Considerations in Establishing a US Approval Pathway for Biosimilar and Interchangeable Biological Products

Bruce Babbitt and Cecil Nick

On 23 March 2010 the US health care reform legislation, the Patient Protection and Affordable Care Act (PPAC), was signed into law, providing a legal pathway for approval of biosimilar products via the Biologics Price Competition and Innovation Act of 2009 (BPCI). Since that time FDA has been working on implementation. Their first step was to form the Biosimilar Implementation Committee (BIC), co-chaired by Dr. Janet Woodcock, Director of the Center for Drug Research and Evaluation (CDER), and Dr. Karen Midthun, acting Director of the Center for Biologics Evaluation and Research (CBER). The agency also set up two review committees: the CDER Biosimilar Review Committee and the CBER Biosimilar Review Committee. FDA then held a two-day public meeting on 2 and 3 November 2010 to gain public viewpoints on the Act’s implementation, including specific issues and challenges. A variety of constituents representing patients, health care professionals, industry associations, and biopharma manufacturers, among others, presented constructive information and concerns. Docket FDA-2010-N-0477 remained open for comments through December 31, 2010.

It is expected that in the 2011 timeframe, FDA will utilize this input in the development of related guidances and rulemaking and, with that, a regulatory pathway will be begin to take shape. This paper represents key concepts that were presented and submitted to FDA in consideration of the US biosimilars pathway by PAREXEL Consulting.

Overarching Principles
Patient safety is paramount, and the safety of biosimilars can ultimately only be established through clinical trials followed by a postmarketing risk management program. The more complex question is the extent to which comparative efficacy needs to be established through clinical trials to demonstrate biosimilarity and interchangeability. If the preclinical and phase 1 programs already demonstrate biosimilarity, there may be no need to repeat the entire original clinical program since the structural and biological similarity would have already been demonstrated to the reference product. Thus, only certain clinical data may be required to confirm similar safety as well as efficacy claims.

A key principle underlying our thinking for streamlining biosimilar development programs is to acknowledge that, unlike when a new biological entity is first developed, much is already known about the structure-function relationships, mechanism of action, and safety and efficacy profile of the innovator product in specific patient populations. In addition, there has been a vast improvement in physico-chemical and biological methodologies such as MS-MS, chemoluminescence and surface plasmon resonance, that can be used for characterization of therapeutic proteins and which can accurately detect structural or functional differences between the products. Furthermore this progress continues apace with the developments of methods such as hydrogen/deuterium exchange nuclear magnetic resonance which can monitor dynamically the three dimensional structure of proteins.

Another key concept is the "totality of the data." By this, we mean that physico-chemical and biological characterization, coupled with nonclinical and clinical pharmacological data, already provide a solid basis for establishing biosimilarity. It is important to balance the value and relevance of these data against the need for clinical safety and efficacy data, particularly where minor or modest differences between the biosimilar and reference products are unlikely to translate into any meaningful differences in the clinic. This approach is similar to that applied to comparability testing data from developers of new biological entities following major changes in manufacturing. While for minor process changes the innovator can draw heavily and successfully on
previous experience, major changes such as the introduction of a new expression construct closely parallel the challenges of developing a biosimilar.

**Selection of Reference Products**

In Europe, at the present time, pivotal safety and efficacy of biosimilars need to be demonstrated against reference products approved and sourced from within the European Economic Area (EU plus Norway, Iceland, and Liechtenstein) as comparator. If the US were to take a similar position and mandate that therapeutic equivalence trials be performed against a US-sourced reference product, this would lead to redundant replication of large phase 3 studies. This, in turn, could raise ethical concerns and would likely have a negative impact on the cost and viability of biosimilars in the US.

Thus, we consider that in cases where it cannot be proven that ex US-sourced reference product and reference product available in the US originate from the same production process and are released to the same specifications, it ought to be quite sufficient to confirm their similarity using state of the art analytical methods to compare both physico-chemical and biological (in vitro and in vivo) properties across these batches. In these circumstances, it ought to be acceptable to support marketing approval using phase 3 safety and efficacy studies against reference products sourced from this ex US source. Where there is the potential for uncertainty, human pharmacokinetic (PK)/pharmacodynamic (PD) equivalence data should, in general, be adequate to confirm the comparability of reference product derived from international sources with US-sourced reference product.

