Anchor-based Thresholds for Meaningful Within-Patient Change in the Phase 3 Trial to Evaluate Tislelizumab for the Treatment of 2/3L NSCLC

Shauna McManus[1], Charlie Iaconangelo[1], Daniel Serrano[1], Lauren Podger[2], Yiyuan Ma[3], Lin Zhan[4], Boxiong Tang[4], Gisoo Barnes[4]

[1]Open Health, 4350 East-West Highway, Suite 1100 Bethesda, MD 20814, US

[2]Open Health, 20 Old Bailey, London, England, EC4M 7AN

[3]BeiGene, Inc, BeiGene, China

[4]BeiGene, Ltd., Emeryville, CA, US

Background and Objective

Recent FDA guidance has emphasized the importance of an anchor-based approach to estimating thresholds of meaningful within-patient change (MWPC)

- FDA: There is insufficient evidence to support the position that 10-point thresholds define a clinically meaningful within-patient score change in the European Organization for Research and Treatment of Cancer (EORTC) instruments[1]
- Anchor-based methods should be used to estimate thresholds
- There is little published work reporting results based on the implementation of an anchor-based approach to EORTC instruments[2-5]

This research applies the anchor-based approach to estimating MWPC for the EORTC QLQ-C30[4] and EORTC QLQ-LC13[5]

[1]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/downloads/Drugs/NewsEvents/UCM620708.pdf
[2]Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. JNCI: Journal of the National Cancer Institute. 1993;85(5):365-376. [3]King M. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Quality of life research. 1996;5(6):555-567. [4]Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-376. [5]Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. European journal of cancer (Oxford, England : 1990). 1994;30a(5):635-642.

Trial Overview

BGB-A317-303 (NCT03358875)[1]

- Phase 3 randomized, placebo controlled clinical trial
- Evaluated tislelizumab (anti-PD-1 antibody) for the treatment of 2/3L non-small cell lung cancer (NSCLC) vs docetaxel

Clinical Characteristics

- N=805 patients
  - Aged ≥ 18 years
  - Histologically confirmed Stage IIIB or IV NSCLC
  - Either squamous or non-squamous histology
  - Failed prior therapy

Patient-Reported Outcomes (PRO) Collected

n=326 patients who had the required PRO data at baseline and follow-up

The following PRO instruments were utilized:

- EORTC QLQ-C30
- EORTC QLQ-LC13

[1]Zhou C, Huang D, Yu X, Liu Y, Fan Y, Shu Y, et al. Results from RATIONALE 303: A global phase 3 study of tislelizumab (TIS) vs docetaxel (TAX) as second- or third-line therapy for patients with locally advanced or metastatic NSCLC. [abstract]. In: Proceedings of the 112th Annual Meeting of the American Association for Cancer Research; 2021 April 10-15. Philadelphia (PA): AACR; 2021. Abstract nr CT039.

Methods

Anchor-based approach to calculating MWPC was implemented:

1. EORTC QLQ-C30 Global Health/QoL Status scale (2 items, averaged together) used as anchor
   - Anchor change score was calculated, patients accordingly assigned to 1 of 5 anchor categories (Deteriorated, 2+ categories; Deteriorated, 1 category; Maintained; Improved, 1 category; Improved, 2 or more categories)
2. Thresholds were computed for each anchor category
   - Mean and median PRO scores were calculated stratifying on anchor change groups
3. Anchor groups and thresholds were visualized using empirical cumulative distribution functions (eCDFs)
   - Not a statistical test; visual examination of the eCDFs to see if there is noticeable separation between curves representing each anchor category at a given threshold value

What Constitutes Interpretable Meaningful Change?

