

Table 14: Comparability Concepts and Case Studies

Facilitators –

Karen Sitney, *AbbVie Inc.*

Sonia Taktak, *Pfizer, Inc.*

Scope:

Manufacturers must ensure that the quality, safety, and efficacy of drug product are not impacted by changes to the manufacturing process throughout the product lifecycle by demonstration of comparability. In today's landscape where many promising therapeutics are eligible for accelerated regulatory pathways, the level of product and process knowledge to evaluate process changes may be limited. In addition, novel modalities and advanced therapeutics present additional challenges for defining CQAs and demonstrating comparability. This roundtable session will discuss comparability concepts used across the industry and will identify instructive case studies.

Questions for Discussion:

1. How does your company approach the risk assessment for manufacturing changes? When do you utilize internally approved comparability protocols, post-approval or always?
2. For which types of programs and at what phase do you typically seek agency feedback about comparability plans, if at all? What feedback has been received?
3. What approaches have you used for setting comparability criteria? When are statistically derived criteria appropriate? Do you apply sampling, or is data set typically first 3 post-change batches? How have you dealt with limited supply of pre-change material in terms of comparability assessments, particularly in the context of C>?
4. What stability data do you include in comparability assessment? Do you use GMP stability data, and if so in what context? When do you perform side-by-side forced degradation studies?
5. For what types of changes if any have you used clinical PK studies to mitigate risk/remove residual doubt?

To support this discussion, please provide an example for a late phase manufacturing change with challenges (setting of appropriate comparability criteria, differences in one or more QA, differences in stability profiles (accel/stress) for pre- and post-change material)? Would like to

identify at least one case study each for a conventional mAb, an accelerated program and an advanced therapeutic.

Discussion Notes:

January 25 and 27, February 2 and 4, *combined* –

1. How does your company approach the risk assessment for manufacturing changes? When do you utilize internally approved comparability protocols, post-approval or always?

- **Risk assessment process**
 - Risk assessment is a formal process. Technical team (analytical team) assesses product quality impact of the change.
 - Comparability is part of every CMC conversation especially in the Cell and Gene Therapy area. We use risk assessment for all changes.
 - As part of our change control system, the change owner must do a risk assessment prior to development of comparability protocol.
 - Different sites have different risk assessment approaches. Difference in development compared to commercial, release and stability. acceptance criteria, tighter than release.
 - Risk Assessment is always done.
 - We have a streamlined approach to risk assessment and a review council.
 - We would have Risk Assessment and comparability protocols. May look slightly different depending on stage.
 - We always have risk assessment. For Commercial post-approval changes (PAC) we use an FMEA type approach. The number of changes and type of changes dictate scope of comparability package, e.g., site change vs. process changes.
 - Risk assessment is very much a part of development and always part of evaluating any change. Process risks and analytical risks are evaluated to develop a grid for the comparability package to be developed. e.g. site to site ==> full package of release characterization and forced deg; for a like-to-like change ==> slimmer package proposed.
 - Comparability is a collaborative assessment of process changes with cross functional analysis for changes across functions. Each change is assessed for potential impact on CQAs, and prior knowledge is used to further support the risk assessment. We have developed a template and tailor it to fit each case.

- For late stage/commercial product, a risk assessment is always used. We are starting to use prospective comparability protocols more and more now.
- The approach to the risk assessment depends on the type of change and what phase. At phase 3, we use a full FMEA approach, it is internally reviewed and agreed, and we would discuss sufficiency of the comparability protocol with FDA if major change.
- We use the risk assessment to set up and plan the comparability protocol, either during development or post-approval.
- It is a sliding scale rating severity (minor, moderate, major) versus phase (pre-proof of concept, pivotal study, etc.). This may not be a formal process if the changes are minor, at later phase this will be formal. From mid-phase thereafter we would engage with regulators, discuss planned changes.
- Assessment up front to ensure bases all covered.
- We used a templated comparability protocol, not always a risk assessment.
- We leverage the risk assessment, considering stage and type of change. For major changes, we do a full-scale study with release data, extended characterization, forced degradation. For minor changes, this would be streamlined.
- Would do risk assessment for impacts to both DS and DP, even with only a DS change.

