SCIENTIFIC WHITE PAPER

Reference Standards for Potency Assays

AUTHORS

Laureen Little, *Quality Services and BEBPA*
Dorota Bulik, *Ultragenyx Pharmaceutical Inc.*
Sian Estdale, *BEBPA*
Sebastianus Kolen, *Byondis BV*
Marie Gottar-Guillier, *Novartis*
Mike Merges, *Autolus Ltd.*
Peter Rigsby, *Medicines & Healthcare products Regulatory Agency UK*
Jane Robinson, *BEBPA*
Mike Sadick, *Precision Biosciences*
Perceval Sondag, *Novo Nordisk*

2024

Overview

This paper presents an overview of current practices common in the biopharmaceutical industry for the development of in-house reference standards for use in bioassays to determine the potency of biopharmaceutical preparations. It does not necessarily represent the views of every contributor nor of the organizations to which they are affiliated. It is intended to highlight some of the most common issues encountered in the development of reference standards and the ways in which some organizations address these. It should not be interpreted as an instruction to adopt a particular procedure or methodology. Relevant current regulatory guidelines should always be consulted and discussion with regulatory authorities undertaken when necessary.

Contents

Introduction

Reference standards (RS) are critically important for the execution of bioassay methods as the potency of the test sample (e.g. drug product or drug substance) is determined as a relative potency by comparison to the potency of a reference standard. Consequently, determination of reference standard potency throughout a product's lifecycle must be accurate, precise, reliable, and consistent across time and across multiple batches of reference standard standards. This is a particular challenge early in product development when often only Interim RSs are available and no higher-order standards (e.g. a Primary RS) are available for comparison.

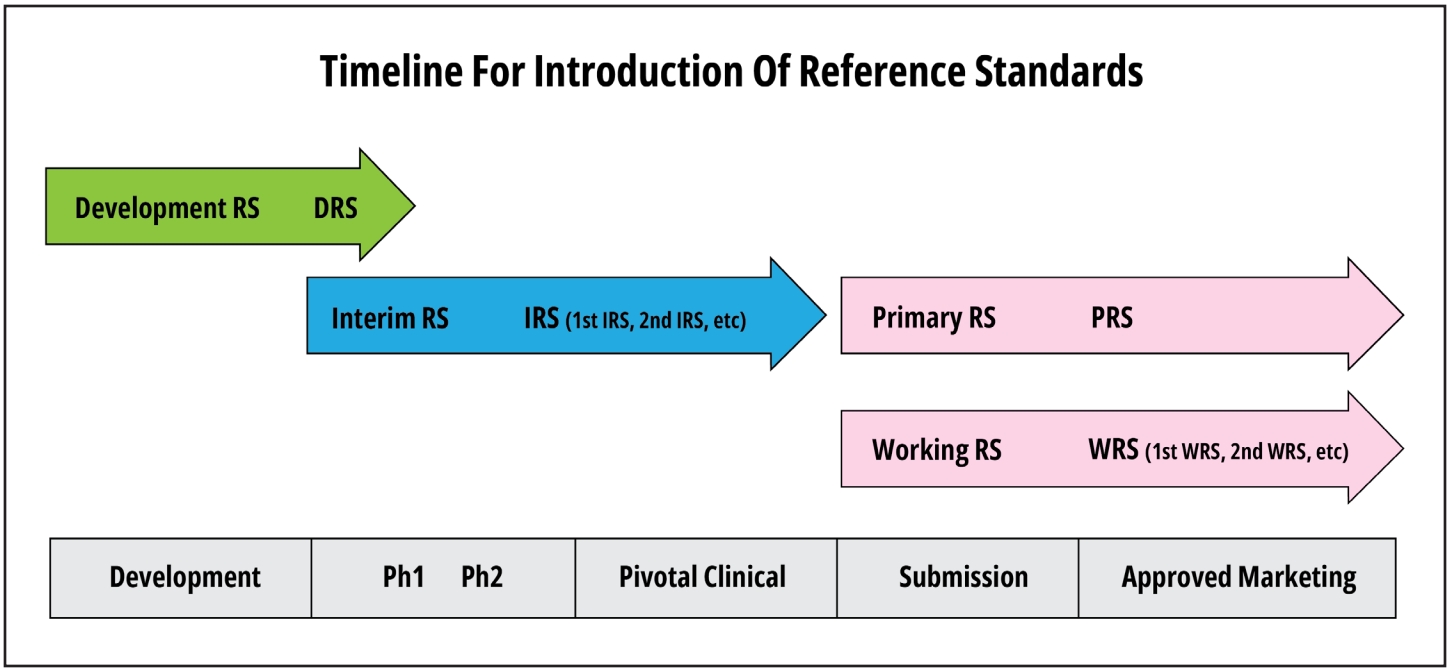
The biological activity of a complex biologic drug cannot be quantitatively determined using physicochemical tests alone, and the potency determined in a bioassay is measured relative to that of another preparation which might itself change. Consequently, a multifaceted approach is required to monitor for potency change of the reference standard and while no single test can prove that potency has not changed, a preponderance of evidence from multiple approaches can support the conclusion whether or not the potency of the standard is changing over time.

In most cases, and the ideal situation, the same reference standard is used for potency tests and physicochemical tests. However, in some cases, this is not feasible. This paper considers specifically the requirements for reference standards for potency tests. With the recent increase in the development of cell and gene therapy products, it has become evident that, in addition to the general considerations involved in the development of reference standards for potency testing, there are issues specific to these Advanced Therapy Medicinal Products. These specific issues will be covered in a future white paper.

Summary Of Abbreviations Used For Various Classes Of Reference Standards

For more detail, see Glossary

Abbreviation	Standard Title	AKA	Description
DRS	(Method) Development Reference Standard	Research Reference Standard	used in early development of a potency assay; may be used to release early clinical material
IHRS	In-House Reference Standard		RS used within a company, in contrast to an external or official reference standard
IRS	Interim Reference Standard	Clinical Trial Reference Standard	prepared from representative clinical lots or from material used for quality control purposes during product development; GMP or non-GMP; suitably characterized
ORS	Official Reference Standard		RS from a source recognized by regulatory authorities, eg. World Health Organization
PRS	Primary Reference Standard		ICH Q7: "a substance that has been shown by an extensive set of analytical tests to be an authentic standard that should be of high purity. ..." ICH 6B: "an appropriately characterized in-house primary reference standard [should be] prepared from lot(s) representative of the production and clinical materials."
RS	Reference Standard		in this white paper, a preparation of a biopharmaceutical used in a bioassay to determine the relative potency of a test sample of the biopharmaceutical
WRS	Working Reference Standard	Secondary Reference Standard	suitably characterized standard prepared from representative cGMP clinical or commercial lots(s); established by comprehensive analysis against a PRS; used in each assay



Usually several batches of potency reference standard are generated throughout the lifecycle of a biologic product. During clinical development this standard will be representative of the current clinical product. Although it is expected that during product development the characteristics of the clinical product will change, (impurity profiles will improve, formulations will change, etc.) it is standard practice to utilize an Interim Reference Standard (IRS) for as long as possible until a two tier reference system (Primary Reference Standard, PRS, and Working Reference Standard, WRS) is adopted. How long a given IRS is used is determined by the quantity of standard available, its stability and whether it remains representative of the clinical product being released. Given the critical function of the IRS, great care must be taken in its selection.

General Principles and Considerations

The procedures adopted for the development of in-house potency reference standards, particularly during the early stages of product development, have to be a compromise between the scientific ideal and the practically feasible. As the product progresses from early development through to license submission, regulatory requirements for the reference standard become increasingly stringent. At early stages of product development, knowledge of the product is limited and, commonly, the bioassay also will be in early development. For the reference standard, investigation of factors such as optimum formulation, container, processing, storage, etc. and level of characterization will usually be limited by the resources, amount of product available and capability of the bioassay. As product and bioassay development progress, greater characterization of the reference standard becomes necessary and it may become apparent which factors require further investigation and investment.

From the outset, maintenance of documentation is absolutely essential to the reference standard development program. In addition, the following points are worth bearing in mind. They will be considered in more detail later in the paper at the stage of product development when they become regulatory requirements or are most commonly found to be necessary.

- The requirements for the reference standard are not necessarily the same as those for the drug substance or drug product. Special formulation, processing and storage may be needed to ensure, for example, long-term stability and compatibility with the bioassay.
- Usually, and ideally, the same reference standard should be used for the potency assay and other analytical techniques. However, it is possible that different analytical techniques may require a special preparation or formulation of the reference standard.
- Ideally an aliquot of the reference standard should be used for only a single potency assay. A single potency assay may require only small quantity of product which is less than that filled

in the vial or ampoule of drug product. In this case, preparation of the reference standard in small aliquots in small containers may be advantageous, consuming less product and requiring less storage facility.

- Samples should be laid down as early as possible to permit real-time stability studies under proposed storage conditions.
- Replacement reference standards will almost certainly be required. Retained samples should be kept at all stages even though the degree of continuity of a program that they provide may be limited at early stages of product and reference standard development.
- Stocks of a reference standard should be carefully monitored and a replacement prepared while there is still sufficient stock for comparison with the candidate replacement and for retained samples.
- At any stage, a candidate reference standard should selected as being representative of the product as currently manufactured.
- Data from physicochemical analyses must be used in combination with potency assay data in selecting a reference standard that is representative of the currently manufactured product and in monitoring the stability of the reference standard.
- Candidate replacement standards can be run alongside the existing standard in routine assays, commonly as control samples, permitting accumulation of extensive data to judge their suitability as future reference standards.
- On replacement of a reference standard, direct comparison of the new standard with its immediate predecessor is essential and comparison with retention samples of earlier standards can help provide continuity of data.
- If an official reference standard (eg. WHO, USP, etc.) is available and suitable, it can prove useful from the early stages of the RS development and there are regulatory expectations for its use at later stages.
- Stocks of current reference standards and retained samples should be stored in two or more independent sites. Although the stocks would normally be held under nominally identical conditions, periodic testing can verify comparability between sites.

