

2023 CASSS CGTP Summit

Statistical Perspectives on Analytical Comparability Studies for Autologous Cell-based Therapies

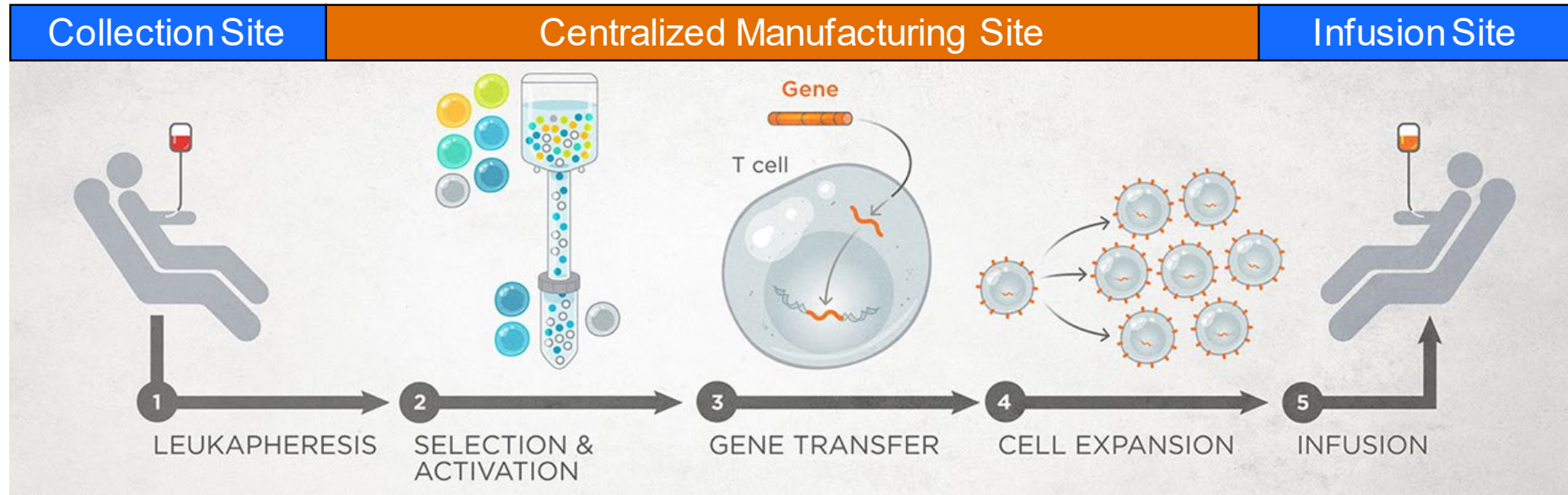
June 26, 2023

Kedar H Dave

Overview

- International Conference on Harmonisation (ICH): Comparability of biotechnological/biological products subject to changes in their manufacturing process **Q5E** (2004).
 - *A determination that a product is “Comparable” indicates that products before and after a manufacturing change are highly similar and that no adverse impact on the quality, safety or efficacy of the drug product has occurred*
 - does not mean that pre- and post- change products are identical or indistinguishable
- FDA’s three-tiered approach and EMA’s reflection paper
- Incorporating unique features related to autologous cell-based therapies
- FDA’s views and EMA’s Q&A for ATMPs

Autologous CAR T Cell Therapies at Bristol Myers Squibb



FDA approved March 2021
EMA approved August 2021



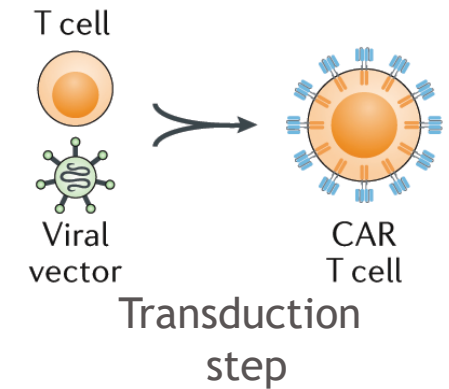
FDA approved February 2021
EMA approved January 2022