- **Internal protocol**

- We use formal comparability protocols in early and late stages, but use another term (e.g. "testing plans"). These are not submitted to HA.
 - Written protocol for comparability studies, based on learnings from the past.
 - Comparability protocols are not used all the time. Looking right now on how to use CP in Gene Therapy products.
 - Typically use a generic comparability protocol without formal risk assessment.
 - Company has experience with site changes. Comparability protocol signed by QA.
 - Internal development comparability protocol is separate from submission comparability protocol.
 - Different if technical development (early development vs late phase). In technical development, previous approach is no protocol, then discuss the data in context. Currently, recommend using agreed upon protocol using assessment criteria (not acceptance criteria) due to regulatory feedback.

- In development share protocols with regulators. Include as part of original BLA. Biologic co-formulated product. Changes that are planned once BLA approved are shared with regulator at pre-BLA.
- Must have protocol approved before PPQ start, possibly review by regulatory agency depends on the change.
- Internal protocol in development and commercial. Depends on complexity of material Example: BLA was accelerated and moving fast. We prepared a few comparability protocols for changes to be made post approval and consulted with agency. This was for biosimilar programs.
- CRO - for biosimilars, sponsors generate some initial data and use it to write protocol, then engage with HA.
- Comparability protocol needs to be approved prior to PPQ. Pre-approval of protocol by HA depends on the change.
- We would have risk assessment and internal protocols. Could occasionally have pre-engagement meetings .
- We would have risk assessment and comparability protocol. Engagement with agency depends on program acceleration and risk.
- We perform the comparability protocol, with the type of acceptance criteria dependent on phase and amount of available data. A risk assessment is applied retrospectively if differences are observed.
- **Major DS changes - how do you strengthen your comparability approaches?**
 - Vaccines - would use orthogonal characterization methods
 - Characterization - sometimes ADCC
 - Strengthen characterization package, but have justification for why certain methods are/are not intended for routine release testing
 - For major changes, repeat forced deg studies
 - Side-by-side characterization
 - Focus on no change to attributes affecting mechanism of action

2. For which types of programs and at what phase do you typically seek agency feedback about comparability plans, if at all? What feedback has been received?

- **General approach to requesting authority feedback**

- Authority engagement on comparability reserved for high profile changes. We take the opportunity to engage as we want the late-stage development to run as smoothly as possible. Proposed comparability protocol is typically the basis of discussion, as we usually have this before data are available. We usually approach this as a stand-alone CMC meeting, but often discussion needs cross functional input from clinical/medical safety. If we make a statement that criteria are clinically relevant, they can support.
- We plan a CMC-driven interaction to discuss comparability acceptance criteria, but also specifications, tests. It is typical for the Agency to agree in principle but say “it will be a review issue”. PACMPs are also negotiated with Agency meeting, e.g., proposed to not include HCP, DNA, Protein A clearance as part of PACMP. Data supported removal of these from the comparability panel, and that was approved.
- We have asked agencies multiple times in development. The feedback is often that “it's a review issue”. Also now we often use a protocol with “assessment criteria” rather than acceptance criteria, because of the limited amount of data, and we discuss the results in terms of impact on safety and efficacy if they are outside of the assessment criteria in the report. Assessment criteria are usually based on platform knowledge (we have a database and make use of prior knowledge). If we are forced to, we will define tests, number of pre-change batches.
- Negotiation and open line of communication is encouraged. I am willing to be a pest with the FDA project manager to ensure this.
- Would present the protocol at a Type C meeting. Recent example for an early but post-proof of concept study, we proposed full release data and extended characterization only. We had done forced degradation, but this was not discussed (back pocket data in case of questions). Feedback was that the protocol was acceptable, but FDA requested study on Fc effector function. Although this was an IgG2, not expected to have effector function, FDA wanted this demonstrated in context of the comparability study.
- Unless we have a simple change or one for which we have significant platform experience, we will seek feedback. We generally do not do forced deg upfront, but wait for agency request. We would try to negotiate this out. For major changes we would try to have 3 pre- and 3-post change lots, release and characterization panel with head-to-head testing as appropriate, accelerated condition stability data. What is the rationale to not provide forced deg? Risk of differences, not relevant but hard to explain away.