Reference Standard for Method Development (Development Reference Standard, DRS)

The establishment of a potency reference standard during early product development can be overlooked until a potency assay is requested by a regulatory agency to release clinical trial standard. In practical terms, this is much too late. It is best if a reference standard is selected as soon as a biological assay is being developed. Initially, many sponsors call this working standard, secondary standard, working development reference, a “control” sample or interim reference. In this document we will refer to it as the Development Reference Standard (DRS). A key activity during early development of the potency assay is the selection of a DRS and the aliquoting of this standard into single use vials for storage under suitable conditions. This reference standard is often sourced from non-GMP material, obtained from the Research Department. It is best if this standard is run on every plate during method development.

Once a dose response curve is obtained and a method determined for quantitating a potency value using a comparison of the test sample against the reference standard, the DRS can be re-evaluated and may be replaced with a RS derived from a batch of product which is more representative of the current manufacturing process. This typically happens during safety/toxicology studies or Phase I clinical trials. During the transition from the DRS to clinical trial reference standard, referred to here as the first Interim Reference Standard (IRS), it is wise to continue to include the DRS in all assay runs as a positive control. However, it is important to retain aliquots of the DRS as it may be required later in development for stability or bridging studies.

First Interim Reference Standard (First IRS)

During late clinical development or in the commercial phase, the selection of the reference standard is a more straightforward (although not easy) task. A lot of product that is representative of the product lots currently in production is selected and the prescribed testing and characterization are performed. However, for the First IRS it can prove difficult to find a lot that is truly representative and, in fact, it is likely that the lot selected will not be representative of the product as ultimately developed for commercialization.

First Interim Reference Standard (First IRS) Cont.

Aspects to focus on in early development are that the standard represents the product as it is currently produced and that it meets some of the following requirements/suggestions:

1. Known biological activity. The standard should show activity in known animal models and/or critical cell-based models or other relevant activity assay, even if these procedures might not yet be able to report rigorously defined activity units. The First IRS should be demonstrated to have the appropriate biological activity.

2. Proposed First IRS lots should be microbiologically free of adventitious agents. The presence of adventitious agents often will destabilize product. This standard will be tested using the appropriate sterility tests, which may or may not be compendial (such as USP <71>, etc.) This might also include bioburden testing. IRS lots that are stored in deep frozen conditions might not require this testing.

3. It is not necessary that the First IRS be manufactured under GMP conditions. In fact, the first IRS standard is sometimes the toxicology batch and not manufactured under GMP. However, the First IRS lot should be manufactured in a clean facility, with good documentation.

4. Accelerated stability studies are typically carried out, including multiple freeze/thaw cycles. Appropriately representative stability data may be utilized. If the First IRS is not stable to frozen storage, formulation might be undertaken to stabilize the standard. It is not necessary that the First IRS be formulated as the clinical material for use in humans. However, it is important that the formulation does not significantly alter the shape of the dose-response curve in any relevant activity assays.

5. Physical/chemical characterization studies are instituted as soon as the appropriate methods become available, such studies may include: (Note these are common analyses, which are not required, nor are they the only acceptable methods):
   a. Amino Acid Analysis
   b. Carbohydrate analysis
   c. Impurity profile
   d. Immunoreactivity
   e. Aggregation/high order structure
   f. Appearance testing
   g. pH
   h. concentration analysis

Treatment Of The First IRS

Once selected, it is critical that the First (and all subsequent) IRS be handled with care.

1. Prepare the standard in single use aliquots. Careful consideration should be given to the storage of these aliquots. Some companies, knowledgeable about the stability of their product type store at -70°C and consider that as stable and sufficiently safe. When it is known that the standard is not stable to freeze/thaw the best storage condition should be selected. This could include, but is not limited to: storage at -20°C in a solution designed to prevent freezing, lyophilization, storage at other temperatures. If there is a paucity of knowledge about the stability profile, it is recommended to generate data at multiple storage conditions.

2. Designate a minimum of 20 to 100 (depending on practical requirements) retention samples designed for use just prior to commercialization. These may prove critical during the bridging studies needed to determine final commercial specifications.

3. Designate enough standard to support real time stability studies. This should include enough vials for a 5 to 10 year study and enough vials to perform complete characterization at least once or twice per year based upon historical knowledge.

Releasing Clinical Lots With The First (or other) IRS

The first IRS is often assigned a value of 100% potency if no suitable comparator exists at this stage of product development. Setting the potency as 100% (in absence of other existing standards) is the most common practice. However, if successive standards are found by comparison with their immediate predecessor to be 100% potency within set limits and so each successive potency is set at 100%, then drift can occur unnoticed. Using in-house potency units and assigning the actual measured potency can make any drift more obvious.

The following data may support the assignment of 100% potency.

- The First IRS is sourced from a representative batch of clinical batches. As stated above, if this standard is manufactured by the current manufacturing process utilizing the appropriate cell substrate it can be assigned 100% potency.

Releasing Clinical Lots With The First IRS Cont.

- The First IRS is analyzed in the existing potency assay(s) and compared to the DRS and other prior product batches. It is also compared to other candidate batches manufactured with the current method. A bioassay expert will need to document that the RS results in an acceptable dose-response relationship in the bioassay.
- The First IRS must comply with all applicable existing release specifications.
- Available physicochemical tests are also run to determine that the proposed First IRS is representative of the current production batches. These test results are compared to all other candidate First IRS batches.

If the First IRS meets an appropriate subset of the above criteria, then it may be used to release early clinical batches with an assigned potency of 100%. Other scientifically sound and robust approaches may also be used to qualify an IRS to release the first human clinical material.

The critical aspect is that the IRS assigned a 100% potency must represent product batches manufactured by the current process.

Re-qualification and/or Monitoring of an IRS

To demonstrate continued suitability for use and acceptable stability, IRS batches are tested at periodic intervals. In the case of IRS batches, there is no higher-order RS with which to compare unless an official reference standard is available and suitable. In addition, the property of biological potency cannot be directly correlated to physicochemical test methods. As a result, a multifaceted approach is used to re-evaluate these reference standards. The combined evidence of all reasonable tests to demonstrate potency stability allows the conclusion to be reached that the reference standard remains suitable for use. The following examples could be considered as part of this justification.

- Comprehensive physicochemical testing shows no changes that could result in a change in potency
- Monitoring and trending the potency of an independent plate control relative to the IRS.
- API and/or drug product release and stability result show no obvious trends unless these can be explained by something other than a change in the reference standard.
- Evaluation of any supporting stability data and assessment of molecular properties that may impact biological activity.
- Similarity of the dose-response curves of the IRS (Slope, Upper and Lower Asymptotes) when tracked chronologically.

In addition to data from the multifaceted approach, all reasonable measures should be taken to protect the RS from degradation or other changes that could impact potency, such as the following.

- Overprotective packaging and storage temperature
- Avoidance of freeze-thaw cycles
- Source the RS from highly representative batch

Subsequent IRS

The First IRS is used as long as there is sufficient stock and it is deemed representative of the manufactured clinical material to be used in the human clinical studies. A replacement of the reference standard is needed if the IRS stock is approaching exhaustion. This often occurs during Phase I and Phase II product development as batch sizes tend to be smaller and many exploratory potency bioassays will be needed to support product development. Therefore, it is important to develop and understand a process to replace interim reference standards.

Replacement of an Interim RS which is still Representative of Clinical Trial (CT) Material

- It is preferred that the new reference standard be derived from a clinical trial (CT) drug substance batch to ensure it is representative of the clinical material already (or to be) released for use in the clinic. If a new IRS batch is generated, its potency will be determined by comparison testing relative to the reference standard being replaced.
- The value assignment of the new IRS can either be value assigned from the IRS being replaced or if the value is sufficiently close to 100%, it may be given that value.
- The variability of the potency assay will be used to determine the minimum sample size that will result in a sufficiently accurate potency assignment.

Replacement of an Interim RS which is still Representative of Clinical Trial (CT) Material Cont.

- It is important that none of the reference standards be completely depleted during product development, thus it is important that a set of retention samples be established immediately during the manufacture of the IRS.

Replacement of an Interim RS which is no longer Representative of Clinical Trial (CT) Material

If at any point it is determined that the First IRS is no longer representative of the CT drug product or if the First IRS was not derived from a batch that is representative of CT drug substance, a second IRS is usually established with representative CT batch even if the first IRS has not reached a low stock situation. For example, if there is a significant process change (e.g., a cell line switch) a second IRS may be established even if no change is observed in the dose response curve in the potency assay. In this case the second IRS is qualified by comparison of the new standard with the first IRS. This process is described in the next section.

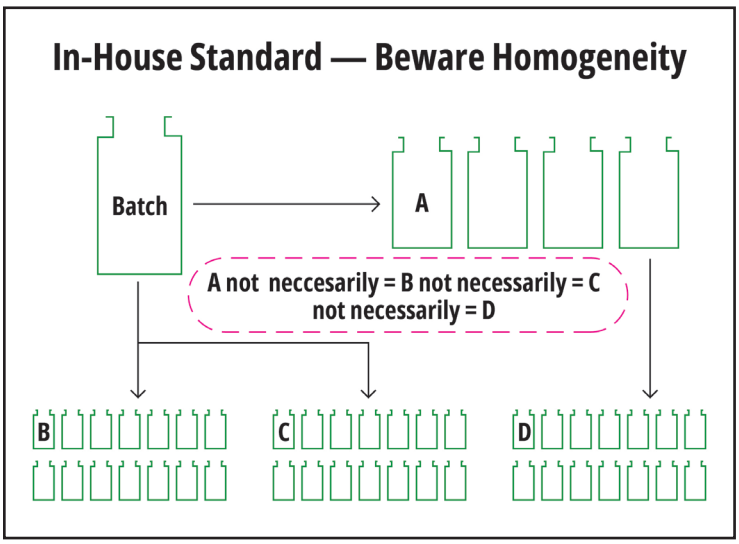
If a significant process change results in a change in the biological activity of the clinical standard then a new IRS must be established to release CT drug substance/product produced by this process. If there is a shift in the new IRS dose-response curve characteristics, such as a change in the curve shape, change in the ED50 or alteration of the asymptotes, in the existing potency assay, a new IRS to release CT drug substance cannot be qualified by a comparison to the First IRS. This mandates the development of a representative Second IRS. The qualification of this new IRS is a difficult situation as the new IRS cannot be compared quantitatively to the prior IRS (First IRS or the DRS). However, it is important that all three (e.g. Second IRS, First IRS and DRS) be compared head-to-head to demonstrate the need to change to the Second IRS. The selection of the specific lot for the replacement IRS in this case will be similar to that of the First IRS. All candidate CT lots are analyzed in the existing potency assay and all available physical/chemical tests are completed. The entire dose-response curve in the potency assay from all candidate lots are compared and the most representative batch should be selected. In determining which lot is the most representative of the currently manufactured product. It may be necessary to assign different weightings to various properties. The potency, or specific activity, is a major, but not sole, consideration. A lot which is at the extreme of the range for any property should be avoided.