Certainly, if differences were to exist between the differently sourced reference products, these differences would most likely be detected using the more sensitive physico-chemical and biological test methods, together, where necessary and feasible, with comparative PK/PD assessments in humans, than by the much less sensitive approach of safety and efficacy clinical trials. Therefore, the combination of physico-chemical and biological testing, coupled where relevant with clinical PK/PD assessments, generally represents the most rigorous way to show comparability between the biosimilar candidate and reference products sourced from regions outside the US.

The data package bridging the biosimilar candidate drug to both the US and ex US-sourced reference products that would result from the above-mentioned combination of in vitro (physico-chemical/biological) and in vivo (human PK/BE) studies is quite similar to what has been required to support non-trivial changes to the manufacturing process that may be introduced late in clinical development or post-approval. In many cases, the innovator company has not been required to perform even a formal bridging PK equivalence study, due to the perceived strength of the in vitro comparability data.

**Extrapolation Between Indications**

Clinical trials on new biological entities were designed to demonstrate efficacy against placebo; however, placebo trials will often no longer be viable, at least for serious and life-threatening diseases for which there now will exist effective therapies. Therefore, for follow-on biologics, it is necessary to perform non-inferiority or equivalence trials. In order to ensure adequate retention of efficacy, such trials generally need to be several-fold larger than the initial placebo-controlled trials performed by the originator. If equivalence trials were to be performed in each indication, the development program would thus need to be many fold larger than the original program, negating the whole concept of a biosimilar pathway.

As long as the mechanism of action is relatively well understood across indications, it should be possible to demonstrate comparable safety and efficacy of the biosimilar and reference product in one adequately sensitive indication, and then taking into account all prior analytical, nonclinical and clinical comparability data, to conclude therapeutic equivalence for all indications, ie, extrapolate the results across multiple indications. By adequately sensitive, we mean a patient population where differences between the biosimilar and reference products, if they exist, are most likely to be detected and that generates a dataset with a relatively low coefficient of variation.

Even in cases where the target indications are highly diverse, for example, as is the case with many approved monoclonal antibody products, extrapolation should be considered reasonable. Monoclonals can theoretically exert multiple effects in vivo, and the relative contributions of these effects might well not be the same across indications. Nevertheless, if all relevant biological effects are understood to be involved in generating a therapeutic effect in the selected target population, trials in this population should be an adequate indicator for general therapeutic equivalence.
Primary Endpoints
For those innovator products where there exists a strong and well defined pharmacodynamic effect, for example, as with insulin and G-CSF, PD markers may trump clinical endpoints in terms of precision and could be equally or more relevant. When using PD endpoints as a surrogate for clinical endpoints, the PD effect needs to be dose-sensitive and correlate to the therapeutic effect. If this correlation can be adequately demonstrated from the literature, then a PD marker will be a more sensitive measure of efficacy than a clinical endpoint.

The use of surrogate clinical endpoints can also play a key role in the development of biosimilars. For example, in oncology, overall survival is considered the gold standard, but this is often not a practical endpoint for non-inferiority trials. Moreover, demonstration of survival becomes extremely challenging as new next-line therapies become available to treat disease progression. Progression-free survival is perhaps the next best endpoint, but can also take many years to demonstrate equivalence. Therefore, a more practical approach would be to use a response parameter as the primary endpoint, and to include progression-free survival, and where possible, overall survival, as secondary endpoints. There is already regulatory precedence for this, for example, in the approval of liposomal doxorubicin versus doxorubicin.

Equivalence Margin/Non-Inferiority Margin in Therapeutic Trials
The draft FDA guidance for industry on non-inferiority trials states that it “is common in NI trials for the test drug to be pharmacologically similar to the active control. (If they were not pharmacologically similar, an add-on study would usually have been more persuasive and more practical). In that case, the expectation of similar performance (but still requiring confirmation in a trial) might make it possible to accept a single trial and perhaps could also allow less conservative choices in choosing the non-inferiority margin.”

