studies in a small number of animals, especially for monoclonal antibodies, as a measure of their molecular structure rather than toxicity. According to the recent statements from WHO [54], “based on regulatory experience gained to date in marketing authorization applications for biosimilars, the need for additional in vivo animal studies would be expected to represent a rare scenario”. However, the guidelines in India take a very different view, stating, “Regarding the animal models to be used, the applicant should provide the scientific justification for the choice of animal model(s) based on the data available in scientific literature. However, toxicity studies need to be undertaken in rodent or non-rodent species if the pharmacologically relevant animal species are unavailable and appropriately justified”. This requirement was implemented because India requires at least one animal toxicology study, and no studies are allowed on monkeys for religious reasons.

Human and animal cells, organoids, organs-on-chips, and in silico modeling are alternatives to animal testing models, enabling us to create better and more predictive scientific methods. In addition, to reflect changes in animal protection legislation, non-clinical in vivo testing has been substituted by in vitro assays in the previous ten years. These measures can help reduce animal use. They also align with the EMA’s Regulatory Science Strategy for 2025, aiming to create a more adaptive regulatory framework that promotes human and veterinary health [80].

Animal toxicological studies can be misleading if they rationalize discrepancies in impurities, post-translational modifications, or antibody responses since an animal model can justify these differences. For example, animal data were submitted in biosimilar applications [66] to substantiate such variability, but the FDA refused to accept the animal data.

The EMA and FDA have approved more than 130 products. None of them have failed animal toxicological testing because they cannot, being least sensitive in detecting any difference between a biosimilar candidate and its reference product. These observations and conclusions are widely accepted as scientifically sound arguments, but among sponsors, there is always fear that study results will be rejected eventually. This would cause a delay in market access at a high cost, and therefore sponsors like to stay on the safe side by overpowering their studies.

There is great awareness of the futility of the animal testing of biosimilars, but this will soon become a moot point, as the US Senate is considering a bill to remove the animal testing of biosimilars [81]. Section (bb) is amended from “(bb) animal studies (including the assessment of toxicity)” to “an assessment of toxicity (which may rely on, or consist of, a study or studies described in item (aa) or (cc)); (aa) is analytical assessment and (cc) is clinical testing. This bill sponsored by Senator Lujan of New Mexico is now on the table in Senate and expected to be signed soon [81].

Other agencies will follow suit.
9. Clinical Evaluation

The need for clinical efficacy, as described in the FDA guidance, has brought much confusion and misunderstanding [44]:

“As a scientific matter, FDA expects a sponsor to conduct comparative human PK and PD studies (if there is a relevant PD measure(s)) and a clinical immunogenicity assessment. In certain cases, the results of these studies may provide adequate clinical data to support the conclusion that there are no clinically meaningful differences between the proposed biosimilar product and the reference product. However, if residual uncertainty about biosimilarity remains after conducting these studies, an additional comparative clinical study or studies would be needed to evaluate whether there are clinically meaningful differences between the two products.”

First, “clinical” does not mean testing in patients, and it could well be an in silico pharmacokinetic study, as stated in the FDA’s Biosimilars Action Plan [82]. However, without defining the “residual uncertainty”, without clarifying what is meant by “clinical”, the regulatory agencies and the developers have assumed it to mean clinical efficacy testing as a requirement; hundreds of such studies have been conducted, and none of these studies failed, just like the animal toxicology testing, they cannot fail.

10. Clinical Pharmacology

It is now well-established that demonstrating comparable pharmacokinetics is critical in the successful development and approval of most biosimilars [83]. The US, Canadian, and EU guidelines require comparative PK and, if relevant, PD studies between the candidate biosimilar and its reference product by default [84]. The current WHO guidelines also recommend this approach because PK(PD) studies are sensitive to detecting potential product-related differences in vivo. Scientific reports underline the importance of well-performed and robust PK studies with proper power calculation avoiding too optimistic calculation of the inter-individual variability. Such PK studies may already provide sufficient data on safety, including immunogenicity. In any case, EMA, Health Canada, and FDA guidelines request safety and immunogenicity data for all clinical studies [85,86].

The FDA states, “Comparative human PK, PD, clinical immunogenicity, clinical safety, and effectiveness are required. In addition, at least one study with equivalence design and adequate power is required to evaluate any clinically meaningful difference between the proposed biosimilar product and the reference product”.