- Had a successful interaction for a late stage program that had gone dormant. We had to compare back to 10 yr old material which was done with a comparability protocol pre-discussed with FDA.
- PPQ batches DS and DP planned 2 weeks apart. 2 month accelerated data on first PPQ batch (T=0 on 2 batches and 1 batch accelerated). Agency had no questions about proposed change. Was there a commitment to provide stability data? Additional stability will be provided during review of the CBE 30 upon request.
- HA feedback is helpful if you are seeking a bracketing approach.
- Regarding submissions, Agency will sometimes ask for protocols, raw data - seeking to understand and feel comfortable that there is no impact to product quality.
- **Global submissions**
 - In commercial space, for significant change, strategically select subset of countries to solicit advice.
 - Upfront consider what requirements are global. Engage early with HA.
 - How do you handle comparability requirements across global markets? May assess which markets might be the most challenging and engage those HA.
 - Noted some differences on this topic regarding China expectations. Experience with China requiring in-country testing, material sourcing, etc.
 - For process changes during development we would request feedback from a variety of agencies (not just FDA); agencies are not always aligned, would try to have a globally acceptable comparability protocol.
 - For process changes we in all cases ask for feedback on the strategy. We seek advice from multiple agencies, it is interesting what different things the different agencies pick up on. Often we have requests for MoA directed assays as part of extended characterization if not part of release. We have reached understanding not to have quantitative criteria early on, with expectation that we would have quantitative criteria later.
- **Canada specific interactions**
 - For Health Canada, a common question is what bracketing approach to use? They like to see the data and approach. Have discussed rolling submission, stability plans and forced degradation. May be able to leverage development data. Container closure (feedback on testing).

- Health Canada have received feedback that forced degradation, extended characterization can be helpful. Looking for rationale for comparability (e.g. how criteria are set), also interested in consistency in method transfer between sites.
 - Health Canada encourages engagement on comparability approach, we have pursued engagement particularly for changes to marketed product and sometimes on clinical programs (but should have some data to discuss).
3. What approaches have you used for setting comparability criteria? When are statistically derived criteria appropriate? Do you apply sampling, or is data set typically first 3 post-change batches? How have you dealt with limited supply of pre-change material in terms of comparability assessments, particularly in the context of C>?
- **Comparability criteria**
 - Assessment criteria versus acceptance criteria
 - In comparability Protocol: we used terms ‘Comparability Acceptance Limits and Comparability Alert Levels.’
 - With stress condition, how to compare? If try to do statistics, may be constrained. Look at slope of degradation instead.
 - Have had feedback to refer to EMA Reflection paper on statistical methodology. Have used qualitative criteria/descriptive statistics in one case all the way to BLA. Visual presentation of the data and visual presentation of chromatograms can be very persuasive.
 - Always use numerical criteria, release criteria or tighter. Agencies will state that release criteria are not appropriate, but for some attributes we have had this accepted.
 - We typically use the specification acceptance criteria for qualitative attributes and tolerance interval for quantitative. We have had some experience applying equivalence tests. All of this is for a legacy product where we had the luxury of much data pre-change. This can help if you have few post-change batches.
 - The release specification is a component of our comparability. We generally can’t create statistically meaningful criteria for characterization tests. If there are differences, we would explain if they were meaningful (e.g., data, literature to justify). Pre-BLA we generally do not use statistically derived criteria. For post-approval might use a confidence interval or set clinically relevant limits. For example, we had a high HCP limit, but this was set based on maximum clinical exposure.

- Early on with less manufacturing experience, we would not use statistics. Later on we will use statistically derived criteria, but as internal criteria. We find the agency will want to make their own assessment regardless of the criteria, and even if they have stated in feedback that the criteria appear to be appropriate. If there are formal criteria they are included in the submission, but we would still justify differences in certain cases.
- **Number of batches**
 - Based on 125 batches
 - Finishing up 1 round of comp study. 2 pre-change and post-change (1 Eng and 2 GMP). 95% prediction interval for pre change. If new lot fall in the range it is accepted.
 - It is acceptable to use Eng. batch if limited supply
 - CRO - some sponsors use entire manufacturing history
 - Use of statistics to set criteria - some use at least 30 batches as "rule of thumb"
 - Minimum number of batches depends on available data
 - Can be a challenge in early development
 - Generally acceptable to use what you have (in the case of early clinical programs), come back and reassess with more data if necessary
 - For statistical tools we have used confidence intervals, tolerance intervals - it depends on the number of batches you have. With one or two batches, we would not set numerical criteria. We want to be in a position that if a change is observed for a non-CQA, we could discuss and justify.
 - We will use a mix of clinical and engineering pre-PPQ lots (process locked) to support comparability. This helps to add some variability and keep from setting too tight criteria. The lots which have been used in the clinic are thus bridged to the commercial process.
 - If we have limited number of lots, we include engineering and non-GMP tox batches.
 - For commercial products, we may use all batches, or may select which batches to use. Batches would need to span the lifetime of the process, to capture the all broad spectrum of variability over time.
- **Accelerated programs**