Analytical Comparability for Manufacturing Process Lifecycle Management

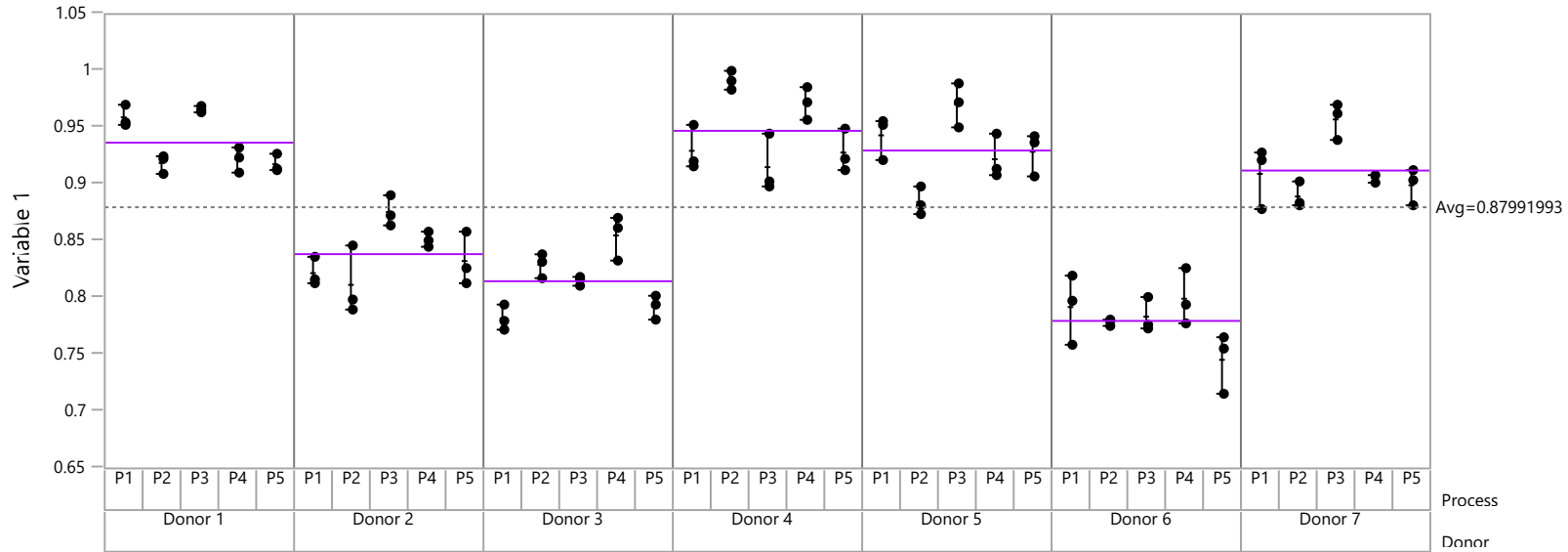
- Change and transfer scenarios:
 - Apheresis & collection process version / site
 - Cellular drug product processing version / site
 - Critical reagent
 - Lentiviral vector process version / site
 - Cryopreservation step
 - Assays¹⁷
- Not practical to develop a one-size-fits-all methodology

Unique features of Analytical Comparability Studies for Autologous CAR-T

- Healthy donor used as a surrogate starting material during development and comparability studies
- Clinical manufacturing and many development studies occur at-scale
- Variability in Critical Quality Attributes (CQAs):
 - starting material is usually the most dominant source
 - lentiviral vector (LVV) lot-to-lot and dosing variability also important
- Release specifications are often “too tight” (relative to traditional drugs)
 - based on larger sample sizes
 - using 95% confidence, **but less than 99%** coverage Tolerance bounds
 - do not allow a buffer for maneuverability
 - high Out-Of-Specification (OOS) rates even under common cause variability
 - could become a *de facto* patient selection mechanism



Variance Components Analysis¹



- Donor variability
- Process variability
- Analytical variability
- Total variability

Variance Components							
Component	Var Component	% of Total	20	40	60	80	Sqrt(Var Comp)
Donor	0.00439270	82.4					0.06628
Process[Donor]	0.00065038	12.2					0.02550
Within	0.00029072	5.5					0.01705
Total	0.00533380	100.0					0.07303

Donor variance makes up 82.4% of the total variance

Variability as a standard deviation for each of the variance components

$$\sigma_{Total}^2 = \sigma_{Donor}^2 + \sigma_{Process}^2 + \sigma_{Analytical}^2 = 0.00439 + 0.00065 + 0.000291 = 0.00533$$

Variability of patient-specific products¹¹

- Patient-specific products use source material from each patient, and every patient is different, therefore every lot will be different-so variability is both expected and acceptable, right?

Yes and No

- For Specifications with wide tolerances, it is difficult to rely on just lot release specifications to show consistency and comparability
- What you end up with for a final product lot should reflect what you started with

When a product has substantial inherent variability
You should aim like this...

