Using a Condition-Specific Measure of Patient-Reported Outcomes to Derive Utilities in Myelofibrosis

Neil S Roskell,[1] Estella Mendelson,[2] Diane Whalley,[1] Christopher J Knight[1]

[1]RTI Health Solutions, Manchester, United Kingdom;
[2]Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States

BACKGROUND

- Myelofibrosis is a chronic myeloproliferative disorder affecting the bone marrow
- Patients often suffer from spleen enlargement (splenomegaly) and constitutional symptoms such as fever, night sweats, and weight loss, which can detrimentally affect a patient's quality of life
- Ruxolitinib, a potent and selective oral inhibitor of Janus kinase 1 (JAK1) and JAK2, provides rapid and durable improvement of splenomegaly and disease-related symptoms in patients with myelofibrosis, and offers marked clinical benefits that are independent of *JAK2* V617F mutational status and myelofibrosis subtype
- Generic preference-based measures of health can be used to support the analysis of utility gains from treatments. The EQ-5D™ is the preferred preference-based measure of the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom (UK). However, the limitations of generic measures in disease areas such as oncology are widely recognized
- Condition-specific measures offer more relevant assessments of health-related quality of life and can be used to derive utilities

OBJECTIVES

- Primary: to use data collected with the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire for Cancer (QLQ-C30) in a myelofibrosis clinical trial to derive utilities
- Secondary:
  - To investigate the pattern of utility changes over time by treatment group
  - To estimate the gains in utility associated with response to treatment in patients with myelofibrosis, where response is defined by spleen volume reduction and absence of constitutional symptoms

METHODS

- A pivotal, phase 3, open-label clinical trial, COMFORT-II (Controlled Myelofibrosis Study with Oral JAK Inhibitor Therapy) was designed to assess the efficacy and safety of ruxolitinib (n = 146) versus best available therapy (BAT; n = 73) in the treatment of patients with myelofibrosis. Alongside clinical measurements, patients completed the QLQ-C30 at baseline and weeks 8, 16, 24, and 48
- Two different algorithms for determining utility values were applied to the patient QLQ-C30 data collected during the trial:
  - The first algorithm, devised by McKenzie and van der Pol,[2] maps scores from the QLQ-C30 to utility values derived from the generic EQ-5D health state classification system (EQ-5D algorithm)
  - The second algorithm, devised by Rowen and colleagues,[3] uses QLQ-C30 scores directly to generate utility values based on the condition-specific EORTC-8D health state classification system (EORTC-8D algorithm)

Treatment group analysis

- Mean and change in utility values over time by treatment group were derived for 2 patient populations:
  - Observed dataset: patients' utility values were included at time points for which they had valid observations
  - Complete dataset: patients were included only where they had valid utility values at every time point between baseline and week 48
- Utility values were examined for different response outcomes in terms of spleen volume, constitutional symptoms, and adverse events
  - Spleen volume: the cutoff points for spleen volume reduction groups were 10%, 25%, 35%, and 50%. Patients with disease progression (death, leukemic transformation, splenectomy, splenic irradiation, spleen growth) were included in the less than 10% group
  - Constitutional symptoms: the absence or presence of weight loss, fever, or night sweats was determined by patients' responses to relevant items of the Functional Assessment of Cancer Therapy for patients with lymphoma (FACT-Lym)

METHODS *(continued)*

  - Adverse events: the presence or absence of grade 3 or 4 (severe or life-threatening) adverse events
  - Ten response outcome categories were further defined by a combination of percentage spleen volume reduction and the absence or presence of constitutional symptoms
  - For the response outcome summaries, the utility value was considered the unit of observation, such that individual patients could provide multiple data points to each of the utility summaries
  - Because individual patients could contribute more than 1 data point, means and standard errors (and confidence intervals [CIs]) for response groups were computed using a simple random effects model treating the patient as a random effect