The selected batch can be assigned 100% and used to release future lots.

Establishment of this type of Second IRS often occurs at a clinical trial interface such as going from PI to PII. It may also occur going from PII to PIII or PIII to commercial. Information about these significant changes is typically submitted to regulatory agencies prior to initiating the clinical studies. Any change of potency RS which cannot be directly bridged to the prior RS should be included in this regulatory briefing

Preparation of Further Aliquots of Reference Standard on Depletion of Stock

It is sometimes possible to prepare a further stock of aliquots of reference standard from the same batch of product as used previously. However, the handling and storage history of this later set of aliquots will not be identical to that of the first. Rigorous testing is necessary to demonstrate similarity of the initial and later sets of aliquots. Aliquots must be clearly identified as to which set they belong and the identity of which set is used must be recorded in every application.



Case Studies

Case 1: Supply of First IRS is running low, but is still representative of the clinical standard. The following approach can be taken:

1. Check the similarity of the dose response curve of the First and Second IRS.
2. Complete all characterization studies as outlined above.
3. Compare the proposed IRS head-to-head with the original IRS. The number of runs required will depend on the precision of the method.
4. The new IRS may be value assigned off the original standard if a sufficiently precise biological activity assay has been developed at this time. However, if this is not the case because the original or current bioassay is too imprecise to perform a quantification of the new interim off the First IRS, the proposed replacement lot may be considered to be 100%. In this case it is critical that sufficient retention lots be laid down to do a retrospective value assignment.

Case 2: Manufacturing changes result in CT product which no longer has dose response curves similar the First IRS. The following approach can be taken if early PI to PII:

1. Complete all characterization studies as outlined above for the First IRS.

   a. It can be advantageous at this point to determine if the proposed IRS lots are microbiologically free of adventitious agents

   b. Accelerated stability studies should be carried out, including multiple freeze/thaw cycles. If the standard is not stable to frozen storage, formulation might be undertaken to stabilize the standard. (NB it is not necessary that the IRS be formulated for use in humans, however, it is important that the formulation does not significantly alter the shape of the dose-response curve in any relevant activity assays)

   c. Full characterization studies should be available or instituted as soon as the methods become available. Such studies include but are not limited to the following:

      i. Amino Acid Analysis

      ii. Carbohydrate analysis

      iii. Impurity profile

      iv. Immunoreactivity

      v. Aggregation/high order structure

      vi. Appearance testing

      vii. pH

      viii. concentration analysis

2. Work with toxicology, development and preclinical safety groups to confirm no new impurities have been introduced into the proposed clinical trial standard.
3. Perform head-to-head potency assays with the First and Second IRS to demonstrate that the dose-response curve has been altered.
4. Provide an analytical assessment of the First and Second IRS which includes biological and physical chemical characterization.
5. Determine if there have been any formulation changes which might account for changes in the dose response curve. The formulation may affect the response measured in a bioassay. To test whether a formulation change is responsible for an observed change in dose-response curve, the following steps can be followed provided the test sample and reference standard undergo a sufficiently large dilution on addition to the bioassay:

   a. adjust the assay medium composition of the reference standard to give the same final composition in the assay as the test sample of the new clinical material.

   b. compare the dose-response of the reference standard in the original assay medium with that in the new assay medium

   c. compare the dose-response curve of the reference standard in the new formulation of assay medium with that of the new clinical test material

6. If possible, determine the rationale for the altered dose response curves (for example perhaps an alteration of the glycosylation patterns in the active ingredient, removal of process related impurity, etc.) this can be done by purifying various heterogeneous active moieties to determine if there was an altered dose-response for dominant and minor components. This information may be available in a product development report.
7. Prepare a scientific package for regulatory approval for utilization of the newly altered CT IRS in future studies.

8. Screen all available CT lots for the Second IRS.
9. Choose a "middle of the road" potency standard which meets all the same criteria as required for the First IRS. If necessary multiple lots may be pooled, but this should be avoided if possible.
10. Assign this lot a potency of 100%.

Case 3: Manufacturing changes result in CT IRS which no longer has dose response curves similar the previous current IRS. The following approach can be taken PII to PIII:

1. Complete all characterization studies as outlined above for the first and second IRS.

   a. Determine whether the proposed RS lots are microbiologically free of adventitious agents.

   b. Accelerated stability studies should be carried out, including multiple freeze/thaw cycles. If the standard is not stable to frozen storage, formulation might be undertaken to stabilize the standard. (Note: it is not necessary that the IRS be formulated for use in humans, however, it is important that the formulation does not significantly alter the shape of the dose-response curve in any relevant activity assays)

   c. Full characterization studies should be instituted as soon as the methods become available such studies include but are not limited to the following:

      i. Amino Acid Analysis or some form of sequence analysis
      ii. Carbohydrate analysis
      iii. Impurity profile
      iv. Immunoreactivity
      v. Aggregation/high order structure
      vi. Appearance testing
      vii. pH
      viii. Concentration analysis

2. Work with toxicology, development and preclinical safety groups to confirm no new impurities have been introduced into the proposed clinical trial standard.
3. Perform head-to-head potency assays with the first and second interim RS to demonstrate that the dose-response curve has been altered.
4. Provide an analytical assessment of the first and second interim RS which includes biological and physical chemical characterization.
5. Determine if there has been any formulation changes which might account for changes in the dose response curve.
6. Prepare a scientific package for regulatory approval for utilization of the newly altered clinical trial standard in future studies.
7. Screen available all available lots for the second interim RS.
8. Choose a "middle of the road" potency standard which meets all the same criteria as required for the first IRS.
9. Assign this lot a potency of 100%.

Case 4: Establishing IRS for PIII:

1. Complete all characterization studies as outlined above for the first and second IRS.

   a. Determine the proposed RS lots are microbiologically free of adventitious agents as needed.

   b. Accelerated stability studies should be carried out, including multiple freeze/thaw cycles. If the standard is not stable to frozen storage, formulation might be undertaken to stabilize the standard. (Note: it is not necessary that the IRS be formulated for use in humans. However, it is important that the formulation does not significantly alter the shape of the dose-response curve in any relevant activity assays)

   c. Full characterization studies should be instituted as soon as the methods become available such studies include but are not limited to the following:

      i. Amino Acid Analysis
      ii. Carbohydrate analysis
      iii. Impurity profile
      iv. Immunoreactivity
      v. Aggregation/high order structure
      vi. Appearance testing
      vii. pH
      viii. concentration analysis

Case Studies Cont.

2. Confirm that the impurity profile does not contain new impurities
3. Perform head-to-head potency assays with prior interim RS to demonstrate that the dose-response curve has not been altered.
4. Screen available all available lots for PIII IRS.
5. Choose a "middle of the road" potency standard which meets all the same criteria as required for the First IRS.
6. Determine an appropriate number of runs required to value assign the PIII IRS from the PII IRS. This is a power calculation which will depend upon the precision of the potency method and the required confidence interval for the PIII

Reference Standard to Release Commercial Product

Release of commercial products typically requires a two-tier reference system. These two tiers consist of a primary reference standard (PRS), sometimes referred to as the gold standard, and a working reference standard (WRS), sometimes referred to as the secondary reference standard.

Primary Reference Standard (PRS)

The First PRS should be of sufficient quantity to last for as long as possible, preferably the entire commercial life of the product. If stock of a PRS batch does approach exhaustion or cannot be made stable for this length of time, it is necessary to establish a replacement PRS. Subsequent PRS batches are tested using the PRS that is being replaced and are made using representative commercial drug substance/product.

To reduce the risk associated with loss of a reference standard, stock should be stored in at least two locations that are as independent as possible of each other. Sufficient stock should be held in the second location to permit studies to establish a replacement standard if necessary. This is particularly important for the PRS.. Although the stocks would normally be held under nominally identical conditions, periodic testing can verify comparability between sites.

Potency Assignment Strategies

The PRS and WRS must have characterization strategies to assign the potency and to monitor for stability. The following sections describe an approach for initial establishment, re-qualification, and replacement of IRS with a PRS for release of commercial drug substance/product.

A PRS is implemented to serve as the basis for establishing all future reference standards. The PRS is considered the standard to which all subsequent reference standard batches are linked.

WRS batches are qualified using the PRS as the higher-order standard. The source material for WRS batches is from representative commercial drug substance or drug product (DS/DP) production. It is recommended that a WRS should not be prepared from the same production batch as the PRS. Having the WRS and PRS sourced from different batches can assisting in identifying potential issues such as problems with stability.

The transition from the IRS to the first commercial product reference standards (PRS and WRS) is a key activity. From the bioassay perspective it is critical that both the appropriate batch of PIII clinical standard or engineering runs be selected and that a formal program for qualifying and monitoring reference standard be established.

At this stage both the PRS and WRS are established. Ideally the PRS comes from a PIII clinical trial batch. This is a practical selection as this standard has been "clinically validated", meaning that it was part of the CT studies which demonstrated that the product was efficacious.

The following steps are commonly taken in establishing this PRS:

- A PIII batch is selected based upon available volume, use in the clinic, and the fact that it is representative of the DS/DP to be manufactured in the proposed commercial process. If a PIII batch is unavailable then one of the engineering batches is sometimes substituted.
- Most companies typically target for a minimum of 20 years supply for the PRS.