The EMA states, “PK/PD studies, followed by clinical efficacy and safety studies, confirmatory studies for demonstrating clinical biosimilar comparability are required. In addition, at least one study with adequate power and equivalence design in a sensitive population is required to detect potential differences concerning efficacy and safety”.

Clinical pharmacology comparisons comprise the most relevant testing to support biosimilarity. First, however, misconceptions should be addressed; unlike the development of new drugs, here the purpose is to compare [87,88], which means that limited inclusion criteria can reduce the study size. For example, Sandoz’s first study, GP17-101, failed to demonstrate PK biosimilarity between the GP2017 test molecule, the EU-sourced Humira, and the EU- and US-sourced RPs. Therefore, the company reconsidered the study design and conducted another trial, GP17-104. The revised design increased the sample size and was restricted to male patients, randomized, and stratified by body weight. GP17-104 was successful in demonstrating PK profile similarity [89].

The FDA Biosimilar Action Plan also recommends employing in silico methodologies to compare biosimilars, including immunogenicity assessments. Since immunogenicity is entirely structure-dependent, better analytical assessment techniques give greater confidence in reducing or eliminating anti-drug antibody testing. In addition, impurities and
aggregates induce extrinsic immunogenicity, which may be easily measured and compared to a reference product as part of the analytical evaluation.

11. Clinical Safety and Efficacy Studies

The issue of clinical efficacy testing becomes more compliance-based in the US, where the legislation has created two classes of biosimilars, one that is biosimilar with “clinically meaningful difference with the reference product”, and another that is further tested in multiple switching and alternating studies with the reference product, to receive the designation of the interchangeable biosimilar. Thus, the FDA has approved two interchangeable products—insulin glargine and adalimumab. Hundreds of clinical studies on switching and alternating have shown that there is no difference [90]. Decades of switching biological drugs such as insulin have shown no reason to assume that switching will result in a different response or safety issue. Moreover, alternating makes little sense, as there is no rationale for returning to the switched product.

Apart from the waste of resources, the more significant loss is that this legislation creates a lack of confidence in the biosimilar that is not interchangeable testing. This difference will be used by larger companies, as many of the products not part of the substitution program are conducting these studies. However, removing this category of biosimilars will take a legislative action currently underway, but like other such changes, it will take time.

FDA, Health Canada, and EMA guidelines provide some flexibility regarding the phase III-type “confirmatory” clinical efficacy and safety studies if certain requirements are met, especially the availability of PD markers that are relevant markers or even surrogates for efficacy. The Health Canada guideline states, “The non-clinical and clinical programs should be designed to complement the structural and functional studies and address potential areas of residual uncertainty”. The FDA guideline presents points that must be addressed if the confirmatory efficacy and safety study is considered dispensable.

However, for most products, especially biosimilar mAbs, resource-intensive, phase III-type confirmatory studies with an equivalence design are still expected, a major hurdle in the entry of biosimilars.

The US guidelines advise confirming efficacy and safety studies to be performed if residual uncertainties remain after the previous development steps. However, this advice is not rational. No animal or efficacy testing can resolve an issue at the analytical assessment stage since no data are available to correlate the analytical differences with safety and efficacy; moreover, the efficacy studies are the least sensitive. It is this perception that needs to be removed. Justifying differences in the analytical properties can only lead to the entry of more hazardous biosimilars. The dose responses for biological drugs are broad and unknown; as a result, none of these studies have ever failed. Clinical testing requires establishing a pre-specified margin of difference that is always arbitrary; this includes bioequivalence testing and clinical efficacy responses, making all these studies much less sensitive to identifying differences between a biosimilar product and its reference product. Hundreds of clinical efficacy testing conducted show that none has failed.

The MHRA states: “although each biosimilar development needs to be evaluated on a case-by-case basis, it is considered that, in most cases, a comparative efficacy trial may not be necessary if sound scientific rationale supports this approach. Therefore, a well-argued justification for the absence of an efficacy trial should be appended to CTD Module 1 of the submitted application [38].