RESULTS

Patient Baseline Characteristics

- **Table 1** shows that the baseline characteristics for patients were reasonably similar across the 2 treatment groups, including spleen volume and the presence of constitutional symptoms

Table 1. Characteristics of the Patient Population at Baseline

Characteristic	Ruxolitinib n = 146	BAT n = 73	Total N = 219
Age in years, mean (SD)	65.1 (9.7)	65.2 (10.3)	65.2 (9.9)
Sex, n (%)			
Male	83 (56.8%)	42 (57.5%)	125 (57.1%)
Female	63 (43.2%)	31 (42.5%)	94 (42.9%)
Baseline spleen volume[a] in cm^3, mean (SD)	2,662.1 (1,351.26)	2,631.1 (1,405.27)	2,651.7 (1,366.31)
Baseline constitutional symptoms, n (%)			
Present	101 (69.2%)	46 (63.0%)	148 (67.6%)
Absent	45 (30.8%)	27 (37.0%)	71 (32.4%)
Baseline prognostic risk group, n (%)			
Intermediate-2–risk patients	58 (39.7%)	29 (39.7%)	87 (39.7%)
High-risk patients	88 (60.3%)	43 (58.9%)	131 (59.8%)
Missing	0	1 (1.4%)	1 (0.5%)

SD = standard deviation.

[a] The average spleen volume for a healthy adult[4] is approximately 100 to 200 cm^3.

- Utility summaries were performed overall and by baseline prognostic risk group (intermediate-2 and high). No important differences in utilities were seen by baseline prognostic risk; thus, results in this poster are overall rather than by baseline prognostic risk group
- **Table 2** contains overall utility summaries for all patients for each algorithm

Table 2. Utilities by Algorithm

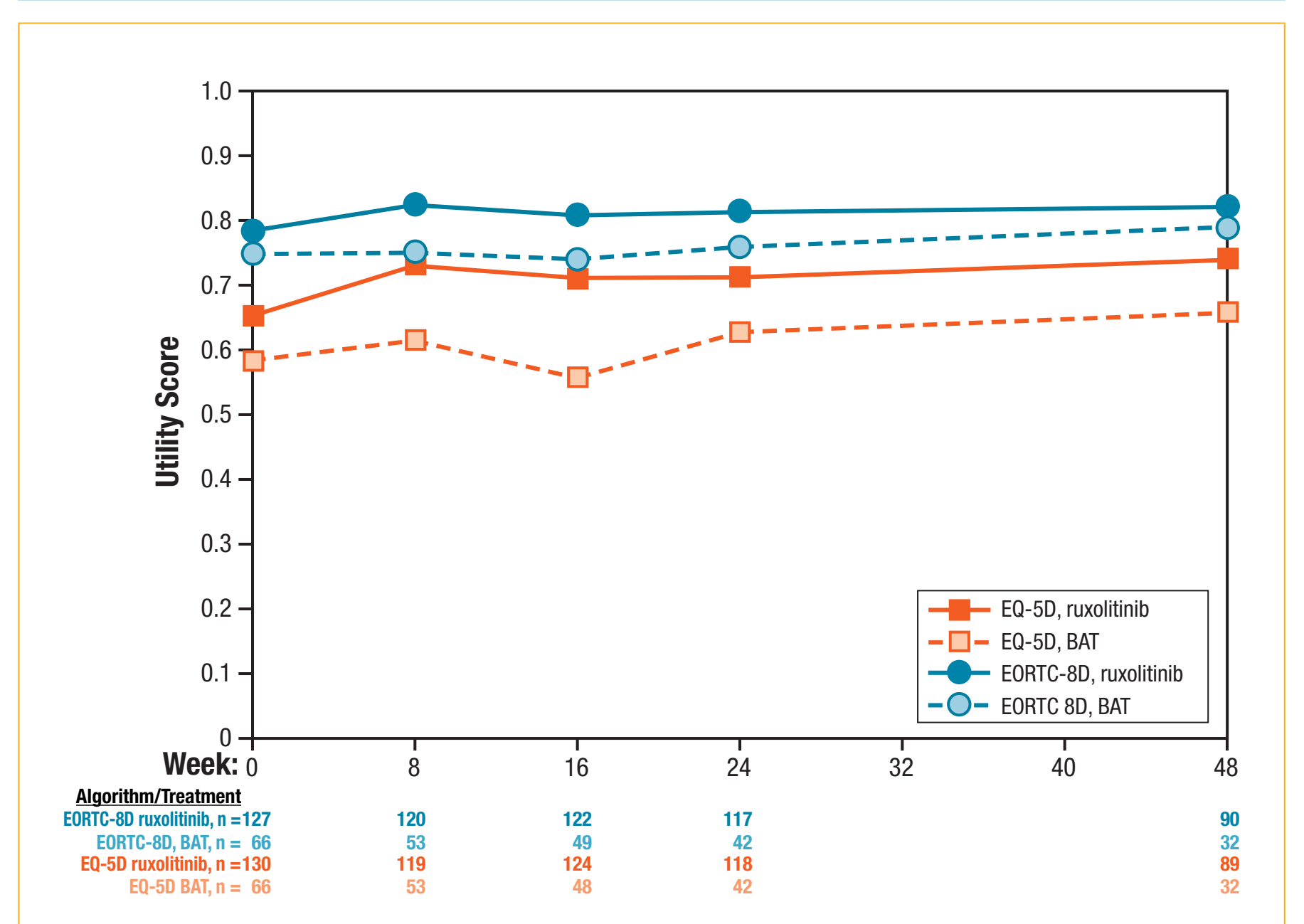
Category	Univariate Statistic	EQ-5D Algorithm	EORTC-8D Algorithm
All patients/ utilities	N (observations)	822	817
	Mean (95% CI)	0.659 (0.626, 0.693)	0.785 (0.767, 0.802)
	Range	-0.181, 1.048	0.322, 1.000