The following steps are often taken when establishing the WRS

- If this is the first WRS, it is possible, but not generally recommended, to use the same batch used to establish the PRS. If additional time is required to establish an independently sourced WRS, it may be possible to make a larger batch of PRS and use it for routine testing until the

WRS is established. If this is done, it is critical to not allow the WRS to stock below a level where it will last for the minimum targeted period (e.g., 20 years).

- To ensure adequate stability of the WRS, it may be necessary to select formulations, containers, processing and storage that are different from those used for the final DP.
- If first WRS source material is the same as that for the PRS, then the WRS batch typically comes from the DS/DP manufacturing stream. Sometimes multiple candidate lots are considered during the WRS selection process. This can permit selection of a lot with a potency towards the middle of the potency range anticipated for the DS/DP. Some manufacturers combine aliquots of multiple source lots to make the reference standard as representative as possible.
- It is possible to qualify and utilize a WRS initially as a control material. This allows for a substantial amount of data to be generated in a routine release environment. The control material can then be repurposed for use as a WRS. However, if this is done it is important that the control material be replaced with a new control material once it (the control material) becomes the official WRS.
- Once the final WRS candidate is selected, it is assigned a potency by comparison with the PRS. The number of runs required for this value assignment is determined by the precision of the potency assay and the desired size of the confidence interval for the final potency assignment.
- If aliquots of the WRS are larger than required for immediate use, i.e., larger than "single use", it may be possible to store thawed or reconstituted aliquots for later use, thus avoiding wastage. Such interim storage may include refreezing liquid formulations, freezing reconstituted lyophilized aliquots, or refrigerated storage, for example 2-4°C. If interim storage is contemplated, extensive bioassay and physicochemical analyses must be conducted to determine the conditions and length of such interim storage that do not impact the suitability of the WRS. Any changes detected should be assessed carefully, even if they do not appear to impact the potency in the current bioassay as they could affect the dose-response in different potency assay.

Handling of the Primary Reference (PRS)

The particular properties required of the PRS, such as long-term stability, may require formulation and storage conditions different from those suitable for the clinical product. Measures undertaken to enhance the stability of the reference standard may include the following, as appropriate to the individual drug, the intended assays and logistical considerations. The effectiveness of each measure should be determined for each individual case.

- A specific container+closure system, different from the clinical product. Prevention of gas exchange, moisture entry, reaction with stopper, and reaction or adsorption on the container surface may enhance stability. Suitable containers range from cryovials, stoppered glass vials to heat-sealed glass ampoules, depending on the individual case. The selected aliquot size will influence the container system chosen. For a small volume of reference standard, a large container will have the disadvantages of: a large headspace (increased evaporation, reaction with headspace atmosphere); large surface area (adsorption, reaction with container material); difficulty of recovering standard; costly storage volume requirements. The container for the reference standard is commonly different from that of the clinical product, in which case, specific studies are required to determine its suitability.
- Reduction of oxygen in the headspace. Filling the headspace with an inert gas may increase stability by reducing oxidative reactions. To be effective, this should be used in combination with a container system that reduces gas exchange.
- A special formulation, different from that used for the drug product formulation. In addition to increasing stability, other drivers for selection of a special formulation for the primary reference standard may include the need for compatibility with the intended analytical methods and including a carrier or bulking agent when each aliquot contains only small concentrations or quantities of the API compared with the clinical drug formulation and container. Sugars and carrier proteins are sometimes used. Difference in formulation between the clinical product and reference standard may cause non-similarity in the dose-response curves. This may be overcome at high dilution or may require compensation in the assay medium.

Handling of the Primary Reference (PRS) Cont.

- Lyophilization. For many biologicals, lyophilisation can enhance long-term stability by removal/reduction of the water involved in chemical reactions. The optimum residual water content and suitable formulation must be determined for each individual case. Too low a residual water content can adversely affect stability. Compositions suitable for liquid formulations may not be suitable for lyophilisation. Stabilizers may be required. Use of lyophilization may require investment in infrastructure and investigation into optimal conditions for the lyophilisation and reconstitution processes but may offer improved long-term stability and permit storage at higher temperatures (for example, a lyophilized material might be stored at -20°C where a liquid formulation required -80°C). A Lyophilized material is not susceptible to freeze-thaw transitions, so this may permit easier handling.

- Storage conditions. In many cases, long term stability can be enhanced by storage at lower temperatures than used for the clinical product. For liquid formulations, the primary reference standard is commonly stored at -80° C. If the clinical product is not stored frozen, specific studies will be required to determine whether the freezing process damages the material. Stabilizers may be required. Freeze-thaw cycles of aliquots should be avoided so the time interval and appropriate holding conditions for a thawed aliquot must be determined. It should be noted that liquid formulations may not be truly frozen at -20°C (despite appearances) and that freeze-thaw transitions occurring around this temperature may adversely affect stability

- Protection from light. This may be achieved by storage in dark facilities, use of amber or opaque containers or secondary packaging. The criticality of protection from light should be determined for each product-formulation combination.

The PRS is a key critical material and it is important that the volume of PRS be sufficient to minimize routine replacement. When considering the appropriate volumes, the following should be considered:

- Aliquot size. The aliquot size should be sufficient to run anticipated assays without unnecessary pooling of aliquots, storage of opened/thawed/reconstituted aliquots or wastage. Typically this quantity might be that required to perform one assay run, allowing for a 10 to 20% percent volume overage of the thawed or reconstituted reference standard. This overage allows for pipetting and robot delivery and losses during transfers.

- Selection of the number of years desired for the PRS. (i.e., 20, 25, 30 years etc.)

- Sufficient volume for long term stability testing (The ICH long term stability testing schedule is often used, however, many firms are not comfortable with annual testing of older standard and often have an increased testing schedule of quarterly or semi-annually.) Some companies and compendial agencies base reference standard reevaluation intervals on historical reference standard reevaluation data, typically with a maximum interval of 5 years between these reevaluations.

- Understanding of the volumes for qualifying working standards. The amount of potency testing is described in an SOP describing the WRS qualification program.

- An overage for the number of aliquots (for example, 30%) of the above volumes to account for qualification of new analytical methods, support of comparability studies to support manufacturing changes and other unforeseen studies.

Handling of the Working Reference Standard (WRS)

- The WRS can be a final product batch which has been selected as a representative batch from the current manufacturing stream. As such this material can be stored under the designated storage conditions for the approved shelf life. However, if desired, the WRS can be a specially manufactured standard which is stored in different single use aliquots. As such all the points made above, handling of the PRS, would apply to the specially filled WRS.

- The WRS can be stored at low temperature eg. -80°C as long as the appropriate freeze/thaw studies have been performed. It is not recommended that liquid formulations be stored at -20° as this temperature is at the transition between frozen and liquid unless there is data available demonstrating there is stability at this temperature. Appropriate formulations can prevent freezing at this temperature. Lyophilized materials can often be stored at -20°C.

- If aliquots of WRS are not stored in single use aliquots, then stability studies to support interim storage of excess WRS for later use should be undertaken

- Storage is often in the dark, in a container which protects it from light and/or closed secondary opaque container.

Volume of WRS

The WRS is a key critical manufacturing reagent and very resource intensive to select, handle and qualify. Companies usually try to balance shorter hold times to reduce unexpected stability problems and longer times to reduce manufacturing costs. When considering the appropriate volumes, the following should be considered:

- Aliquot size. This volume should be sufficient to run the number of assay runs plus a 10 to 20% percent volume overage. This overage minimizes air bubbles forming during the pipetting of the WRS
- Selection of the number of years desired for the WRS. (i.e., 2, 3, 5 years etc.) This number determined by considering the product stability and the frequency of manufacture.
- The number of manufacturing batches planned per year.
- The number of real time testing lots (typically a minimum of three per year, run using the ICH time schedule.)
- Additional volume should be included for real time testing if the product is not manufactured on a minimum of a monthly basis.
- A 30% overage of the above volumes to account for retesting requirements, qualification of new analytical methods, support of comparability studies to support manufacturing changes and other unforeseen studies

Value Assignment of the WRS from the PRS

The WRS batches must be derived from a representative drug substance batch. The potency assignment will be derived by comparison to the PRS. To assign the WRS potency the variability of the potency assay should be considered to determine the minimum sample size for the assessment. It is recommended when possible, the assignment of potency to the WRS include an evaluation of whether a resulting shift in potency would negatively impact the assessment of process capability.

It is recommended that the estimate of the mean relative potency be used to assign potency. The precision of the estimate should be evaluated by assessing the width of the confidence interval for the mean. In the case where the 95% confidence interval for the mean relative potency contains 100%, then the WRS is commonly assigned 100% relative potency. If the confidence interval does not contain 100% relative potency but the sample mean is within 95% RP to 105% relative potency, then the relative potency for the WRS is commonly assigned a potency as the sample mean. In the latter case, an assessment of the study results (and possibly additional testing) should be conducted to confirm an assignment of potency that is different from 100%. In addition, an alternative estimator to the sample mean may be considered in a situation, for example, of highly skewed results. (Note: There is a draft FDA guidance that suggests supporting a 100% assignment by showing that a two-sided 95% CI of the mean fits in a "sufficient narrow range (e.g., 90-110%)". Mathematically this approach is the same as it is written in this document to demonstrate that the mean value is within a 95 – 105% range AND that the two-sided 95% CI of the mean is no larger than +/- 5%.).

If acceptance criteria are not met, an investigation should be conducted to assess assay variability, bias, and other potential causes. Properly identified outliers may be removed from the analysis and / or additional potency results may be generated as deemed appropriate.

The WRS batches should be tested at periodic intervals as described in a formal stability program and all stability-indicating tests, bioassay and physico-chemical, should be run and compared to the PRS as well as being compared to historical values. The potency assignment will also be shown suitable for continued use by comparison to the PRS.

Assessing Stability

During early development and clinical manufacturing, determining the stability interval of a RS is done by estimating the initial stability and then setting a re-test date. During clinical trial manufacture we typically rely on early stability data to set an initial re-test date. This data includes a combination of tests, physicochemical and biological and then relies on probability for setting the target re-test dates. Clinical reference standard programs for biologics often target 5 years minimum at -80°C storage of liquid presentation.