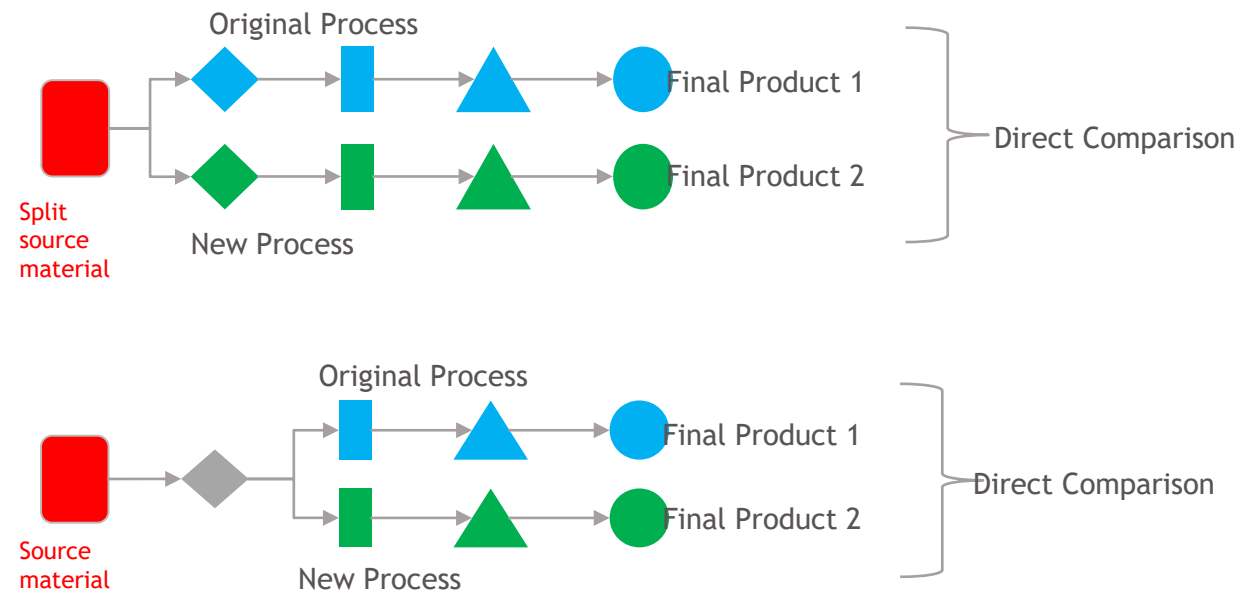


Not this....