Utility Scores by Treatment Group

- **Figure 1** and **Figure 2** present observed mean utilities and changes from baseline in utility derived using each algorithm by treatment and week

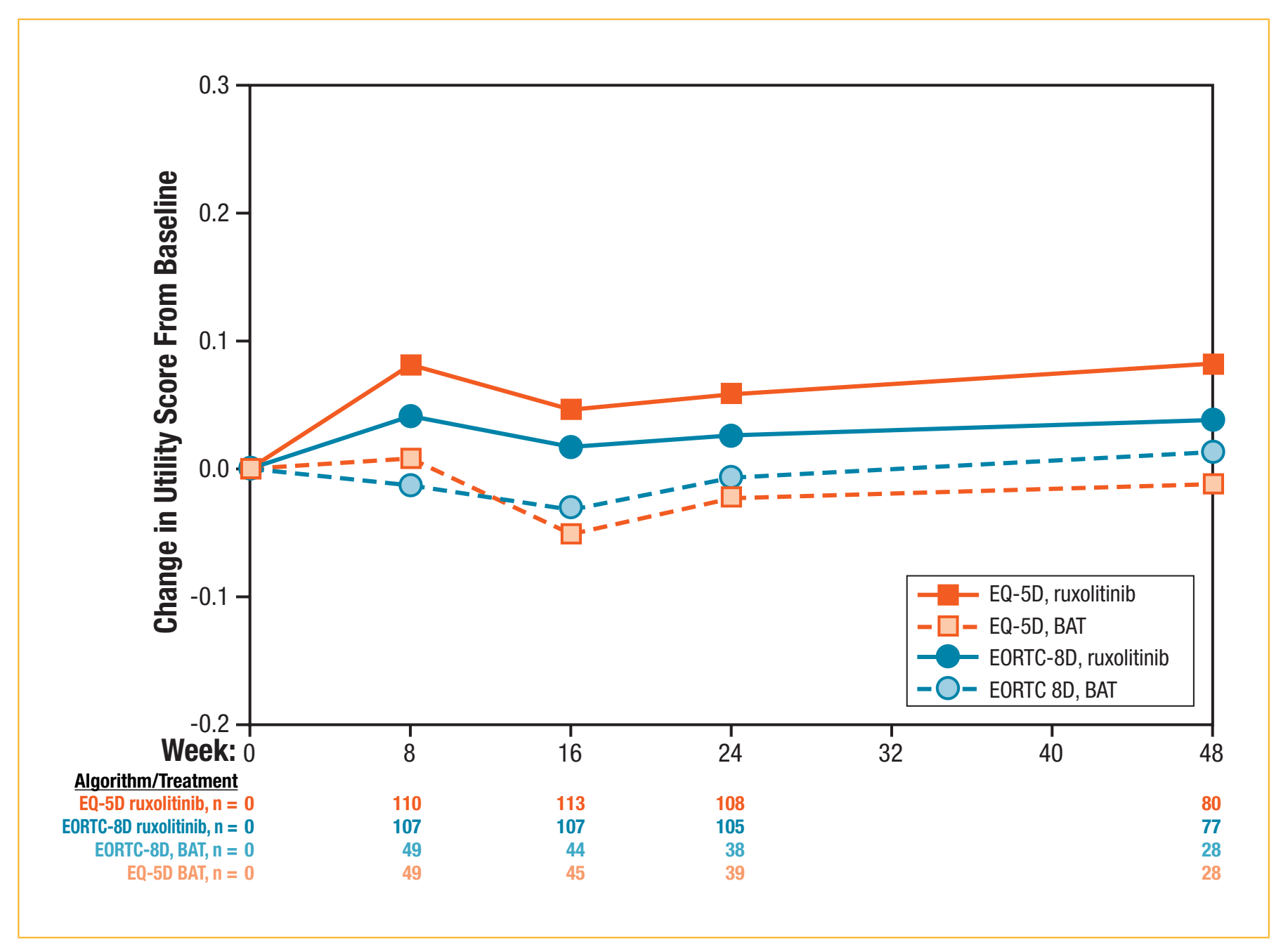
RESULTS *(continued)*

Figure 1. Mean Utility Values by Treatment and Assessment Week



- The results from **Figure 1** suggest overall improvements across time for both treatments using both algorithms
  - Although mean utilities at week 48 were larger for ruxolitinib (EQ-5D = 0.740 [standard error, SE, 0.025], EORTC-8D = 0.822 [SE 0.014]) than for BAT (EQ-5D = 0.658 [SE 0.049], EORTC-8D = 0.791 [SE 0.023]), differences also were evident at baseline (EQ 5D: ruxolitinib = 0.653 [SE 0.021], BAT = 0.584 [SE 0.036]; EORTC-8D: ruxolitinib = 0.785 [SE 0.011], BAT = 0.749 [SE 0.019])