Any change, even if it does not result in an apparent change in the biological activity of clinical interest, is a warning. All such changes need to be assessed to determine the impact, if any, upon the stability of the product. Testing is demanding of resources and

reference standards: companies developing products need to be aware of this and establish a realistic testing program. This program should include the monitoring of RS during routine use.

During early clinical development, data from product development, including forced degradation, is utilized to support stability assessments. Caution needs to be taken as the imprecision of a bioassay may not permit detection of change in activity over a short time period or mild stress conditions.

Stress studies can give some indication of the possible stability of the product and identify possible routes of degradation that will apply to the product under planned storage conditions. Degradation under stress conditions can provide early data but may not reflect the degradation that will occur under normal storage conditions. Real-time studies are necessary to demonstrate what happens under the actual storage conditions.

Common Routes of Degradation

When designing a stability program, it is informative to consider the potential routes of degradation and the sequence/3D structure of your product. Knowing if the product contains hotspots for these types of degradations will help to determine the type of stress studies which need to be done and the types of analytical methods needed to detect these changes. Common degradation pathways for protein products include:

- Deamidation – hydrolysis of Asn and Gln side chain amides
- Oxidation – of Met, His, Cys, Tyr and Trp residues Met, Trp are the typical ones
- Denaturation – loss of 3-D (tertiary) structure
- Aggregation – association of monomers or native multimers – covalent or non-covalent
- Carbohydrate content. Note that a common instability of glycosylation is the hydrolysis of sialic acid residues
- Fragmentation

Once the most likely routes of degradation for the product have been identified, appropriate analytical methods can be developed to detect changes in the product. The potency assay will always be included in this toolbox of methods: Potency in a bioassay is dependent on the integrity of many features of the product and so may demonstrate a change that is not anticipated or identified by the selection of other analytical techniques used. It must be noted that any given bioassay may not be sensitive to a particular change that could be of clinical significance.

Accelerated Degradation Studies

Stress studies (such as high and low pH, oxidation studies and deamidation studies) are performed on representative material prior to establishing routine protocols for routine stability testing of IRS, PRS and WRS. Although forced degradation may follow pathways different from those of degradation under normal storage conditions, they are especially informative when comparing various candidate RS batches. Each study is assayed with its appropriate physical/chemical analyses and the potency assay. This allows the correlation of degradation with impact on the biological activity.

Common forced degradation studies include:

- Thermal
- pH
- Oxidation
- Freeze-thaw
- Agitation, shear
- Light
- Specific solvent and surface interactions

It is important to note whether the stress degradation studies have the expected impact on the proposed RS to determine whether this is a representative batch. For example, if a proposed batch of WRS has a higher degree of degradation than that shown by other batches subjected to the same stress, then this batch is not a viable candidate as it is not a representative batch.

Accelerated Thermal Degradation Studies

Another common accelerated degradation study is accelerated thermal degradation. Extrapolation from rates of degradation observed at high temperatures can sometimes be used to predict degradation rate at lower (storage) temperatures. This can allow for the prediction of stability from short term data. The use of this approach has many caveats, which include the premise that the degradation reaction follows first order kinetics and that only a single degradation reaction is involved. This is often not the case for large biological molecules with complex 3D structure. Although there is no regulatory guidance for biologics for appropriate forced degradation, it can be useful for comparing various batches of reference source material. The prediction of the

Accelerated Thermal Degradation Studies Cont.

stability is commonly made using the Arrhenius equation (ln k = ln A – Ea /RT).

This approach to predict stability is subject to the following caveats.

- Degradation at higher temperatures may follow different pathways from those at lower temperatures. Consequently, rates of degradation predicted from higher temperatures (+45°C, +56 °C) can be greater than that predicted from the elevated temperatures (+4°C, +20°C, +37°C)
- Rates of degradation predicted from short storage times can be greater than those predicted from longer storage times. This can occur if, for example, there is an initial "one-off" non-continuous process, such as one consuming residual moisture
- Rates of degradation can decrease with time if degradation products inhibit degradation, or can increase if degradation products catalyze degradation.

Freeze-thaw studies are an important part of any stability study. It is especially important for reference standards which might be subjected to more than one freeze-thaw cycle. This most commonly occurs when a new batch of WRS is prepared from a larger frozen stock of material or when a single aliquot of WRS is larger than required for one assay and is refrozen for use in a later assay

Real Time Stability Studies

It is important to assess the stability of all reference standards (DRS, First IRS, Second or subsequent IRS, PRS and WRS) in real-time. However, since the potency assay is a relative potency assay, it relies upon the stability of a reference standard within the assay. Care must be taken to provide data which is not self-referential which will hide stability issues.

Strategies for establishing real-time stability testing include:

- Comparison with an external reference standard. This may be possible in the case of 2nd generation products, biosimilar products or products which have a suitable external reference of known stability characteristics. Points which need to be considered include:
  - Is the formulation of the external standard appropriate for use in your assay?
  - Are the dose-response curve characteristics similar to that of the product?
  - Is the stability of the external reference known?
- Use of a battery of bioassay tests to assess the stability of different potency-determining characteristics of the product. This is particularly important for products known to have multiple biological activities
- Physicochemical characterization data
  - Extended battery of tests (not just release tests)
- Accelerated degradation by stress. Many sponsors currently use DOE to select conditions for these studies.
  - Thermal (NB: assumes same path of degradation at different temperatures),
  - pH,
  - oxidative,
  - light,
- Comparison with similar/different standard stored in similar/different conditions which might include other batches, ultra-cold storage, different formulation, etc.
- Longer-term storage of similar standard
- Trending of dose-response curve parameters
- Monitor T=0 data for manufactured batches. If the manufacturing process is stable it can be assumed that these values should be fairly stable. This data can be used as a real time assessment of any changes to the reference. For example, if the T=0 data begins to show increasing potency over time, this indicates that the reference may be losing potency.

Each product will require a unique strategy for determining stability of the RS. Knowledge of the particular product, and possibly similar products, will permit selection of the battery of tests which are most likely to detect degradation of the RS.

Frequency of Real Time Testing

There is a regulatory guidance on real time stability testing which states the following: For long term studies, frequency of testing should be sufficient to establish the stability profile of the drug substance. For drug substances with a proposed re-test period of at least 12 months, the frequency of testing at the long-term storage condition should normally be every 3 months over the first year, every 6 months over the second year, and annually thereafter through the proposed re-test period.

Frequency of Real Time Testing Cont.

This is not necessarily the recommended time schedule for reference standard. Companies should exhibit care when decreasing the real-time testing requirements to biannually or annually for PRS or any RS which has an unknown stability profile. However, reevaluation experience with the IRS may be applicable and support starting with annual testing of the PRS and WRS. There needs to be balance between the risk of an undetected stability failure and the depletion of the stock of RS. This risk may also be reduced if the reference standard is stored at lower temperature and/or in a more protective container closure, such as flame-sealed ampoules with inert headspace stored at deep-frozen temperature.

Parallel Stability Studies

Potency of the RS (or candidate RS) can be compared with different preparations of similar types of material, including multiple product formulations, stored under similar conditions. However, in this approach there is a risk that the different materials may be undergoing similar degradation at similar rates. If comparisons are made between multiple materials, this risk is reduced as it is unlikely that multiple independently produced materials will all suffer from the same failure mode (e.g., comparison between the PRS, WRS, Control Sample, and routine production batches offers four materials that should all have consistent potency over time and would be highly unlikely to undergo the same rate of change over time).

If a suitable official, eg. international or national working standard, is available, this can be used to assess changes in potency. Such official reference standards will have undergone rigorous testing to determine their stability.

One potential problem with parallel stablity studies is the precision of the bioassay may not permit detection of small changes so a large number of assays would need to be performed to increase the precision of the potency measurement and detect small changes. However, if comparisons between multiple materials are made routinely (e.g., relative potency of a Control Sample versus the WRS, relative potency of the WRS versus the PRS, relative potency of production batches versus the WRS), a large amount of data can be accumulated over time, increasing the precision of the potency measurement and the ability to detect small changes.

Real Time Monitoring

Real time monitoring is a key tool for determining the stability of any RS. Typically, this involves a comparison of samples held at normal storage temperature with external reference or samples stored at ultra-low temperature (-150°C, -70°C). This type of study requires some care.

- It assumes the study reflects processes occurring at storage temperature
- It assumes the RS stored at the low temperature sample is stable
- Depending on the precision of bioassay and the small amplitude of changes, a large number of assays may be required.
- Vials in which the RS is stored may not be suitable for very low temperature
- If the RS is already stored at -70°C, this option might be unsuitable

Natural Reference

For some products there may be a natural reference material which can be utilized for monitoring a proposed reference standard. This can be pooled samples from a population where the activity is defined. An example is coagulation factors and inhibitors where 1 international unit = amount or activity in 1 mL of “average fresh normal plasma”

If a natural reference is available care must be taken to confirm:

- The variation between individuals and populations to determine if a pooled source is required.
- Monitor the variation with time across individual and pools. Often this depends upon lifestyle changes such as aspirin use, alcohol consumption, pregnancy, etc.

Predicting/Calculating Stability

- If the degradation products are completely inert in the bioassay, then the degraded material will be functionally similar to undegraded material, but with lower relative potency. The dose-response will be similar to that of the undegraded material but shifted along the concentration axis. A relative potency can be determined.

Predicting/Calculating Stability Cont.

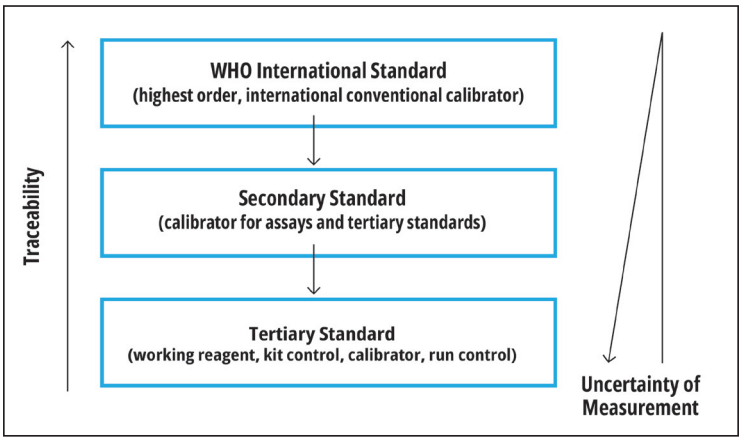
- If the degradation products are not inert: 1) the preparation may not be functionally similar to the original so no relative potency can be calculated; only a change of activity can be noted; 2) if the degradation product is functionally similar to the original, then the measured relative potency may not reflect the degree of degradation. Hence, it is important that for assessment of stability, potency data be used in conjunction with physicochemical data. Any change detected by any analytical technique must be assessed for its potential impact. It should also be noted that a change which does not affect one potency assay might affect the activity in a different potency assay.