- In the case of accelerated development, we have a project that is moving very fast. We are trying to engage with agencies to de-risk the approach. We propose to perform a limited comparability study based just on release data as we cannot wait for characterization data of the post-change batch. Advice from the table was to combine release data with chromatograms from release assays, as well as any characterization data available. As the batch release criteria are often very broad, for comparability you will need discussion of the real data, chromatogram overlays, beyond the usual GMP evaluation of the release results.
- For cases with limited data, we do extensive characterization. In cases of extreme acceleration, plans can be confounded by changes of manufacturing site, testing sites, sometimes happening at the same time. Retain samples on every batch to still have a robust comparison to pre-change material. We need to be prepared to test again if lab changes.
- For a very accelerated program, we had 2 pre-and 1 post-change batch only. We had an extensive panel of characterization assays, plus historical knowledge from similar molecules. We were able to anticipate certain types of changes. Did you refer to other filings? Yes, in this one case, we were able to attribute an odd change to something we had observed before and understood. We also collected fractions enriched for this charge variant on ion exchange to show no impact to potency.
- Acceleration puts more stress on comparability in the CGT arena.
- **End to end comparability**
 - For an accelerated BT product (ADC), forced degradation stability helps as a stand-in for real time data, but with some limitations. We would provide stress stability data in the original submission and a commitment to submit real time during review. They want to see END TO END comparability, meaning linked comparability. For changes at DS level, they request that the DS process change is supported by DS comparability and DP comparability on the resulting DP from the same lots, including stability.
 - China filing. DS process change but no change in DP process. Do we need DP comparability? No, focus on DS (?). China regulation is asking for more. DP confirmation runs. China regulations are changing. Many differences from other agencies and additional requests such as in-country testing, PS-80 China pharmacopeia. When sharing characterization data with China, was asked to make it release.
- **Intermediates and process control**

- We have been asked to compare intermediates. Might be expected for critical intermediates, components of ADC or bispecific molecules, unconjugated or unPEGylated intermediates.
- As part of phase 3 enabling first in class ADC molecule. Burden on certain trace of linker. Highly robust and accurate but very close to LOQ. Statistical analysis on these numbers is difficult. Depends on the assay itself.
- Question from the table about whether the process comparability is also compared. Process controls should not change. Have had a comparability protocol agreed in principle with the agency, but when the amendment was submitted there was a change in the control strategy which invalidated the agreement. We had to change the comparability protocol
- Process performance must also be compared. We provide this along with analytical comparability in S.2.6/P.2.
- Also used SAR activities to show that there is no impact. Easier to add assays than to add lots.
- Have tried binning mass spec and MAM data with respect to differences, e.g., 0-5%, 5-10%. Then we can discuss the bins without discussing numerical differences.
- **Cell and Gene Therapy**
 - For Cell and Gene Therapy, the industry is also learning, this is a hot topic at trade association meetings. The CQA assessment is key, along with QTPP this provides a robust foundation for discussion of clinical relevance.
 - For ATMPs, the powerful, consistent analytical tools disappear. At least one phase of development also disappears. Also, we have very limited material. Much of what we have just discussed is not relevant.
- 4. What stability data do you include in comparability assessment? Do you use GMP stability data, and if so in what context? When do you perform side-by-side forced degradation studies?
- **Depends on the change**
 - Depends on the change whether existing GMP stability data is used, for major changes would do side-by-side testing
 - It depends on process change. If we perform the stability program, it's so we can leverage the shelf life assigned to the pre-change shelf life. Occasionally we came to the conclusion that we would not need a forced deg study, but FDA has insisted

in the past to see this. Even with robust scientific discussion we were not successful to remove the forced deg study to support comparability. Question from the table – was this for all types of change or for e.g., formulation change or device change? All types.