The immunogenicity of biological products is caused by the activation of B cells, which generate T cells to express antibodies. However, anti-drug antibodies can be harmful to healthy subjects in future studies. As a result, the FDA is researching new methods for determining immunogenic potential using tiny fragments of DNA-like molecules called aptamers to test proteins and establish their exact structures to avoid the exorbitant costs of forecasting which particular portions of such proteins will stimulate antibody production [91].
Finally, if the immunogenicity profile differs but cannot impact the disposition profile, the differences will be meaningless, as the FDA has acknowledged in its new guidance on insulins [92].

The European Medicines Agency (EMA) has begun work on a pilot clinical trial program aiming to advise how to decrease or eliminate clinical testing in biosimilar development [93]. Comparative clinical trials are increasingly seen as sloppy techniques for assessing biological agent similarity. As a result, the testing of biosimilars in patients is more of a checkmark than a meaningful indication.

EMA states, “generally, clinical data aim to address slight differences shown at previous steps and to confirm the comparable clinical performance of the biosimilar and the reference product”. Clinical data cannot be used to justify substantial differences in quality attributes [94]. Therefore, the first argument relates to identifying “slight differences” or, as the FDA labels it, “residual uncertainty”.

Thus far, no biosimilar product has been rejected based on clinical efficacy and safety testing if they passed the rest of the testing. This means the products were biosimilar, or the testing was too insensitive to detect any difference [95–102]. In both cases, this testing becomes irrelevant. This concept of real-time testing is now also questioned by the FDA, which stated that clinical efficacy testing is “broken” and that new digital technologies and real-world evidence (RWE) are required, as outlined in the 21st Century Cure Act [103].

Biosimilars “may be approved based on PK and PD biomarker data without a comparative clinical study with efficacy endpoint(s)”, according to FDA guidance [102]. Using PK and PD biomarker data in healthy participants or patients enables shorter and less expensive clinical investigations and provides more sensitive testing than clinical efficacy with endpoint(s), as demonstrated with filgrastim [103]. The FDA acknowledged this and granted approval for filgrastim-aafi, filgrastim-sndz, pegfilgrastim-jmdb, pegfilgrastim-cbqv, and epoetin alfa-epbx based solely on PD evaluation. Furthermore, the FDA identified the features of PD biomarkers in its advice to assist sponsors in using PD biomarkers as part of biosimilar development programs [104].

Another reason why the clinical efficacy testing of biosimilars can be fallacious is due to the testing models used: equivalence or non-inferiority. In the equivalence testing mode, we first determine the M1 or total efficacy value of the reference product—a highly variable but available parameter; second, we select an acceptable range of difference, the M2, based on a clinical judgment that usually cannot be definitive—at best, it is an arbitrary choice. As a result, since both products are expected to be identical, equivalency studies are least likely to fail. On the other hand, non-inferiority testing is contraindicated because a biosimilar product showing a higher efficacy may also have more safety issues.

The study’s power weakness is a simpler explanation of why the clinical efficacy is less sensitive. Table 2 lists the clinical efficacy testing reported on clinicaltrials.gov that was completed; the study size is generally larger than the originator product used in its placebo-controlled study. Still, the studies have very low power and never fail. It is not too complex to understand this assertion. At a given study power and alpha value, the study size depends on the anticipated difference; that works fine for placebo-controlled studies but not when two products are tested that are supposed to be similar. The only purpose these studies serve is a stereotype proof of concept requirement, even if it is an irrational approach.

Many drugs, including anticancer drugs, require the homogeneity of the study population, which is unlikely. Patients are inevitably exposed to multiple drugs and treatment modalities; additionally, anticancer drugs have a low efficacy rate, further reducing the statistical probability of identifying any difference. Oncology or other terminal illness treatment efficacy studies face specific hurdles, such as enrolling a comparable group of naive patients. Such investigations further fail due to the brief lifespans of patients, which can disrupt the study design.
Table 2. Number of subjects in clinical studies of biosimilars compared to the originator study (www.clinicaltrials.gov (accessed on 1 July 2022)). (References [95–101] report the values for the reference product).