Official Reference Standards (ORS) / External Reference Standards

There are various officially recognised sources of external reference standards for biological activity. Examples include Pharmacopoeia such as USP or PhEur, government agencies such as NIST or NIBSC and public entities such as WHO. Use of an external reference standard may be specified by a regulatory agency prior to approval and ICH Q6B states that "results of biological assays should be expressed in units of activity calibrated against an international or national reference standard, when available and appropriate for the assay utilised". This requirement has been particularly common for products which require pharmacopeial monograph potency tests.

External reference standards are not usually available in sufficient quantities for routine use. They typically have specified units of biological activity (e.g. most WHO International Standards are assigned international units, IU) and are intended to be used to calibrate the references used in routine assays. Although some sources may supply sufficient reference standard for routine product release this is becoming less common and companies should be establishing their own internal reference programs. An essential part of this program will be to compare the product to the required external reference. Similarity of the dose response curves must be demonstrated before the product reference standard can be value assigned using (and be traceable to) the external reference.

A consequence of value assignment against an external reference standard is that the measurement uncertainty in the value should also be assigned. Alternatively, the calibration exercise may be used to demonstrate that the measurement uncertainty is negligible in comparison to the precision of the lot release assay or less than some pre-justified maximum level, in which case the uncertainty is not formally assigned to the reference standard. It is likely that statistical guidance may be required regarding this estimation of uncertainty, although several published examples exist.



An important consideration for users of such standards is evaluating the impact when an existing external reference standard is replaced. Although attempts to ensure continuity of the external standard unitage are made during replacement, there remains a potential impact on the potency labelling of existing products. Companies should therefore have an internal program which monitors the external reference supply. When an existing standard is due to be replaced it may be beneficial to perform head-to-head studies with both external references and existing in-house reference standards in order to appropriately inform any corrective actions that may be required. In some cases, standards which appear similar to the external supplier will not be similar in a particular potency assay. If similarity in dose-response curve characteristics is not evident studies should be undertaken to determine if this is due to a formulation change or alteration of a specific biological activity.

Transitioning from an Older Reference Program to a Two-Tier Approach

In some cases, older established products may have been licensed without a two-tier (PRS and WRS) system in place. When a two-tier reference system is planned from the outset of the drug development program, ideally the PRS is selected from a PIII clinical trial batch. This means that the PRS has been

Transitioning from an Older Reference Program to a Two-Tier Approach Cont.

"clinically validated" in that it was included in the studies which demonstrated that the product was efficacious. However, it is sometimes the case that if no PIII batch is available, an engineering batch is used as the PRS. The PRS is a key material so it is important that the stock of PRS established is sufficient to minimize the need for replacement.

In an older reference standard program without a two-tier system, there will usually have been a series of WRS and no PRS. A series of WRS will generally have been established as replacement WRS were required due to exhaustion of stock, insufficient stability of the WRS or changes in the product which left the WRS unsuitable for continued use. This means that the link back to the product batches tested in the clinical trials is less direct than when a single, or small number of, PRS have been used.

The basic requirements and properties of the PRS will be as already discussed in the sections "General Principles and Considerations" and "Reference Standard to Release Commercial Product", but there will necessarily be procedural differences when introducing a PRS for an established product. The steps required to establish a two-tier system in a particular case will depend on the individual circumstances.

The PRS must be representative of the product as currently manufactured, as should be the WRS. The candidate PRS will require extensive testing to obtain a thorough characterization. A battery of physicochemical and bioassay analytical techniques should be applied, not simply the tests used for batch release. It may be appropriate to use additional techniques that were not originally applied to the characterization of the product as novel techniques may have been developed and knowledge of the product will have increased. The PRS should be selected as being towards the middle of the range for properties as measured by both bioassay and physicochemical analyses.

There is the advantage that at this time there will be extensive knowledge of the product, its critical quality attributes and its stability. The bioassay also will be well established. Formulation and storage requirements for a RS will be known, though the PRS may require longer term storage than may have been achieved for the WRS.

If an ORS has been used to assign potency to WRS batches, this will facilitate the establishment and potency assignment of the PRS.

Continuity must be ensured for release of the product batches, so selection of the PRS will require extensive comparison with the current WRS. Depending on how replacement WRS have been established and whether retention samples of previous WRS are available, candidate PRS should be tested against previous WRS as an additional precaution against any anomalies in the current WRS. Future WRS will be tested and potency assigned against the PRS.

Potency would normally be assigned on testing against the current WRS. This may be dependent also on testing against an ORS, if available and suitable.

If an assay control sample has been used routinely in bioassays and the batch meets all requirements including sufficient batch size, this may be a suitable candidate PRS for which there is already extensive data.

Similarly, if candidate replacement WRS have been tested by routine inclusion in assays, one of these may be a suitable candidate PRS.

As a general principle, the PRS and WRS should be derived from different production batches. Having preparations derived from two different batches can provide a measure of control to reveal anomalous behavior by one or the other, for example, problems with stability affecting one of the preparations

Biosimilar Products

Differences For Biosimilar Development

Although the requirement for a two-tier reference standard program is the same for innovator and biosimilar products, there are some distinct differences and challenges for biosimilar development. Difference include:

- The originator product is available to serve as a reference material until development of an in-house reference standard (IHRS). In this use, the originator product is referred to as the Reference Product (RP). This should not be confused with the biosimilar developer's reference standard, RS, to be used in laboratory testing to release clinical and commercial batches.
- Often there is a challenge in procuring sufficient number of lots of originator product that would be representative of manufacturing variability.
- Originator product can be sourced from different markets (e.g. EU, US, JP)

Biosimilar Products Cont.

- Development activities can be started based on one or several originator product batches (or a mixture). This reference material can be utilized until the first in-house reference standard (IHRS) is produced. To do this the following needs to be determined:
  - Determine the number of batches (appropriate statistical power), or batches to be pooled for the preliminary reference material.
  - Qualify the first IHRS against the RP as part of the demonstration that the IHRS is suitable. This means the assignment of potency value to the IHRS with a tight acceptance criterion: e.g. 95-105% potency against the preliminary reference material.

Although the existence of the originator product helps establish what will make an acceptable IHRS there are challenges faced by biosimilar developers:

- Often they have "immature" potency methods at the time of qualification and the precision and accuracy of these assays may cause problems.
- The IHRS might be from a manufacturing process which is not validated and might not be a representative lot of the final clinical standard.

BioSimilar Reference Specifics

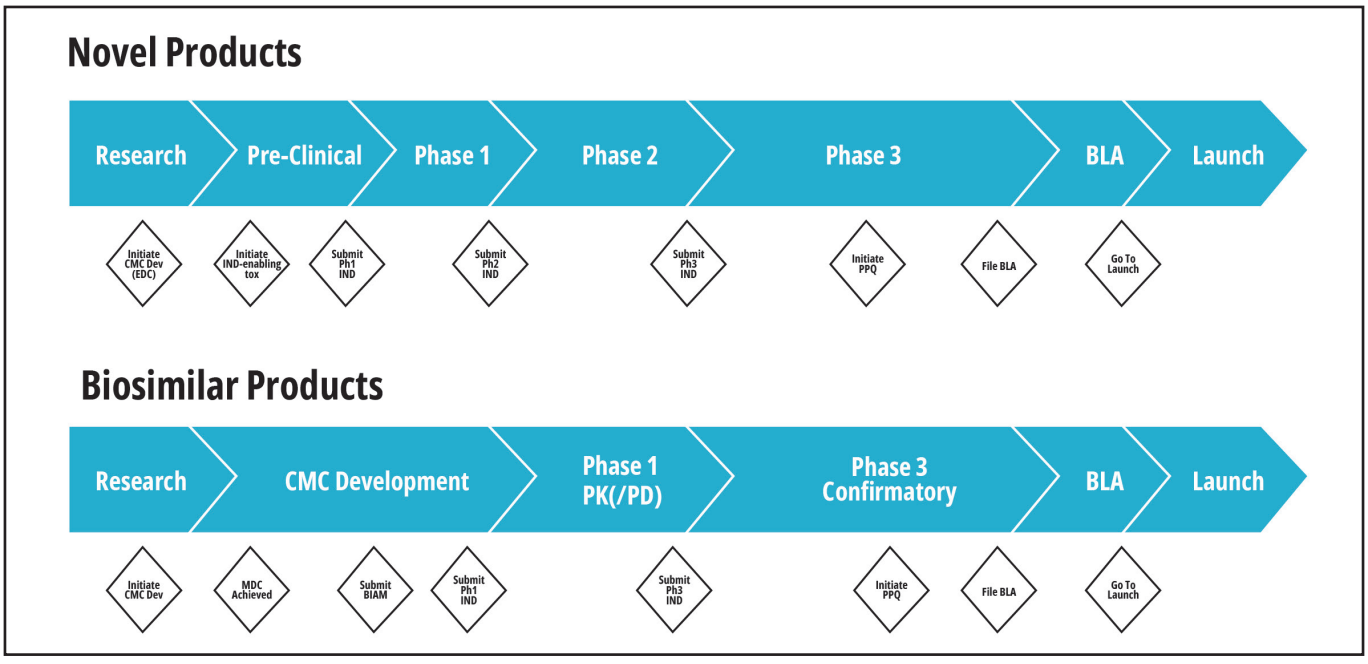
Similar to innovator product development, during biosimilar development: the IHRS is used to assess similarity, and to release clinical and commercial standards. In ideal cases all these data are generated against a single IHRS batch during entire development and commercial phases. This decreases the number of interim reference lots required during development.

