# Visualizing Meaningful Change in Small Sample Sizes

Charlie Iaconangelo, PhD[1], Shauna McManus, MPH[2], Daniel Serrano, PhD[3], Bryant Barnes, BS[4]

*[1]USA OPEN Health Group, New York, NY, USA; [2,3,4]OPEN Health Group, Bethesda, MD, USA*

## BACKGROUND

- In the field of clinical outcome assessments (COAs), current practice is to estimate Meaningful Within-Patient Change (MWPC) using an anchor-based approach.[1]
- FDA Guidance 3 and 4 recommend using empirical cumulative distribution functions (eCDFs) to visualize meaningful change and establish thresholds.[2,3]
- The results are increasingly used to inform regulatory decisions and in drug packaging.
- However, when sample sizes within anchor categories are small, the eCDF may obscure, rather than illuminate, the relationship between COA change score and external anchor.
- Improving best practices in characterizing MWPC is crucial for regulatory decision-making and is an important consideration to help clinicians effectively communicate with patients.

## OBJECTIVE

- **Explored when an eCDF may be difficult to interpret and how such challenges can be circumvented by using scatterplots to visualize MWPC.**

## METHODS

### Scatterplot Approach

- Subject scores are plotted with baseline scores on the x-axis and follow-up scores on the y-axis (**Figure 1**), rendered in a scatterplot.
- The scatterplot is stratified by color to represent anchor category-specific scores.
- Subjects who do not respond at follow-up due to toxicity/progression/death are plotted in a separate color, using baseline response and an assigned follow-up score reflecting maximum severity.
- A 45-degree line represents "No Change" and the vertical distance between subject score and the 45-degree line represents the "Observed Change."
- Additional lines representing the mean/median change score of each anchor category are plotted parallel to the "No Change" line.

### Comparison of eCDF and Scatterplot Approach

- A simulation study was conducted to compare the eCDF and scatterplot approaches.
  1. A large dataset was generated to reflect real trial data.
  2. Thresholds were computed and the eCDFs were plotted.
  3. A series of smaller datasets were selected/subsampled from the full dataset.
  4. Threshold values were computed and both eCDFs and the scatterplot approach were plotted.
- The eCDFs and scatterplots were compared in terms of which approach yielded more interpretable results under small sample sizes.

### Evaluation of Full Dataset

- For each item, response distribution was evaluated by computing the number and percentage of subjects responding in each category at each timepoint.
- The COA score was computed as a sum of 5 items, with a possible range of 0-15.

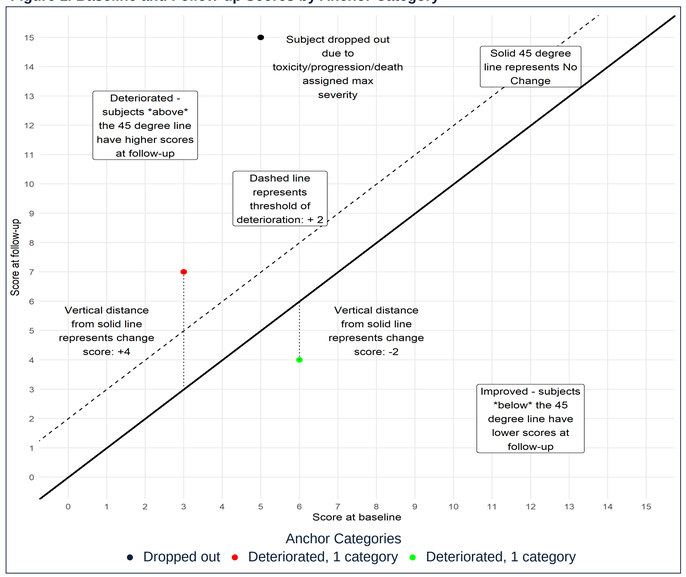
### Subsampled Datasets

- To compare the two approaches, subsampled datasets with n=500, n=100, n=50, and n=30 were created by randomly selecting subjects from the full N=1000 dataset.
- Using the n=30 subsample as an illustration, subsampled data were created as follows:
  - Randomly selected 30 subjects from the full N=1000 dataset.
  - 10% of the subsample were randomly selected as dropped-out due to intercurrent event (e.g., n=3 out of the n=30 in the subsampled dataset).
    - These subjects were assigned the maximum severity at follow-up, a score equal to 15.

### Evaluation of Subsampled Dataset

- The eCDF and the scatterplot were each used to evaluate the responsiveness of the COA score to changes in the patient condition, as reflected in the anchor categories.
- The full dataset was used to provide reference values of the thresholds and overall performance.