<table>
<thead>
<tr>
<th>Product</th>
<th>Reference Product (N)</th>
<th>Biosimilar Product (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adalimumab (Humira)</td>
<td>70</td>
<td>200–860</td>
</tr>
<tr>
<td>Bevacizumab (Avastin)</td>
<td>307</td>
<td>450–763</td>
</tr>
<tr>
<td>Rituximab (Rituxan)</td>
<td>161</td>
<td>256–629</td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>233</td>
<td>225–875</td>
</tr>
<tr>
<td>Ranibizumab (Lucentis)</td>
<td>37</td>
<td>312–712</td>
</tr>
<tr>
<td>Infliximab (Remicade)</td>
<td>63</td>
<td>199–584</td>
</tr>
<tr>
<td>Aflibercept (Eylea)</td>
<td>180</td>
<td>366–576</td>
</tr>
</tbody>
</table>

Another argument against clinical efficacy testing is the extrapolation of indications allowed for the biosimilar product. If there are doubts about the safety or efficacy, they should be tested in all indications, not just one selected by the developer, even where the modes of action are the same. A good example is conducting a psoriasis study for adalimumab approval instead of testing in psoriatic arthritis.

Several detailed analyses have confirmed that analytical assessments alone or combined with human PK/PD studies can demonstrate comparable efficacy and safety [102–110]. Comparative in vitro functional testing confirms the similarity of active substances’ higher-order structures and functions. PK studies prove similar exposure from the final (formulated) product and provide information on safety and immunogenicity. These are the essential elements of biosimilar development. In contrast, the phase III-type efficacy and safety studies are regarded as “confirmatory” [83,84,111], yet they hardly confirm anything more than what has already been established. More than 130 biosimilars approved in the EU and US show that none of the products was rejected or questioned based on clinical efficacy studies—they all passed.

The observations listed above make the interchangeability status of biosimilars established after multiple switching and alternating between the biosimilar product and the reference product a moot point.

The author has submitted an amendment to the BPCIA, where it removes the sections related to “Interchangeability” and replaces “clinical data” in (aa) with “clinical pharmacology data”.

12. Label

The FDA and EMA provide the detailed structure of prescribing information (label) in biosimilars and agree that all risks associated with the reference product must be stated. However, the WHO leaves it to the discretion of regional agencies [112].

13. Summary of Harmonized Guideline

It is unlikely that regulatory agencies will agree on a single guideline document. Still, they can certainly agree on the basic principles that are scientifically driven to assure the safety and efficacy of biosimilars without any unnecessary or illogical testing.

- A uniform terminology. Products labeled as “biosimilar” are “approved” using “analytical assessment”. “Comparability” confuses the use of ICHQ5E; it is “comparative similarity”.
- The reference product should be the first approved as a new drug in the US, EU, UK, or Japan. The test samples must be secured from the source country, not from any other country where the same product is distributed unless the product is approved using the same dossier; no bridging study is required.
- The expression system type (e.g., mammalian, bacterial, yeast, etc.) used should preferably be the same as used by the reference product, notwithstanding differences in the type of expression species that may not be known. In addition, there should be no antibiotic contamination.
- Must demonstrate that process is controlled and reproducible. Use ICH guidelines.
• All excipients should be free of animal products, and no novel excipients should be added to the formulation.
• Pharmacopeia test methods can be used for validating test methods only, not for comparative testing. The specifications in pharmacopeia should not be used to establish biosimilar specifications.
• Product-specific monographs and the specifications suggested cannot be used to support the claim of biosimilarity.
• If legally possible, the formulation should be the same as the reference product; or one with fewer components.
• It should have the exact mode of action, same concentration, same dose, same route, and same indications.
• Release specifications should be based on testing of reference lots for only non-legacy attributes. Legacy attributes may use compendial specification. All test methods to establish specifications must be validated. Tolerance intervals may be preferred to establish specifications.
• Product-related attributes that are not included in the release specification should be compared using a suitable, not necessarily a valid, method. Compendial test methods can be used if available.
• Analytical assessment of data subject to statistical testing must be derived from at least 9–10 lots of the reference and biosimilar products. The data must be normally distributed. Where applicable, the 3Sigma approach is recommended. The visual comparison is applied wherever the output is a non-quantified graphic display.
• Forced stability testing should be part of the analytical assessment. It should include at least one lot at a commercial scale and manufactured under cGMP compliance that will be used as a clinical lot.
• No unmatched impurities unless a safety profile is established for the same impurity, not the same type. It is preferred to modify the downstream process to remove unmatched impurities.
• Process-related impurities must be fixed with process optimization.
• No animal testing.
• Pharmacokinetic/pharmacodynamic testing in healthy subjects or patients at two dose levels; one at the full dose and the other at half the dose. A parallel-group switching in the second cycle allows for immunogenicity evaluation and cross-reacting antibody evaluation. Use restricted inclusion criteria to reduce study size. A lower dose is likely to be more sensitive in establishing similarity [87].
• No changes in the product are allowed after the PK/PD study has been conducted. ICHQ5E does not apply to biosimilars until they are approved. The clinical lot must be GMP and at scale.
• The label must conform to the reference product label; use the FDA format available. Must include all indications and no new indications [112].
• Post-market safety reporting is required.