An International Standard might be available or become available; there is an increasing need for harmonization of the IHRS with a public standard that can serve as a benchmark for the originator and all biosimilars

EMA's recommendation is to use an International Standard or PhEur standard as primary reference to characterize biologics (EMA/CHMP/BWP/534898/2008 Rev. 2). Sponsors can use in-house standards. However when and appropriate international or national standard is available it is recommended to calibrate in-house standards against it (ICH Q6B).

For biosimilars, several potency assays in addition to the release potency assay(s) is (are) performed. Several potency assays reflecting all known Mechanisms of Action (MoAs) are developed and run to ensure a complete similarity evaluation. This complete characterization of the biological activity is considered a key product development step. The IHRS is often the reference standard utilized for this complete characterization.

The main challenges for biosimilar development are very aggressive timelines for the drug development making it difficult to have methods which are qualified and validated in time for the registration. Very often a short P1 leads directly to PII/PIII where the manufacturing process is validated and for which all supporting methods need to be validated. This means that the choice and qualification of the IHRS occurs early and this can result in some problems at later stages.



Acknowledgement

The authors thank Patricia Koutz, BEBPA Speaker Coordinator, for recording a summary of the discussions at meetings of the BEBPA Reference Standards Focus Group and for collating the texts contributed by the authors to this paper, and Ryann Engel, BEBPA Head of Outreach, for designing the layout of the final document.

APPENDICES

Appendix 1. Terminology

Assay Control Sample (ACS): a sample prepared and tested in the same fashion as the test article(s)/ unknown sample(s) and Reference Standard. The ACS should be included in every assay, ideally on every plate in a multi-plate assay. The source of the ACS should be different from the source of the Reference Standard (eg. from a different lot).

Assay Run (or Unit): The execution of the SOP that is completed by a laboratory (eg. by one analyst) in a time frame determined by the method (eg. one day) with a single preparation of assay standards, samples and controls.

Assigned Value is an accepted potency value, normally assigned by a national or international organization to an official reference standard based on their analyses, and can differ from True Value due to the presence of method variability and/or less than perfect accuracy.

Clinically Validated: A reference standard which has been used during clinical trials to determine product efficacy. It is often the lot used as the primary reference standard.

Correction Factor: A unitless factor applied to adjust relative potency determinations on change to a new reference standard or change in potency of an existing reference standard over time. Use of correction factors is not generally recommended.

Development Reference Standard, also known as Method Development Reference Standard or Research Reference Standard: a standard utilized early on during the development of a potency assay, typically a lot that is obtained from the research group, and may or may not be used to release clinical material. If it is not used to release clinical material it may transition into the first control sample.

Direct comparison: a comparison made by inclusion of preparations in the same assay run. Direct comparison is typically used to define the potency of a new candidate reference standard against a Primary RS or a previous Interim RS that is being replaced.

Drift, reference standard: the change over time that results in a trend of increasing or decreasing potency test results when using one or more batches of reference standard. Drift can be caused my many situations, including the following:

- Degradation of a reference standard that decreases in potency without a corresponding re-assignment of the Defined Potency. Over time, this will cause test results when using the reference standard to increase.
- Accumulated shifts in the relative potency from one batch to the next. For example, if reference standard batches are replaced by chaining one batch to the next by direct comparison, and the bioassay being used has a negative bias, each replacement standard will have a lower measured potency than the previous even if the standards being used actually have the same true potency. Over time, the reference standards will become less and less accurately assigned causing a drift in test results.

External Reference Standard: see Official Reference Standard

Forced Degradation is the degradation of a drug product and/or drug substance by chemical, heat, light, etc., more stressful than typically occurs during accelerated degradation conditions. Forced Degradation is required to demonstrate attribute functionality of stability indicating methods (ex. bioassay) and provides evidence of the degradation pathways and products of the drug substance.

Higher-order RS: A reference standard that is used to evaluate subordinate reference standards. Primary and Compendial reference standards are examples of higher-order standards.

In-House Reference Standard: A reference standard that is used within a company, in contrast to an external or official reference standard. The In-House Reference Standard may be a Development Reference Standard, Interim Reference Standard, Primary Reference Standard, etc.

Interim Reference Standard: A reference standard that is suitably characterized for use and prepared from a non-GMP or GMP representative clinical lots or from material used for quality control purposes during a product's development stage. The establishment of the Interim Reference Standard is based upon appropriate characterization, inherent to the biologic, and not in comparison to an Official Reference Standard or Primary Reference Standard.

Multifaceted approach: The use of multiple sources of evidence to demonstrate a property. The multifaceted approach for demonstrating stability is particularly important in cases where there is no higher-order reference standard with which to compare. No single piece of evidence is sufficient to prove that biological potency is not changing.

Official Reference Standard: (sometimes called External Reference Standard) are defined by ICH Q7 as being from an "officially-recognized source" by regulatory authorities. Examples are Compendia such as USP or PhEur, government agencies such as NIST and NIBSC and public entities such as the WHO. The term 'Reference Standard' is reserved by some organizations to refer to Official Reference Standards.

Pooled Batches: the result of pooling two or more lots of material; in this paper, for Reference Standard production, to ensure sufficient volume or representative attributes that may vary from lot-to-lot.

Primary Reference: or 'Reference Standard, Primary' per ICH Q7 is defined as "a substance that has been shown by an extensive set of analytical tests to be an authentic standard that should be of high purity. This standard can be: (1) obtained from an officially recognized source, (2) prepared by independent synthesis, (3) obtained from existing production material of high purity, or (4) prepared by further purification of existing production standard." ICH 6B is more specific for biologics; instead of being derived from a ultra-purified small molecule-like standard, "an appropriately characterized in-house primary reference standard [should be] prepared from lot(s) representative of the production and clinical materials."

Reported Value or Reportable Value: The relative potency estimate derived from different combination calculation methods for independent assay results. The bioassay Standard Operating Procedure (SOP) defines the derivation of the Reportable Value. The Out of Specification for the SOP is based on the Reportable Value.

Reference Material: Term used by some organizations to denote various reference standards other than Official Reference Standards.

Research Reference Standard: see Development Reference Standard

Shift, reference standard: Reference standard shift describes a one-time change in relative potency that can occur when a reference standard is replaced with a new batch. Shift is typically caused when the measurement uncertainty associated with the defined potency is large and/or the number of replicate analyses is too small. If the replacement reference standard has a different relative potency compared to the existing reference standard, a shift in test results will be observed.

Secondary Reference Standard: often referred to as the 'Working Reference Standard' is a suitably characterized standard prepared from representative cGMP clinical or commercial lots(s) to support relatively longer-term predictable quality control testing of product lots for release and stability. The Secondary Reference is typically established by comprehensive comparative physicochemical and biological assay analysis against a Primary Reference Standard.

Standard Operating Procedure (SOP) or Method is a complete step-by-step and highly detailed description of the process steps required to perform predicable assay runs. SOPs in general are a company's business approach to documenting processes. SOPs help achieve higher communication standards, performance and quality uniformity, regulatory compliance and increased efficiency.

True potency: The actual potency of a standard which is unknown but can be estimated by averaging replicate analyses. The true potency could only be known using a "perfect" method (perfectly accurate (no bias) and no variability).

Two-Tiered Reference Strategy includes both a Primary Reference Standard and a Secondary Reference Standard. According to ICH Q6B "In-house working reference standard(s) used in the testing of production lots should be calibrated against primary reference standard." The implementation of a Two-Tiered Reference Strategy with an appropriate Reference qualification SOP when transitioning from one Reference to another will minimize method drift over the lifecycle of a biopharmaceutical program.

Working Reference Standard see Secondary Reference Standard

Appendix 2. Sample Size Calculation

2.1. Sample size calculation for value-assigning potency

While advanced methodologies such as MCMC (Markov chain Monte Carlo) simulations are optimal to calculate the minimum required sample size to value assign the potency of an IRS, the below formula will provide a reasonable starting point:

$N_{min} = [Z(\alpha/2) * \sigma / E]^2$

where

- $Z(\alpha/2)$ is the critical value of the standard normal distribution at the desired level of confidence (e.g. 1.96 for a 95% confidence level), with $(1-\alpha)$ being the confidence level.
- σ is the standard deviation of the log-potency
- E is the desired margin of error (MOE) in the log scale.
- Nmin is the minimum sample size

Note that this formula will usually not provide an integer number. As this is the minimum required, it is advised to round the value up regardless of the result. For example, if Nmin=5.03, the sample size should be 6 rather than 5.

2. Sample size calculation for equivalence testing

Once again, advanced methodologies may be preferred to simulate different scenarios when a statistician experienced in bioassays is available. However, in the absence of a statistician, the following formula can be applied to estimate the required sample size to demonstrate equivalence between the potency of two reference standards. This is the formula for equivalence tests when the difference is assumed to be zero (Julious SA, 2004).

$N_{min} = 2 * [(Z(\alpha/2) + Z(\beta)) * \sigma / \Delta]^2$

where

- $Z(\alpha/2)$ is the critical value of the standard normal distribution at the desired level of confidence (e.g. 1.96 for a 95% confidence level), with $(1-\alpha)$ being the confidence level.
- $Z(\beta)$is the critical value of the standard normal distribution at β level of power (which is the minimum probability of success of the test in case of true equivalence).
- σ is the standard deviation of the log-potency.
- $\pm\Delta$ is the equivalence margin.
- Nmin is the minimum sample size.