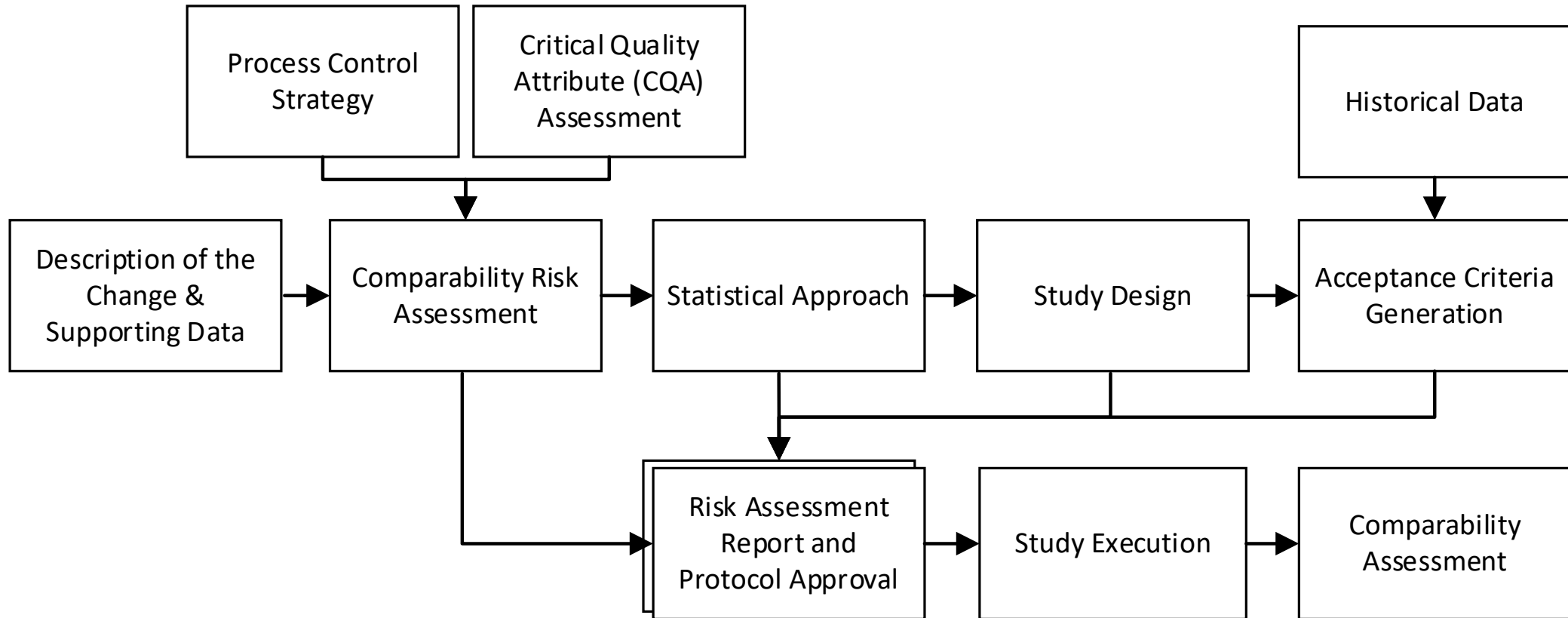


Paired Run Studies^{2,4,12,13,14}

Paired (or “blocked”) run studies split either at the source material or further downstream depending on the nature of change may be used to establish comparability. Paired studies with the same starting donor material may be used to establish comparability to remove patient/donor variability from the assessment.



Risk-Based Comparability Approach^{2,13}



Comparability Risk Assessment - Tier Assignment^{2,5,7,18}

Tiers are assigned to individual attributes based on the phase of the product lifecycle, the nature of the attribute (CQA vs. non-CQA, appropriateness for specific tier evaluation), and the assigned risk score

Tier	Description
Tier 1	A Tier 1 approach assesses analytical comparability using an equivalence test, where equivalence of attributes is assessed by testing the difference in means between the pre and post change attribute.
Tier 2	A Tier 2 approach assesses analytical comparability using a quality range approach, where a quality range is defined based on historical experience and analytical comparability is demonstrated if a sufficient percentage of test lot values fall within the defined quality range.
Tier 3	A Tier 3 approach assesses analytical comparability through an assessment of visual displays and subject matter determination of comparability.

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2017, VOL. 27, NO. 2, 197–205
<http://dx.doi.org/10.1080/10543406.2016.1272606>



Development of statistical methods for analytical similarity assessment

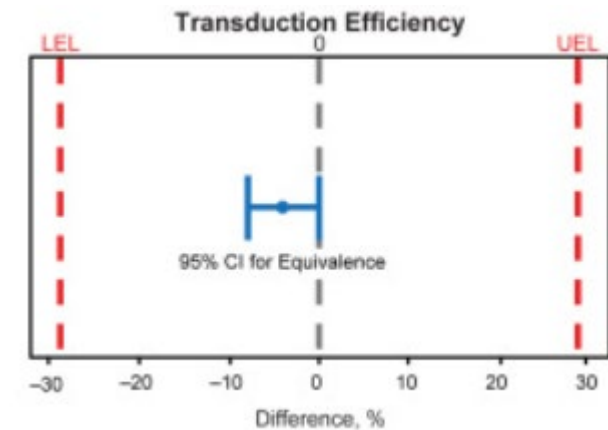
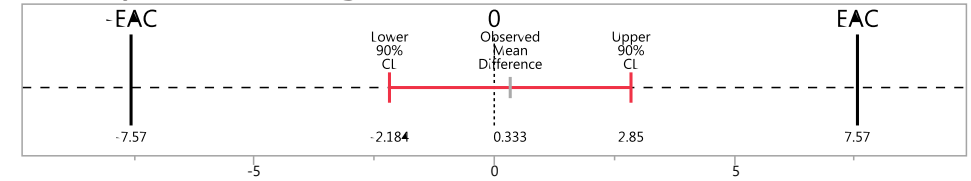
Yi Tsong^a, Xiaoyu Dong^b, and Meiyu Shen^a

^aOffice of Biostatistics, Office of Translational Science, CDER, FDA, Silver Spring, Maryland, USA; ^bGlobal Biostatistical Science, Amgen Inc., Washington, DC, USA

Equivalence Test Approach - 1^{2,13,15,18,19}

- “Two one-sided tests” (TOST) procedure
- Equivalence Acceptance Criteria (EAC) is the practically meaningful difference
 - Determined *a priori* by Subject Matter Experts based on Process and Product Understanding
 - Estimates of process and analytical variability from paired runs are often used to set EAC (i.e., $EAC = k \times \sigma_{Ref}$, where $k = 1.5$ or 2)
 - σ_{Ref} should reflect unexplained, common-cause experimental variability
- Benefits with approach:
 - Where appropriate, utilizes paired runs to remove donor variability
 - Study size can be statistically justified
 - Ensures rigorous comparability evaluation for the highest risk attributes