- Forced degradation is useful as stability takes time. We assess from the risk assessment if we will gain knowledge from doing the work. Not usually needed for simple, low risk changes. Otherwise, it is worth doing the study as you may be asked for it to mitigate risk.

- **Method variability**

- Need to consider method variability and storage of material if you're going to conduct side-by-side testing later on (i.e. want to avoid degradation of comparator batches prior to use in side-by-side testing)
- Side-by-side not always performed, but if methods have changed since pre-change assays, or if pre-change data was from CRO or from an acquisition, the stress study would be best if performed side by side.

- **Previous knowledge**

- For biosimilars don't have GMP stability data, so would rely on forced deg studies
- We always include forced degradation. For the original BLA we typically provide the whole 9 yards (real time, accelerated, forced deg). Post-approval, depending on your knowledge base, you might not need to include forced deg, but only accelerated and long term. We would go back to our risk assessment and base this on the totality of the changes. For minor changes, e.g., different suite, different line, same equipment we could omit and submit with minimum 3 months real time data, at least for US.
- If we don't have a lot of stability data, we include 1-2 months of forced deg. The study can be ongoing at time of IND submission in preparation for questions.

- **Depends on the agency**

- It depends on the agencies. EU asks for 6m but we try to submit for 3m for both accelerated and stressed. For long term we don't include it if there are no trends over shelf life at long term storage.

- **What conditions?**

- Question from the table, has anyone submitted without 3 stability conditions? The collective experience was “no”, but the discussion moved on to say that it could be justified.

- The risk of forced degradation, if the conditions are too severe, we may see differences, e.g., even a minute pH difference can cause a different degradation profile. Especially if DS is not fully formulated. If the long-term storage is -80°C, what is the accelerated or the stress condition? Often use only 2 temperatures for DS. The changes at 25°C might be irrelevant, but differences must still be explained.
5. For what types of changes if any have you used clinical PK studies to mitigate risk/remove residual doubt?

- **Depends on the change**

- Biosimilar are the extreme case
- Change in DS process with charge profile changes, might require a comparative PK.
- Change to DS process that did not meet comparability criteria (charge, oligosaccharides, etc). For these, it might require more than PK.
- No experience with clinical bridging but expect that PK would not necessarily be sufficient to justify CQA differences.

- **Depends on the phase**

- For late stage we always include clinical bridging or discuss with FDA with clinician present. It is important to have a safety physician who understands CMC and non-clinically relevant changes.
- For an ADC, we have plans to introduce the commercial process material during a pivotal study. All 3 components (Mab, linker, DS) will have a site change. In the clinical study we have planned a contingent PK study and will also seek authority feedback. This will be a small, stand-alone PK, may not be needed. Question about how much clinical exposure is planned from the post-change material. They have asked a statistician but have not gotten a clear answer.
- For a cell line change post-commercial since this is considered a major change. What analytical data did you provide? Extended characterization, potency assay is important. Demonstrate ADCC activity. Forced degradation needed.

- **Patient exposure**

- We had started phase 3 with material from the clinical site. We planned to have >50% of patients treated with material from commercial site. Due to COVID-19, the trial enrollment was delayed and now we have changed to material from

commercial site (very few patients ended up being treated with the pre-change clinical material).

- Is there an amount of clinical exposure you would expect to provide if material is changed during pivotal study? We had introduced post-change material in the clinical trial and had planned to discuss with FDA, but the project was terminated. We know we need some clinical exposure, but don't know how much. Note that clinical comparability is not the same as clinical exposure in terms of statistical powering.

- **Cell and Gene therapy/limited material**

- For gene therapy in early development - if you do PK studies, do you still have to do analytical comparability? PK alone doesn't give the entire picture (e.g. potency assays)
- Very limited in materials for gene therapy - studies must be carefully considered. PK alone is not going to give the full picture. Potency assays gives comparison assurance.

Additional topics of interest:

Regulatory guidance. Where are we where do we need to go?

Industry can take a bigger role on what best practices are.

Raw material comparability. CGT arena.

Consider splitting comparability topic between early stage development versus commercial/post-approval