Clearly in the case of biosimilars, not only are the pharmacological principles similar, but the molecules are structurally highly similar if not identical. Furthermore, data will have been generated to demonstrate equivalence with respect to receptor binding, biological properties, PK and possibly PD profiles. There will, therefore, often be no scientific reason to expect that comparable efficacy would not be achieved for a biosimilar product, with the possible exception of the effect of enhanced immunogenicity, which cannot be predicted through physico-chemical, biological and single-dose clinical testing and needs to be evaluated in a sensitive patient population. For biosimilars, it is the totality of data that needs to be assessed in order to judge therapeutic equivalence, and confirmatory clinical trials represents just one element of the biosimilarity program. Thus, acceptance of a less conservative therapeutic equivalence margin should be reasonable, so long as some level of efficacy over putative placebo is demonstrated and immunogenicity is measured directly with no evidence suggesting a different immunogenicity profile.

Non-inferiority or Equivalence Design
On the issue of whether to perform a non-inferiority or equivalence trial for demonstration of biosimilarity, where a protein exerts a direct physiological effect and there is a close correlation between dose and therapeutic effect—as is the case for products such as insulin, G-CSF, epoetin and somatropin—clearly supra-activity will be a safety concern and equivalence trials are likely essential.

However, in the case of monoclonal antibody products, clinical effect in inflammatory disease or oncology is generally measured in terms of the proportion of responders and not as a physiological effect. Under these circumstances, it would seem more appropriate to perform non-inferiority trials, since an increase in responders, i.e. crossing the upper bound of the selected equivalence margin, should not in itself be a reason to reject a product as biosimilar, especially given that safety would be monitored independently. It should be stressed that this does not mean a biosimilar can claim superiority, and, in any case, to conclusively demonstrate superiority between two products, which are ostensibly similar, would require a far larger study program than might be required to show non-inferiority. If there was a trend to superior efficacy, this would at the very least need to be explained; one such possibility is reduced immunogenicity, which is certainly a possibility even for two highly similar molecules, due to aggregation or the contribution of impurities, which can act as adjuvants.

Interchangeability
One of the concepts raised by the US Healthcare Reform Law relating to the establishment of a US legal framework for biosimilarity is the issue of interchangeability between the innovator and biosimilar product. According to the new US legislation, a biological product can achieve biosimilarity status without meeting the higher standards of interchangeability. Specifically,
a product could be approved as interchangeable if the submitted information were sufficient to show that the product meets the following three criteria:

• Firstly, clear-cut biosimilarity to the reference product would need to be established;

• Secondly, there would be the need to demonstrate that the biosimilar would be expected to produce the same clinical result as the reference product in any given patient;

• Thirdly, it would need to be demonstrated that there would be no enhanced risk in terms of safety or diminished efficacy when alternating or switching between the biosimilar and reference product, compared with the risk of using the reference product without such alteration or switching.

The first hurdle, which is to meet biosimilarity, has largely been addressed in the above considerations. In some circumstances, the same data required to demonstrate biosimilarity might arguably also be adequate to demonstrate interchangeability. In this respect, consideration needs to be given to what exactly is meant by interchangeability. On the one hand, interchangeability could be interpreted as initial substitution by the pharmacy, after which patients are retained on the same brand of biosimilar. Alternatively, it could involve allowing indiscriminate switching between brands, based purely on the availability of stock at the time.

In general, substitution at the start of therapy ought to be satisfied simply by the demonstration of biosimilarity, since if two products are proven to be similar there should be no reason to choose one over the other. However, whereas indiscriminate switching is accepted for small molecules, the situation is more complex for protein therapeutics. In particular, possible differences in structure and impurities between the reference product and biosimilar are to be expected and, in any case, cannot be proven to be totally eliminated. Any such differences might impact safety, efficacy or immunogenicity, depending on the product in question, and thus increase the risk to patients if they are switched between these products. Furthermore, random switching would make it difficult to link adverse events such as immunogenicity back to the causative product.

Thus, the granting of interchangeable status that permits switching will require more detailed evaluation as part of the development program. The extent and relevance of any risk will vary depending on the properties and uses of the product and the potential differences between the reference product and biosimilar.