Example Results Figure



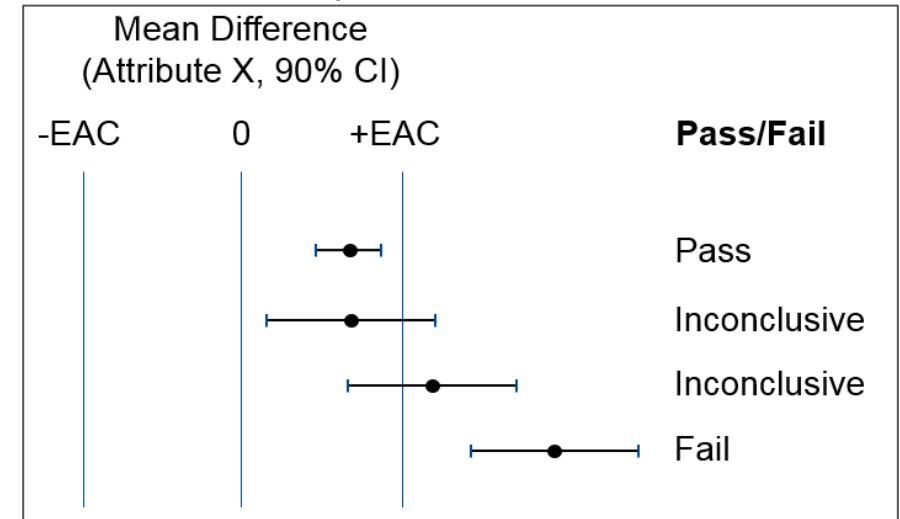
Comments:

- 90% confidence interval for 95% confidence-level
- $\pm 30\%$ might be too wide

Equivalence Test Approach - 2^{2,13,15,18,19}

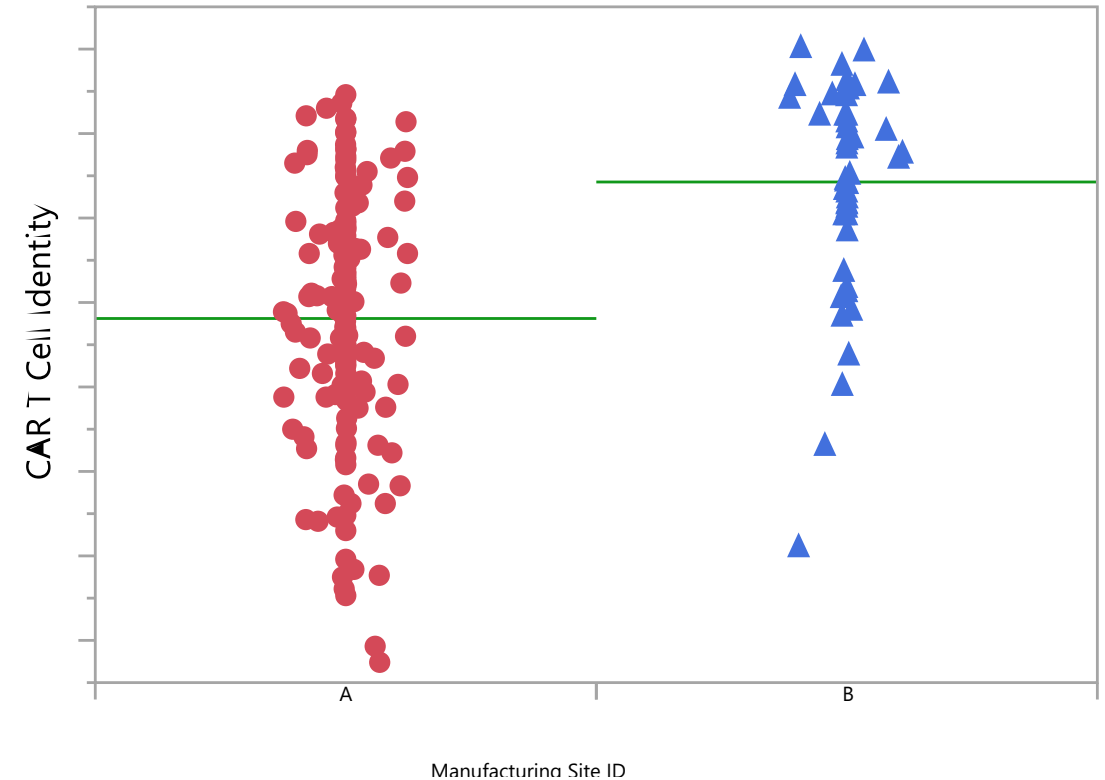
- Inappropriate / insufficient demonstrations for comparability of means
 - *t*-Tests are not appropriate
 - “Point estimate for mean difference inside EAC” is not sufficient
 - FDA’s Bioequivalence testing guidance is not applicable
- CMC quirks
 - “inconclusive” results
 - EAC determined from σ_{Ref}

Overview of Equivalence Test Outcomes



Retrospective Analysis of Manufacturing Data

- Unable to take advantage of paired runs
 - σ_{Ref} based on total variability (i.e., includes starting material variability)
 - If $\text{EAC} = 2 \times \sigma_{\text{Ref}}$ -> too wide
- Large sample sizes
 - Narrow confidence intervals
 - Will fit inside EAC even though there is a shift
- Claim equivalence (based on statistically significant results) even when substantial differences in means exist?