**Figure 1. Baseline and Follow-up Scores by Anchor Category**

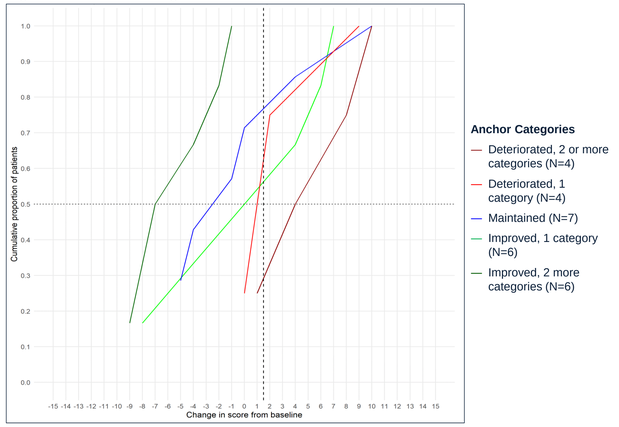



## RESULTS – SIMULATION STUDY

### eCDF, Subsampled Dataset

- Dotted line represents 'Threshold of Deterioration' (**Figure 2**).
- Threshold value was 1.5, defined as the median change of the "Deteriorated, 1 category" anchor group.
- Approximately the same proportion of subjects in the "Improved, 1 category" and "Deteriorated, 1 category" had change scores less than or equal to the threshold.
- **The eCDF visually is not easy to interpret and would likely be considered inconclusive evidence for the responsiveness of the COA score.**

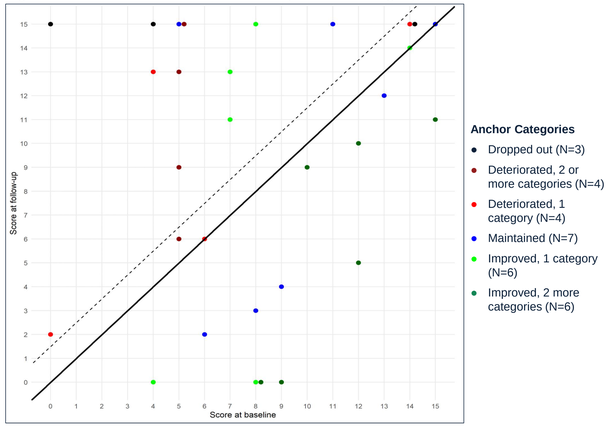
**Figure 2. eCDF of Change in Score From Baseline, N=30**




### Scatterplot, Subsampled Dataset

- Dotted line represents 'Threshold of Deterioration' (**Figure 3**). Threshold value was 1.5.
- Subjects above the solid line of "No change" reported COA scores at follow-up that were greater (i.e., reflecting greater severity) than their baseline score.
- Can see that three green points are problematic; the anchor category is "Improved, 1 category" but the PRO score indicates an increase in severity.
- **The scatterplot offered a transparent view of the relationship between anchor category and baseline/follow-up COA scores.**
- **Missing data is accounted for on the Scatterplot.**

**Figure 3. Baseline & Follow-up Scores by Anchor Category, N=30**




## CONCLUSIONS

### Transparency

- eCDF requires computing the density using the PRO scores.
- With small sample sizes, a few subject PRO scores can influence the density values and thereby obscure the relationship between PRO scores and anchor category in the eCDF.
- Scatterplot avoids the computation of density and therefore ensures that the relationship between PRO scores and anchor category is visible when sample sizes are small.
- At sample sizes below n=50, the eCDF may yield uninterpretable visualization, while the scatterplot approach consistently yields interpretable visualizations of meaningful change.
- The scatterplot approach thus ensures transparency in the evaluation of meaningful change.

### Relationship of Baseline and Follow-up Scores

- The baseline and follow-up scores are included in the scatterplot, allowing for a visual representation of the relationship between the two.

## STUDY LIMITATIONS

- The scatterplot approach proposed does not resolve many of the potential obstacles involved in evaluating MWPC.
- The scatterplot does not fully incorporate the MPWC into the estimand framework, because the scatterplot relies on the same computation of thresholds that is not robust to MAR or MNAR drop-out.
- Approaches to estimating thresholds that adjust for non-random drop-out should be considered in future research.

## REFERENCES

1. Food and Drug Administration. Assessment of the Use of Patient Experience Data in Regulatory Decision-Making https://fda.report/media/150405/Assessment-of-the-Use-of-Patient-Experience-Data-in-Regulatory-Decision-Making.pdf
2. Food and Drug Administration. Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments. https://www.fda.gov/media/116277/download
3. Food and Drug Administration. Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. https://www.fda.gov/media/132505/download

Presented at the 27th Annual Meeting of the International Society for Pharmacoeconomics and Outcomes Research May 15-18, 2022; National Harbor, Washington, D.C.