- FDA Guidance 3[7]:
  - "threshold(s) based on transformed scores may reflect less than one category change on the raw score scale, which is **not** useful for the evaluation and interpretation of clinically meaningful change"
- All EORTC functional and symptom scales:
  - Raw scores 1-4
  - Transformed scores 0-100
  - 1 category change on raw score maps to transformed score change of ±33.3 points

Takeaway: following FDA guidance, threshold estimates of MWPC for transformed EORTC scores should be at least ±33.3 to constitute interpretable meaningful change

Results: QLQ-LC13 Thresholds

EORTC Symptom Scale	Anchor Category	Mean	Median	N	Percent
Cough	Deteriorated, 2+ Categories	9.52	0.00	14	3.2
Cough	Deteriorated, 1 Category	0.00	0.00	50	11.4
Cough	Maintained	-7.48	0.00	156	35.6
Cough	Improved, 1 Category	-10.09	0.00	76	17.4
Cough	***Improved, 2+ Categories***	***-14.44***	***0.00***	***30***	***6.8***
Pain in Arm or Shoulder	***Deteriorated, 2+ Categories***	***14.29***	***16.67***	***14***	***3.2***
Pain in Arm or Shoulder	Deteriorated, 1 Category	-3.33	0.00	50	11.4
Pain in Arm or Shoulder	Maintained	-1.28	0.00	156	35.6
Pain in Arm or Shoulder	Improved, 1 Category	-3.07	0.00	76	17.4
Pain in Arm or Shoulder	***Improved, 2+ Categories***	***-15.56***	***0.00***	***30***	***6.8***
Pain in Chest	Deteriorated, 2+ Categories	2.38	0.00	14	3.2
Pain in Chest	Deteriorated, 1 Category	2.00	0.00	50	11.4
Pain in Chest	Maintained	-2.35	0.00	156	35.6
Pain in Chest	Improved, 1 Category	-7.46	0.00	76	17.4
Pain in Chest	***Improved, 2+ Categories***	***-12.22***	***0.00***	***30***	***6.8***

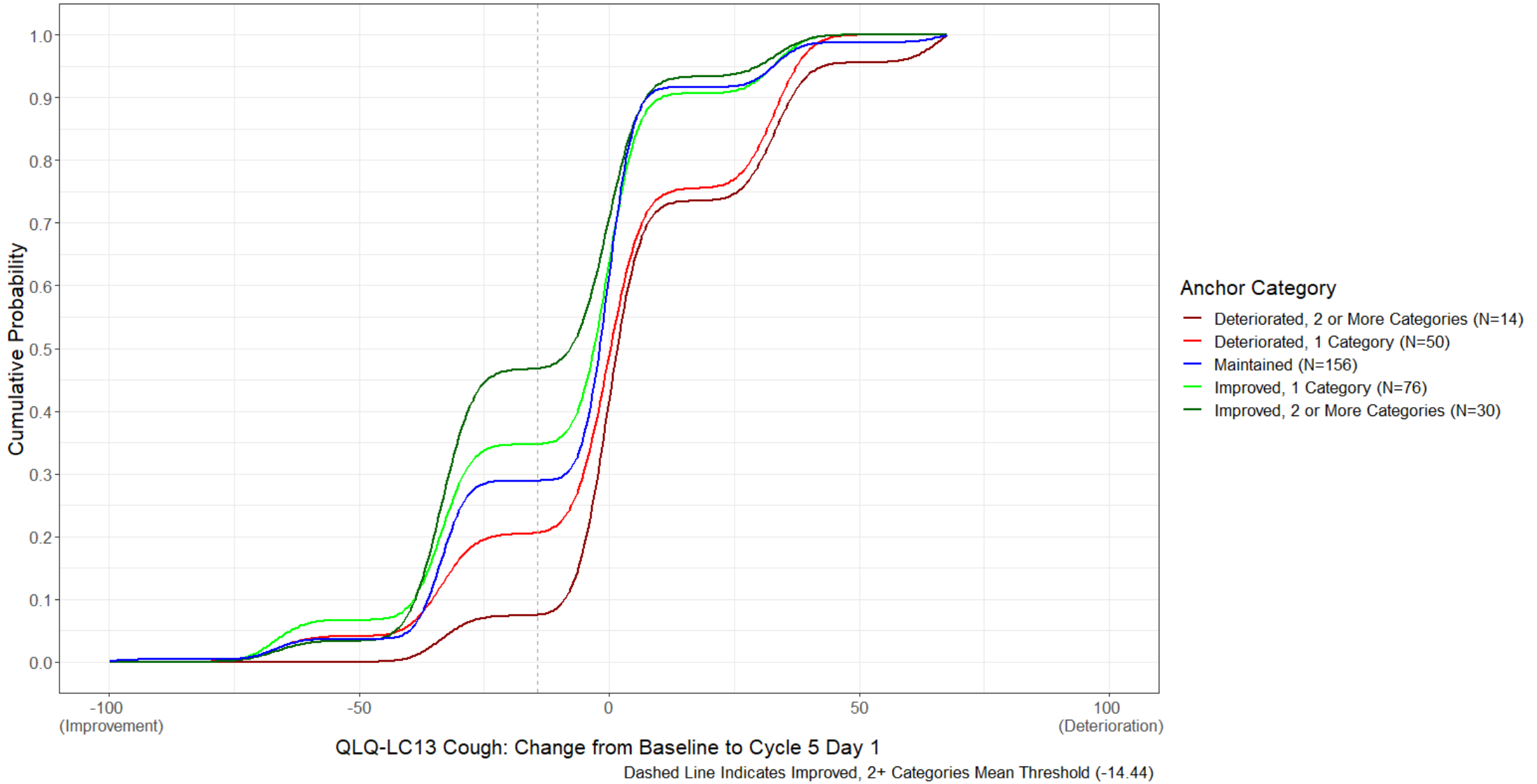
Point estimates that came closest to this definition generally approached +/- 16.667, the transformed score mapping to approximately a 1/2 category change on the raw score; thresholds approaching 16.667 are highlighted