Do these two sites produce comparable product?

EMA Questions and Answers for ATMPs (2019) - 17

Q11: Are statistical approaches appropriate to show comparability of ATMPs?

A: Statistics may provide useful information to support comparability even though any statistical approach has its own limitations and strengths, and those should be well understood and documented before conduct of the comparability exercise and in order to make informed decisions on the comparability utilizing the statistical results.

In any case, it is essential that an appropriate **pre-specified plan** with a justification is provided for the statistical approach chosen and the comparability acceptance criteria proposed for the relevant quality attribute selected according to a risk-based approach. In this regard, it is emphasized that solely meeting specifications is not considered sufficient to conclude on comparability.

A **risk-ranking of CQAs** can be performed to drive the selection of the preferred statistical methodology. A combination of various methodologies can be used to understand the robustness of the chosen statistical approach.

EMA Questions and Answers for ATMPs (2019) - 27

Inclusion of side-by-side analysis of individual values with accompanying descriptive statistics to summarize data (e.g. min-max and 3*sigma ranges) is recommended, particularly when comparing a limited number of samples/batches (e.g., in earlier development phases). Likewise, suitable graphical representations (e.g., individual values scattergrams) could be provided, allowing the identification of possible shifts within the acceptance criteria.

Further consideration should be given to the reflection paper on statistical methodology for the comparative assessment of quality attributes (EMA/CHMP/138502/2017).

Concerning stability aspects, evaluation of comparability between pre- vs. post-change degradation/"time change" rates may be performed e.g., by comparison of the slopes of the time-based regression lines, when applicable. (See also Question 9)

EMA Reflection Paper on Statistical Methodology (2021)⁸

In many instances of QA data comparison, the interval determined by the range between the minimum and the maximum data point observed for the reference distribution is used as 'reference range'. Alternatively, x -sigma intervals (i.e. mean $\pm x \cdot SD$, SD being the standard deviation) are also frequently proposed as reference range. This is often justified by mean $\pm x \cdot SD$ covering a given percentage of the reference distribution (under a normal distribution assumption), e.g. mean $\pm 3 \cdot SD$

As an alternative to x -sigma intervals, tolerance intervals (TI) that take sampling variability into account are sometimes proposed as reference range. A TI is usually derived to estimate a data range by which a specified proportion p (e.g. the central 99%) of the underlying distribution is assumed to be covered with a pre-specified degree of confidence (e.g. 95%). However, although TIs cover the central proportion of the distribution (e.g. the true mean $\pm 3 \cdot SD$ range) with high certainty, this does not imply that the TI is a good (narrow) estimator of this range. In contrary, as TIs take measurement error into account for quantification of uncertainty regarding the estimate of the central proportion of the reference distribution, this does imply that TIs are generally broader when the uncertainty is large (i.e. sample size is small).

FDA Draft Guidance (2019)⁹

801 Appropriate analyses of the comparative analytical data are necessary to support a demonstration
802 that the proposed product is highly similar to the reference product notwithstanding minor
803 differences in clinically inactive components. One approach to data analysis would be the use of
804 descriptive quality ranges for assessing quantitative quality attributes of high and moderate risk,
805 and the use of raw data/graphical comparisons for quality attributes with the lowest risk ranking
806 or for those quality attributes that cannot be quantitatively measured (e.g., primary sequence).
807 The acceptance criteria for the quality ranges (QR) method in the comparative analytical
808 assessment should be based on the results of the sponsor's own analysis of the reference product
809 for a specific quality attribute. The QR should be defined as $(\hat{\mu}_R - X\hat{\sigma}_R, \hat{\mu}_R + X\hat{\sigma}_R)$, where $\hat{\mu}_R$ is
810 the sample mean, and $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots.
811 The multiplier (X) should be scientifically justified for that attribute and discussed with the
812 Agency. Based on our experience to date, methods such as tolerance intervals are not
813 recommended for establishing the similarity acceptance criteria because a very large number of
814 lots would be required to establish meaningful intervals. The sponsor can propose other methods
815 of data analysis, including equivalence testing.

016

Considerations when selecting a comparability approach³

Consideration	Description
1	Protect patients from consequences of concluding comparability when processes are not comparable.
2	Protect sponsors from consequences of concluding lack of comparability when processes are comparable.
3	Incentivize sponsors to acquire post-change process knowledge.
4	Enable decision making with practical sample sizes and reasonable risks associated with considerations 1 and 2.
5	Compare both location and spread of the process distributions.
6	Consider criticality of attribute and align criteria with subject matter expert (SME) knowledge.
7	Select an approach that is transparent and easy to explain to scientists with no formal statistical training.