In situations where it is important to control a physiological function within a tight range, e.g., the use of insulin in controlling blood glucose or, to a lesser extent, epoetin in the treatment of anemia, there is certainly a need to ensure close equivalence in terms of potency, pharmacokinetics and pharmacodynamic parameters, such that the therapeutic response remains within acceptable boundaries following switching. Thus, the 80%-125% PK/PD margins that might generally be expected for acceptance of biosimilarity will need to be justified on a case-by-case basis and, in rare circumstances, tighter bounds may be needed to qualify for interchangeability status. In deciding on and justifying the equivalence margin, there is the need not only to take into consideration clinical considerations but also to formulate an understanding of the batch variability of the reference product in terms of protein content and biopotency. It would not seem logical to set PK and PD equivalence margins tighter than the accepted batch-to-batch variability limits of the reference product or those that have been used historically to demonstrate comparability following major process change. The issue of drift over time in the specifications and properties of the reference product has been raised as a potential concern. However, if the reference product is to retain the same proprietary name, it should meet the same criteria as a biosimilar for interchangeability following process change and excessive drift in product characteristics should, at least in theory, be controlled.

Many products such as monoclonals possess a broad therapeutic margin, which means that application of tight bounds becomes less critical. Furthermore, the high degree of variability in potency and pharmacokinetic measurements and the large batch-to-batch variability observed even within the same process associated with such proteins means that the application of relatively tight margins may also not be practical or meaningful.

From a structural perspective, glycosylated proteins, in general, represent a more complex picture than non-glycosylated proteins. In some cases, more comprehensive clinical testing might be required to grant interchangeability status, as there is likely to be less certainty around the ability to...
demonstrate comparability at the structural level compared to non-glycosylated proteins. Differences in glycoprofile could have a profound effect on potency, PK, PD and/or immunogenicity. For non-glycosylated proteins there is, conversely, often more opportunity to rely on physico-chemical data as a basis for establishing biosimilarity and interchangeability, with variations in the impurity profile representing the most likely potential differences. While these might be expected generally only to impact immunogenicity other toxic effects may not be totally excluded.

Immunogenicity is likely the most critical factor to consider when dealing with multiple switching, in that no matter how similar the biosimilar is to the reference product, there will be differences, at least with respect to process-related impurities, and these differences could, in theory, impact immunogenicity, particularly where levels are relatively high. The risks for immunogenicity will depend on a multitude of other factors, as well, such as the nature of any differences, the characteristics of the product, the route of administration, the expected level of immunogenicity based on experience with the reference product and the impact of increased immunogenicity in terms of safety and efficacy. Because switching exposes patients to two or more sets of process impurities, the probability of immunogenicity might be expected to increase with switching since different patients may respond differently to different impurities or to possible conformational differences. In deciding the nature and extent of data required to support interchangeability, there is therefore a need for a risk analysis on the impact of enhanced immunogenicity, taking into consideration the above factors.

Based on the risk assessment, additional clinical data may be required. Such data may be generated either as a prospective clinical trial or as a post-approval monitoring program following approval as a biosimilar but prior to granting interchangeability status. The design of a prospective clinical trial for demonstrating interchangeability presents significant challenges. A standard crossover design is not likely to provide interpretable data since there will be no way of ascertaining which product triggered immunogenicity. Anti-drug antibodies observed in the second period may have arisen as a result of immunization in the first period, and, due to the highly similar nature of the two products, these antibodies would be expected to cross react. It would also not be possible to disentangle the impact of any period, sequence or carryover effects, and the ethics of performing a trial that in theory might place a patient at risk, but, does not deliver meaningful data, will be highly questionable.

In any case, the question is not whether the biosimilar is more immunogenic or less effective than the reference product, since if this were to be the case, approval even as a biosimilar ought not to be acceptable. Rather, it is whether indiscriminate switching enhances the chances of immunogenicity, jeopardizes safety or reduces efficacy that needs to be explored. In this respect one possible approach might be a three-arm parallel trial comparing, eg, for an anti-TNF product, a 12-month treatment between the reference product (arm 1), biosimilar (arm 2) and patients switched between the two, eg, every three months (arm 3). The primary endpoint would be non-inferiority in terms of immunogenicity and/or efficacy (depending on the protein) between the reference product (arm 1) and the switching arm (arm 3), while the biosimilar product arm (arm 2) is suggested as a baseline check.