Appendix 3. References and Further Reading

Regulatory, guidance and recommendations

1. ICH Q6B: Specifications: Test Procedures and Acceptance Criteria for Biotechnological/ Biological Products. March 1999. https://database.ich.org/sites/default/files/Q6B%20Guideline.pdf

2. ICH Q1A(R2): Stability Testing of New Drug Substances and Products. February 2003 https://database.ich.org/sites/default/files/Q1A%28R2%29%20Guideline.pdf

3. FDA Guidance for Industry: Q7 Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients. September 2016. Revision 1. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073497.pdf

4. USP 35 General Chapter <1010> Analytical Data – Interpretation and Treatment. December 1, 2012 http://www.triphasepharmasolutions.com/Private/USP%201010%20ANALYTICAL%20DATA-INTERPRETATION%20AND%20TREATMENT.pdf

5. USP 32 General Notices and Requirements: 5.50.10. Units of Potency (Biological). May 1, 2009 https://www.uspnf.com/sites/default/files/usp_pdf/EN/USPNF/generalNoticesandRequirementsFinal.pdf

6. USP <1034> Analysis of Biological Assays, August 1, 2012 https://www.drugfuture.com/Pharmacopoeia/usp35/PDF/5186-5201%20%5b1034%5d%20Analysis%20of%20Biological%20Assays.pdf

7. Official USP Reference Standards https://www.usp.org/reference-standards

8. FDA Guidance for Industry: INDs for Phase 2 and Phase 3 Studies Chemistry, Manufacturing, and Controls Information. May 2003 https://www.fda.gov/downloads/Drugs/Guidances/ucm070567.pdf

9. FDA Guidance for Industry: Bioanalytical Method Validation. May 2018 https://www.fda.gov/downloads/drugs/guidances/ucm070107.Pdf

10. FDA Guidance for Industry: Analytical Procedures and Methods Validation for Drugs and Biologics. July 2015 http://www.gmp-compliance.org/guidemgr/files/UCM386366.PDF

11. FDA Guidance for Industry: Referencing Approved Drug Products in ANDA Submissions. Draft. January 2017 https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM536962.pdf

12. EMA Guideline for bioanalytical method validation. 21 July 2011. Effective 1 February 2012 EMEA/CHMP/EWP/192217/2009 https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf

13. EMA Guideline on the requirements for quality documentation concerning biological investigational medicinal products in clinical trials. 27 January 2022. EMA/CHMP/BWP/534898/2008 Rev. 2 Committee for Medicinal Products for Human Use (CHMP) https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-requirements-quality-documentation-concerning-biological-investigational-medicinal-products-clinical-trials-revision-2_en.pdf

14. European Pharmacopoeia 7.0 - Reference Standards. 01/2008 http://www.fptl.ru/biblioteka/standartnye-obrazcy/EP-7_Reference-standards.pdf

15. Guideline for bioanalytical method validation ANVISA's Bioanalytical Guidance RDC 27/2012. May 2012 http://www.e-b-f.eu/wp-content/uploads/2018/05/bcn2012-S60.-3_tavares.pdf

16. Draft Guideline on Bioanalytical Method Validation in Pharmaceutical Development. 15 April 2013, MHLW, Japan http://www.nihs.go.jp/drug/BMV/BMV_draft_130415_E.pdf

17. Recommendations and Best Practices for Reference Standards and Reagents Used in Bioanalytical Method Validation. Bower JF, McClung JB, Watson C, Osumi T, Pastre K. The AAPS Journal 2014 Mar, 16(2): 352-356 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3933579/

18. Potency assignment of biotherapeutic reference standards. Faya P, Borer MW, Griffiths KL, Parekh BS. (2020) J Pharm Biomed Anal 191:113577

19. Recommendations for the preparation, characterization and establishment of international and other biological reference standards WHO TRS No 932 Annex 2, 2004 https://www.who.int/publications/m/item/Annex2-trs932

20. Reference Standards for Therapeutic Proteins: Current Regulatory and Scientific Best Practices (Part 1). Mire-Sluis A, Ritter N., Cherney, B., Schmalzing, D., and Blumel M. BioProcess Intl., March 1, 2014 https://bioprocessintl.com/upstream-processing/assays/reference-standards-for-therapeutic-proteins-350518/

21. Reference Standards for Therapeutic Proteins (Part 2). Mire-Sluis, A. BioProcess Intl. May 1, 2014 https://bioprocessintl.com/2014/reference-standards-for-therapeutic-proteins-351584/

(NOTE: References 16 & 17: CASSS CMC Strategy Forum "Reference Standards for Therapeutic Proteins: Current Regulatory and Scientific Best Practices and Remaining Needs," Gaithersburg, MD, on 15–16 July 2013, summarized by Anthony Mire-Sluis, Nadine Ritter, Barry Cherney, Dieter Schmalzing & Markus Blümel)

22. Sample sizes for clinical trials with normal data. Julious SA. Stat Med. 2004 Jun 30;23(12):1921-86.

International Reference Standards

23. WHO manual for the establishment of national and other secondary standards for vaccines https://apps.who.int/iris/handle/10665/70669

24. WHO Manual for the Preparation of Secondary Reference Standards for in Vitro Diagnostic Assays Designed for Infectious Disease Nucleic Acid or Antigen Detection: Calibration to WHO International Standards. https://apps.who.int/medicinedocs/en/m/abstract/Js23325en/

25. World Health Organization, Expert Committee on Biological Standardization. Fifty fifth report. Recommendations for the preparation, characterization and establishment of international and other biological reference standards. WHO Technical Report Series. 932:73– 130, 2006. http://apps.who.int/iris/bitstream/10665/43278/1/WHO_TRS_932_eng.pdf

26. World Health Organization, Expert Committee on Biological Standardization. Sixty sixth report. Report of a WHO informal consultation on international standards for biotherapeutic products. WHO Technical Report series 999:13–15, 2016. http://apps.who.int/iris/bitstream/10665/208900/1/9789240695634-eng.pdf

27. WHO International Biological Reference Preparations. http://www.who.int/biologicals/reference_preparations

28. Establishment of the first WHO International Standard for etanercept, a TNF receptor II Fc fusion protein: Report of an international collaborative study. Wadhwa M, Bird C, Dilger P, Rigsby P, Jia H, Gross MEB; participants of the study. J Immunol Methods. 2017 Aug;447:14–22. https://www.ncbi.nlm.nih.gov/pubmed/28288790

29. International standards for monoclonal antibodies to support pre- and post-marketing product consistency: Evaluation of a candidate international standard for the bioactivities of rituximab. Prior S, Hufton SE, Fox B, Dougall T, Rigsby P, Bristow A; participants of the study. MAbs. 2018 Jan;10(1):129-142. https://www.ncbi.nlm.nih.gov/pubmed/28985159

Reference Standard Stability

30. Differential scanning fluorimetry: rapid screening of formulations that promote the stability of reference preparations. Malik K, Matejtschuk P, Thelwell C, Burns C. J Pharm Biomed Anal. 2013 Apr 15;77:163-6. https://www.ncbi.nlm.nih.gov/pubmed/23416371

31. A comparison of vials with ampoules for the storage of biological reference standards. Matejtschuk P, Rafiq S, Johnes S, Gaines Das R. Biologicals. 2005 Jun;33(2):63-70. https://www.ncbi.nlm.nih.gov/pubmed/15939283

32. Rapid optimization of protein freeze-drying formulations using ultra scale-down and factorial design of experiment in microplates. Grant Y, Matejtschuk P, Dalby PA. Biotechnol Bioeng. 2009 Dec 1;104(5):957-64. https://www.ncbi.nlm.nih.gov/pubmed/19530082

33. The International Reference Preparation of Tetracosactide for Bioassay: characterization and estimation of its (1-24)corticotrophin-tetracosapeptide content by physicochemical and biological methods. Storring PL, Witthaus G, Gaines Das RE, Stamm W. J Endocrinol. (1984) 100(1):51-60. *(Note:"Bioassay estimates of samples ... which had undergone significant degradation were higher than estimates by HPLC.")* https://www.ncbi.nlm.nih.gov/pubmed/6317783

34. Isomerization of a single aspartyl residue of anti-epidermal growth factor receptor immunoglobulin gamma2 antibody highlights the role avidity plays in antibody activity. Rehder DS, Chelius D, McAuley A, Dillon TM, Xiao G, Crouse-Zeineddini J, Vardanyan L, Perico N, Mukku V, Brems DN, Matsumura M, Bondarenko PV. (2008) Biochemistry. 47(8):2518-30. https://www.ncbi.nlm.nih.gov/pubmed/18232715

    (Note: "Potency change not reflected in small selection p/c analyses")

35. Impact of residual moisture and formulation on Factor VIII and Factor V recovery in lyophilized plasma reference standards. Hubbard A, Bevan S, Matejtschuk P. (2007) Anal Bioanal Chem. 387(7):2503–2507. https://link.springer.com/article/10.1007/s00216-006-0855-x#citeas

36. Thermal stability of the WHO international standard of interferon alpha 2b (IFN-alpha 2b): application of new reporter gene assay for IFN-alpha 2b potency determinations. Caserman S, Menart V, Gaines Das R, Williams S, Meager A. (2007) J Immunol Methods. 319(1-2):6-12. *(Note: Nine years storage IS for IFN-α2b (95/566)* https://www.ncbi.nlm.nih.gov/pubmed/17196611

37. Implications for the assay and biological activity of interleukin-8: results of a WHO international collaborative study. (1997) Mire-Sluis AR, Gaines Das R, Thorpe R. J Immunol Methods 200(1-2):1-16. https://www.sciencedirect.com/science/article/pii/S0022175996001573

38. Applications of Freezing and Freeze-Drying in Pharmaceutical Formulations. Izutsu KI (2018) Adv Exp Med Biol. 1081:371-383

39. Freeze-Drying of L-Arginine/Sucrose-Based Protein Formulations, Part 2: Optimization of Formulation Design and Freeze-Drying Process Conditions for an L-Arginine Chloride-Based Protein Formulation System. Stärtzel P, Gieseler H, Gieseler M, Abdul-Fattah AM, Adler M, Mahler HC, Goldbach P. (2015) J Pharm Sci. 12:4241-4256

Appendix 4. Contributors' Email Contact Details

Dorota Bulik ..... DBulik@ultragenyx.com

Sian Estdale..... sianestdale@icloud.com

Zeban Kolen ..... Zeban.kolen@byondis.com

Marie Gottar-Guillier ..... marie.gottar-guillier@novartis.com

Laureen Little..... Laureen.Little@bebpa.org

Mike Merges ..... M.Merges@autolus.com

Peter Rigsby..... peter.rigsby@mhra.gov.uk

Jane Robinson..... Jane.Robinson@bebpa.org

Mike Sadick..... msadick@imugene.com

Perceval Sondag..... pzsg@novonordisk.com