FDA CBER OTAT Town Hall (2022)¹⁰

For products that are made from donor- or patient-derived starting material, we recognize that the drug product attributes may vary from lot to lot due to the inherent differences in the starting material. And so we recommend that your comparability study isolates the differences in the manufacturing process. And so to do this, you should use a split cellular starting material design. And this will allow you to use a paired difference analysis for the comparability assessment. And so this will take out the variability that's associated with the starting material.

It's often appropriate to use an equivalence approach to evaluate the comparability. When you're really getting into the statistical approach and for this, it is important to have normally distributed data. And this is where talking to a statistician would be helpful if you don't have normally distributed data. And so for an equivalence approach, you can establish a range for the allowable difference in the population means, and in this case, exceeding that range would have an adverse effect on product quality.

FDA CBER OTAT Town Hall (2022)¹⁰

Alternatively, you may use a **quality range approach** where you evaluate post-change results fall within a defined range. And this quality range approach is generally more acceptable for lower-risk attributes, as it's not as robust of an assessment.

In both cases, the equivalence acceptance criteria should be **predefined before you start your study**, and it should be based on your product knowledge. Once again, we do understand that this is a challenge for the field, and I've said a lot of technical things in that response related to some of the statistical approaches that we do recommend, and so we are developing that guidance. And we oftentimes do discuss these approaches with sponsors in the context of the IND, since there are a lot of product-specific considerations that you need to think about as you're designing your comparability study. Thanks.

Comparability: what we can learn from the review of advanced therapy medicinal products⁶

Table 4. Comparability-related issues evident during initial review.		
Type	Theme	Example
Site change	Inadequate comparability data provided for a change of manufacturing site	Kymriah
Surrogate material	Requirement to show that donated starting material used in lieu of patient material is suitably representative for CQAs studied	Libmeldy, Kymriah
Potency assay	Changes and variability in potency test resulting in difficulties for comparative analyses	Zolgensma
	Acceptance criteria not considered suitable for adequate control	Zynteglo
Efficacy data	In the absence of comparability data, some efficacy data was excluded from consideration (primary data)	Kymriah
Non-clinical data	Requirement to show suitable comparability for product used for non-clinical studies and intended for commercial supply	Yescarta
	Limited value of comparative <i>in vivo</i> pharmacology studies in the context of data to support a comparability assessment	Kymriah
Major objection†	Acceptance criteria for potency assay not adequate for mitigating risk of a treatment failure	Zynteglo
	Insufficient comparability information for medicinal product from proposed commercial process and earlier versions of product	Spherox
post-approval measures required	Continued monitoring (trending) of analytical results, e.g., as part of process verification	Libmeldy
	Requirement for additional analyses post-approval	Zolgensma
	Re-evaluate clinical data to understand whether release specification acceptance criteria can assure efficacy and safety	Zynteglo
	Develop an assay to monitor a vector impurity	Strimvelis

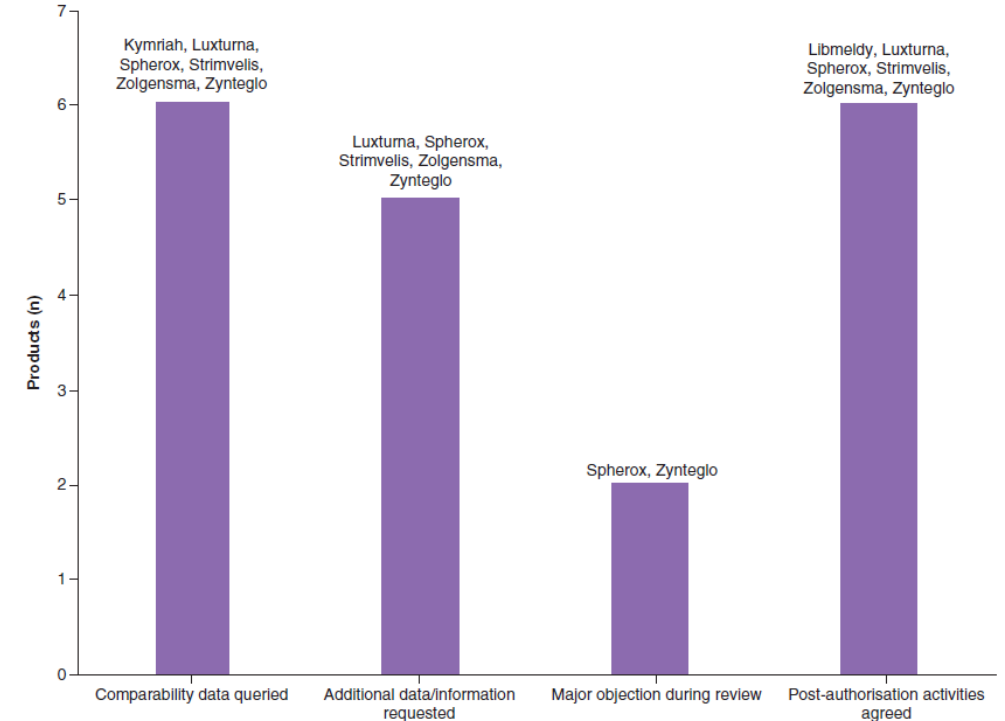


Figure 1. Extent of regulatory concerns identified in relation to comparability for 12 recently approved advanced therapy medicinal products. The source of the information used was the European Public Assessment Reports [39] and