The problem is that such an approach has limitations, because in order to allow for realistic trial sizes, a relatively broad non-inferiority margin would need to be set, meaning that only large increases of risk in terms of safety or diminished efficacy would be detectable. The alternative of applying tight margins would in many cases necessitate very large trials comprising many thousands of patients, often making the attainment of interchangeable status a nonviable option and so defeating the purpose of establishing an interchangeable concept. Furthermore since the trial would be investigating risk without any promise of therapeutic benefit to the patient, the ethics of such a trial is open to question.

There will be times when even a small increment in terms of risk or diminished efficacy will be unacceptable, eg, where there exists the danger of inducing autoimmunity, meaning that for some products the bar for interchangeable status should indeed remain high, even if this precludes any biosimilar version ever obtaining interchangeable status.

In view of the limitations of switch studies, another option might be to establish a safety database following approval of the product as a
biosimilar and then use these data to support safety for subsequent later approval of interchangeable status. Such an approach would also not entirely eliminate the possibility of an increased risk associated with indiscriminate switching, but would at least provide confidence on the safety of the biosimilar such that one could feel reasonably confident that switching between products would not introduce undue additional risk.

**Nomenclature**

There is an interrelationship between interchangeability and nomenclature. One might expect that an interchangeable biosimilar should be considered to have the same active ingredient as the reference product and, therefore, should be granted the same USAN, while non-interchangeable biosimilars would be considered to have a new active ingredient and, on the face of it, would need to be issued with a new USAN to ensure against interchangeability and to facilitate tracking. There are, however, a number of complexities with this approach.

Firstly, the precedent has already been set to some degree in granting the same USAN to related non-biosimilars, for example, somatropin and insulin, which although currently regulated by section 505 of the Food Drug and Cosmetics Act are nevertheless proteins that will in future be classified as a biological under the Healthcare Reform Law.

Secondly, if a non-interchangeable biosimilar were subsequently granted interchangeable status, the USAN would presumably need to change, although the product would not have changed! Thirdly, there is the risk that by giving two highly similar products different USANs, clinicians may be misled into thinking these are products that are substantially different, which might lead to patients unable to tolerate one product being inappropriately switched to the other.

For these reasons, issuing distinctly different USANs for each biosimilar may not be an ideal solution and differentiation might best be achieved through unique branding. However, how interchangeability will be indicated on the label, if not through use of a common USAN, remains a challenge, as there is a need to indicate not only interchangeable status but also with which product the biosimilar is interchangeable. One possible way of doing this might be the issuance of identical USANs for interchangeable products but slightly different USANs for non-interchangeable products, at least for glycosylated proteins. For example, numbers could be added as suffixes such that the first biosimilar to say epoetin alpha would be designated epoetin alpha-2, the next epoetin alpha-3 and so on; however, this would still not work where numbers have already been designated such as interferon beta-1a. Once interchangeability status has been gained, the differentiating number could be dropped. For non-glycosylated proteins, since the precedent has already been set, it would seem reasonable to continue the practice of issuing a common USAN.

**Conclusion**

Patient safety is paramount, and clinical trials followed by a post-marketing risk management program are likely an essential part of any biosimilar program. Nevertheless, physico-chemical, biological, non-clinical, and phase 1 testing provide already a strong indication of similarity and should eliminate the need for a full clinical development program. The assignment of interchangeable status may require more extensive studies. Nevertheless, a pragmatic approach to regulation of biosimilars is essential in order to facilitate the development and approval of biosimilars and thereby enable healthy competition and improve patient access to biological medicines.

---

**Bruce Babbitt, PhD**, is Principal Consultant, PAREXEL Consulting, Waltham, MA. Readers can contact him at Bruce.Babbitt@PAREXEL.com

**Cecil Nick** is Vice President, PAREXEL Consulting, Uxbridge, UK. Readers can contact him at Cecil.Nick@PAREXEL.com