14. Conclusions

Regulatory guidelines should be harmonized and be based on the scientific rationale to assure the safety and efficacy of biosimilars. Furthermore, necessary is the avoidance of studies that might lead to justification of differences that may not be proper such as animal and clinical efficacy testing to justify analytical differences. While the three major agencies have revised the guidelines over time as the safety and efficacy of biosimilars have been well-established, many discords remain that can be readily resolved [113–115].

It is suggested that a single ICH guideline title, “Establishing Biosimilarity of Recombinant Products”, will suffice as described above. Regional guidelines will be more challenging to change due to legal and political rather than scientific reasons. Any product approved under this new ICH guideline should be acceptable by all regulatory agencies if it meets all local legal requirements. This will also reduce the burden on the WHO, which
has struggled to create guidelines and failed due mainly to the democratic nature of the process.

While the above recommendation will apply to all jurisdictions, the case of the US is different, as the FDA has to follow the Congressional Act that describes how biosimilars should be approved. The US Congress has brought several bills to expedite the adoption of biosimilars [116], but none have addressed the core reason for the high price of biosimilars. Senator Lujan of New Mexico presented a bill on 20 May 2022 to “allow manufacturers and sponsors of a drug to use alternative testing methods to animal testing to investigate the safety and effectiveness of a biological product”, as recommended by the author [117]. However, it will require removing all mention and references to “interchangeability, and amending the section as shown below (underline is added word):

(k)(2)(i)(I)(cc)
“a clinical pharmacology study or studies (including the assessment of immunogenicity and pharmacokinetics or pharmacodynamics) that are sufficient to demonstrate safety, purity, and potency”.

Delete (k)(2)(iii) (III)(B) and all references to “interchangeability”.

These changes, as submitted to the US Senate by the author, will limit the clinical studies to only pharmacology studies and eliminate the two-class system of biosimilars. A harmonized ICH guideline, and the changes to the US legislation, will turn the destiny of biosimilars, bringing the perspective of affordable biosimilars affordable into reality.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References
46. FDA. Purple Book. Available online: https://purplebooksearch.fda.gov/faqs#5 (accessed on 4 April 2022).
47. FDA BPCA. Available online: https://www.fda.gov/media/779/download (accessed on 12 July 2022).
50. Moorkens, E.; Vulto, A.G.; Huys, I. An overview of patents on therapeutic monoclonal antibodies in Europe: Are they a hurdle to biosimilar market entry? Mabs 2020, 12, 1743517. [CrossRef]
60. WHO. Call for Consultants on Antibodies. Available online: https://www.who.int/news-room/articles-detail/call-for-consultant-on-monoclonal-antibodies-for-infectious-diseases (accessed on 8 June 2022).
76. Stangler, T.; Schiestl, M. Similarity assessment of quality attributes of biological medicines: The calculation of operating characteristics to compare different statistical approaches. AAPS Open 2019, 5, 4. [CrossRef]
83. FDA. Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product. Available online: https://www.fda.gov/media/88622/download (accessed on 23 February 2021).
91. FDA. Immunogenicity Assessment for Therapeutic Protein Products. Available online: https://www.fda.gov/media/85017/download (accessed on 12 July 2022).
97. FDA. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/125156s0000_LucentisTOC.cfm (accessed on 23 March 2022).
107. McCamish, M.; Woollett, G. The continuum of comparability extends to biosimilarity: How much is enough and what clinical data are necessary? *Clin. Pharmacol. Ther.* 2013, 93, 315–317. [CrossRef]