Publicly available summaries from Marketing Authorization Applications for 12 Cell & Gene Therapy products were evaluated to explore data expectations for product characteristics pre- and post- changes (comparability).

Concluding Remarks

- Comparability studies should be a Business priority
- Plan ahead, avoid comparability studies¹⁶
 - is it possible to combine changes?
- Recommend making changes prior to initiating clinical studies intended to support efficacy for licensure¹⁴
- Stay updated about recent advances
 - be mindful of the uniqueness of autologous cell-based therapies
- Employ and groom CMC Statistician(s)

References

1. Ashton, Roland, et al. “Variance Components Analysis for Autologous Therapy Process Development”, Cell & Gene Therapy Products (CGTP), Bethesda (10 to 12 June 2019).
2. Brust, Erica “Risk-Based Approaches for Autologous CAR T-Cell Therapy Comparability”, Cell & Gene Therapy Bioprocessing & Commercialization, Virtual (19 to 22 October 2020).
3. Burdick, R. K. “Statistical considerations for comparative assessment of quality attributes”, *Statistics in Biopharmaceutical Research*, 13(3), 297-302 (2021)
4. Burger, Scott R. "Demonstrating Comparability in Manufacturing Gene Therapy Products", Workshop B: Best Practices for Comparability Design & Protocol, Hanson Wade - Gene Therapy Comparability, Advanced Cell and Gene Therapy (30 March 2021).
5. Christl, Leah "Overview of the Regulatory Pathway and FDA’s Guidance for the Development and Approval of Biosimilar Products in the US" (2017)
<https://www.fda.gov/files/drugs/published/FDA%E2%80%99s-Overview-of-the-Regulatory-Guidance-for-the-Development-and-Approval-of-Biosimilar-Products-in-the-US.pdf>

References

6. Cockroft, A., & Wilson, A. “Comparability: what we can learn from the review of advanced therapy medicinal products”, *Regenerative Medicine*, 16(07), 655-667 (2021).
7. EMA “Questions and answers: Comparability considerations for Advanced Therapy Medicinal Products (ATMP)”, EMA/CAT/499821/2019 (6 December 2019).
8. EMA “Reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development”, EMA/CHMP/138502/2017 (26 July 2021).
9. FDA Draft Guidance “Development of Therapeutic Protein Biosimilars: Comparative Analytical Assessment and Other Quality-Related Considerations” (May 2019).
10. FDA CBER OTAT Town Hall Series: “Gene Therapy Chemistry, Manufacturing, and Controls” (29 September 2022).
11. Finn, Tom “Early Stage Manufacturing Considerations for Cell Therapy Products”, Cell & Gene Therapy Products (CGTP) (10 July 2018).
12. Markovic, Ingrid “Comparability Assessment and Protocols as Enablers of Change Management over Product Lifecycle Markovic”, CMC Strategy Forum, Rockville (10 July 2016).

References

13. Momonoi, Yoko "Comparability Studies for Autologous Cell Therapy", PDA Cell and Gene Therapy Conference, San Diego (5 to 6 December 2017)
14. Schultz, Kim "FDA Framework for Gene Therapy Development", ASGCT Policy Summit, Washington, D.C., (4 November 2019).
15. Shen, Cjris "Comparability Studies for Autologous Cell Therapy Products", CMC Strategy Forum, Gaithersburg (17 to 18 July 2017).
16. Stanton, Dan "CGT CMC: Plan ahead, avoid comparability studies says Novartis", Bioprocess International, Bioprocess Insider (11 September 2019).
17. Teixeira, Mia "Statistical approach for autologous cell therapy potency method comparability", Non-Clinical Biostatistics (NCB) Conference (20 June 2023).
18. Tsong, Yi, Xiaoyu Dong & Meiyu Shen "Development of statistical methods for analytical similarity assessment", *Journal of Biopharmaceutical Statistics*, 27:2, 197-205 (2017).
19. Tyagarajan, S., Spencer, T., & Smith, J. "Optimizing CAR-T cell manufacturing processes during pivotal clinical trials", *Molecular Therapy-Methods & Clinical Development*, 16, 136-144 (2020).

Thank you