Results: QLQ-C30 Thresholds

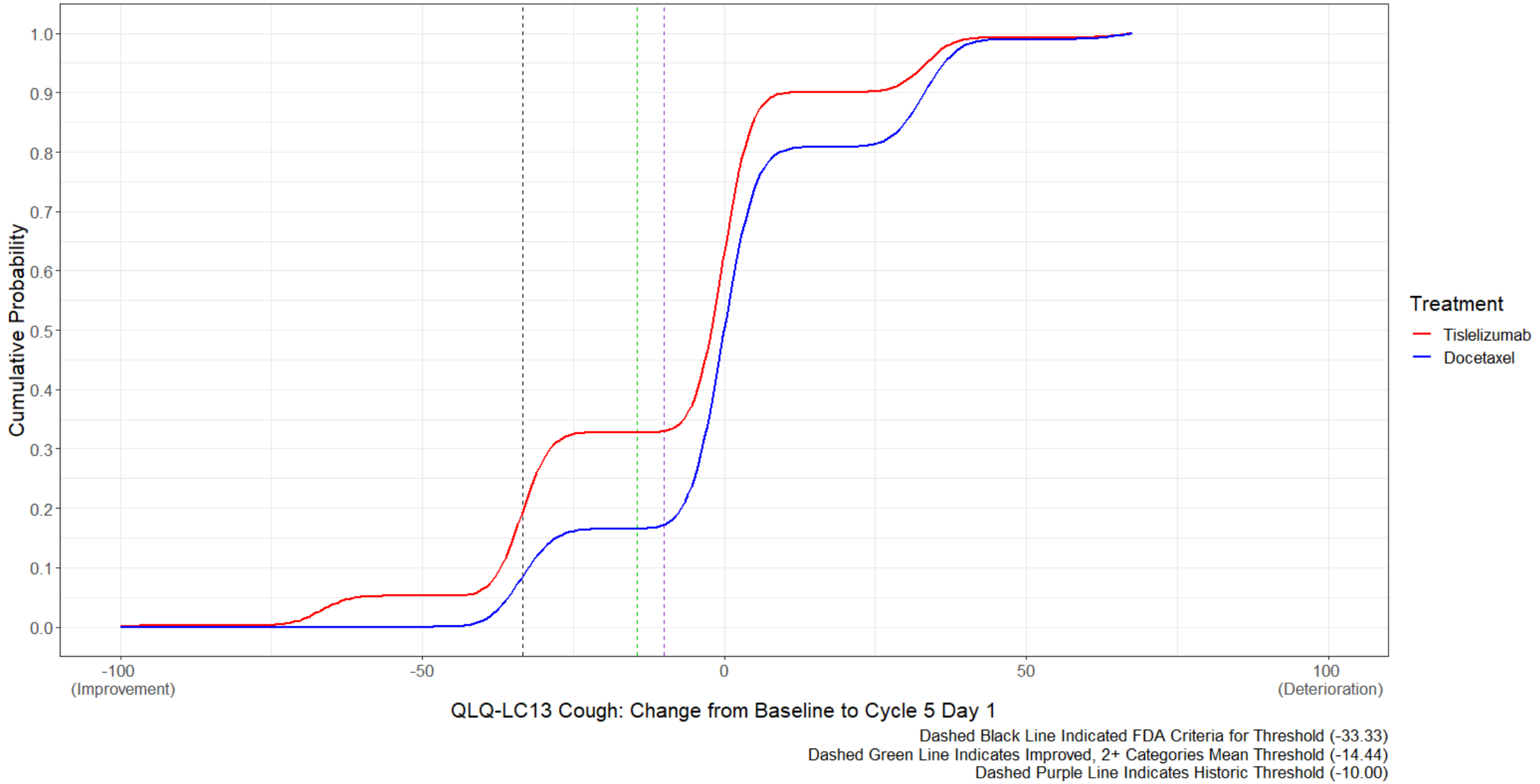
EORTC Symptom Scale	Anchor Category	Mean	Median	N	Percent
Pain	Deteriorated, 2+ Categories	5.95	0.00	14	3.2
Pain	Deteriorated, 1 Category	1.33	0.00	50	11.4
Pain	Maintained	0.11	0.00	156	35.6
Pain	Improved, 1 Category	-7.02	0.00	76	17.4
Pain	***Improved, 2+ Categories***	***-13.33***	***-16.67***	***30***	***6.8***

Point estimates that came closest to this definition generally approached +/- 16.667, the transformed score mapping to approximately a 1/2 category change on the raw score; thresholds approaching 16.667 are highlighted

eCDF of the QLQ-LC13 Cough Single Item, by Anchor Group



eCDF of the QLQ-LC13 Cough Single Item, by Treatment



Summary of Findings

- Thresholds of meaningful within-patient change were generated for QLQ-LC13 and QLQ-C30 functional and symptom scales
- FDA criteria for interpretability requires 1 category change on the raw scale
  - This maps to ±33.3 points on transformed scale
- Four clinically relevant PRO scores had threshold estimates of approximately ±16.7, which maps to a 1/2 category change on the raw score
  - QLQ-LC13 Cough
  - QLQ-LC13 Pain in Chest
  - QLQ-LC13 Pain in Arm or Shoulder
  - QLQ-C30 Pain
- Note that these thresholds exceeded historic ±10-point thresholds

Limitations

Analysis and conclusions limited by the appropriateness of the anchor measure

- FDA recommends the use of global static, current-state, or change measures to generate thresholds of MWPC
  - i.e., anchors patient global impression of severity (PGIS) or change (PGIC) anchors
- PGIS/PGIC were not collected in the BGB-A317-303 trial, instead the QLQ-C30 Global Health/QoL status domain functioned as a global anchor; ***it is recommended for future studies that the PGIS/PGIC measures are included for defining anchor groups***

Conclusions

- Evidence suggests that estimable and interpretable thresholds can be derived that exceed historic ±10-point thresholds
- Observed thresholds exceeded historic ±10-point thresholds but did not exceed FDA criterion
- Observed thresholds were associated with separation between treatment arms in the range considered to be clinically meaningful