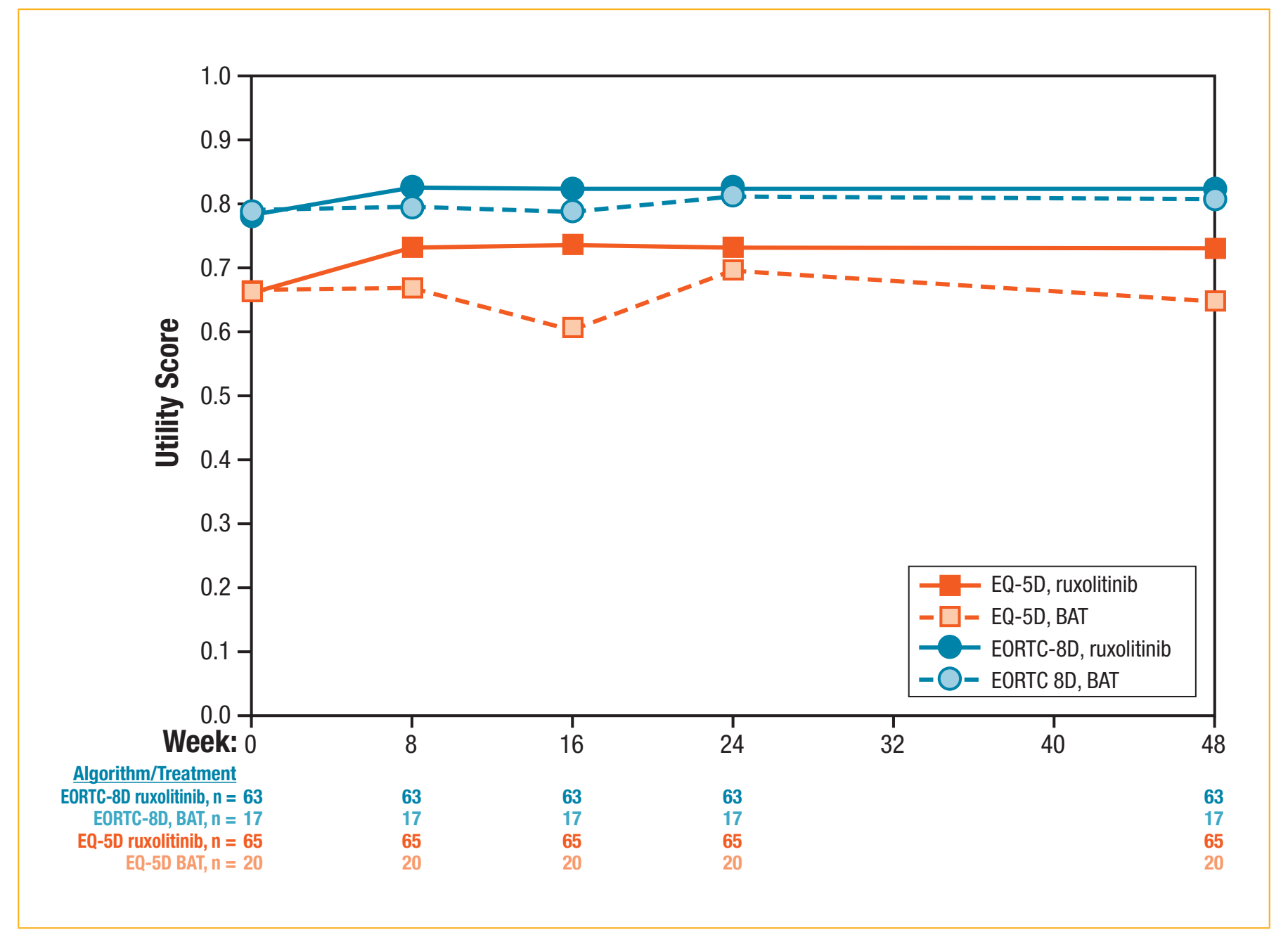
Figure 2. Mean Utility Change From Baseline by Treatment and Assessment Week



  - The mean changes from baseline in utility at week 48 for ruxolitinib were (EQ-5D = 0.082 [SE 0.025], EORTC-8D = 0.038 [SE 0.013]) and for BAT were (EQ 5D = –0.012 [SE 0.040], EORTC-8D = 0.013 [SE 0.021]) (**Figure 2**)
  - Changes in scores indicate that utility derived from both algorithms improved between baseline and week 48 for the ruxolitinib group, while there was little or no change in the BAT group. The difference between treatment groups is more marked in the EQ-5D algorithm
  - The numbers of patients with observed utility scores diminished considerably over time. The percentage of patients with utility scores at both baseline and week 48 was only 61.5% for the EQ-5D and 60.6% for the EORTC-8D in the ruxolitinib group, and only 42.4% for both algorithms in the BAT group

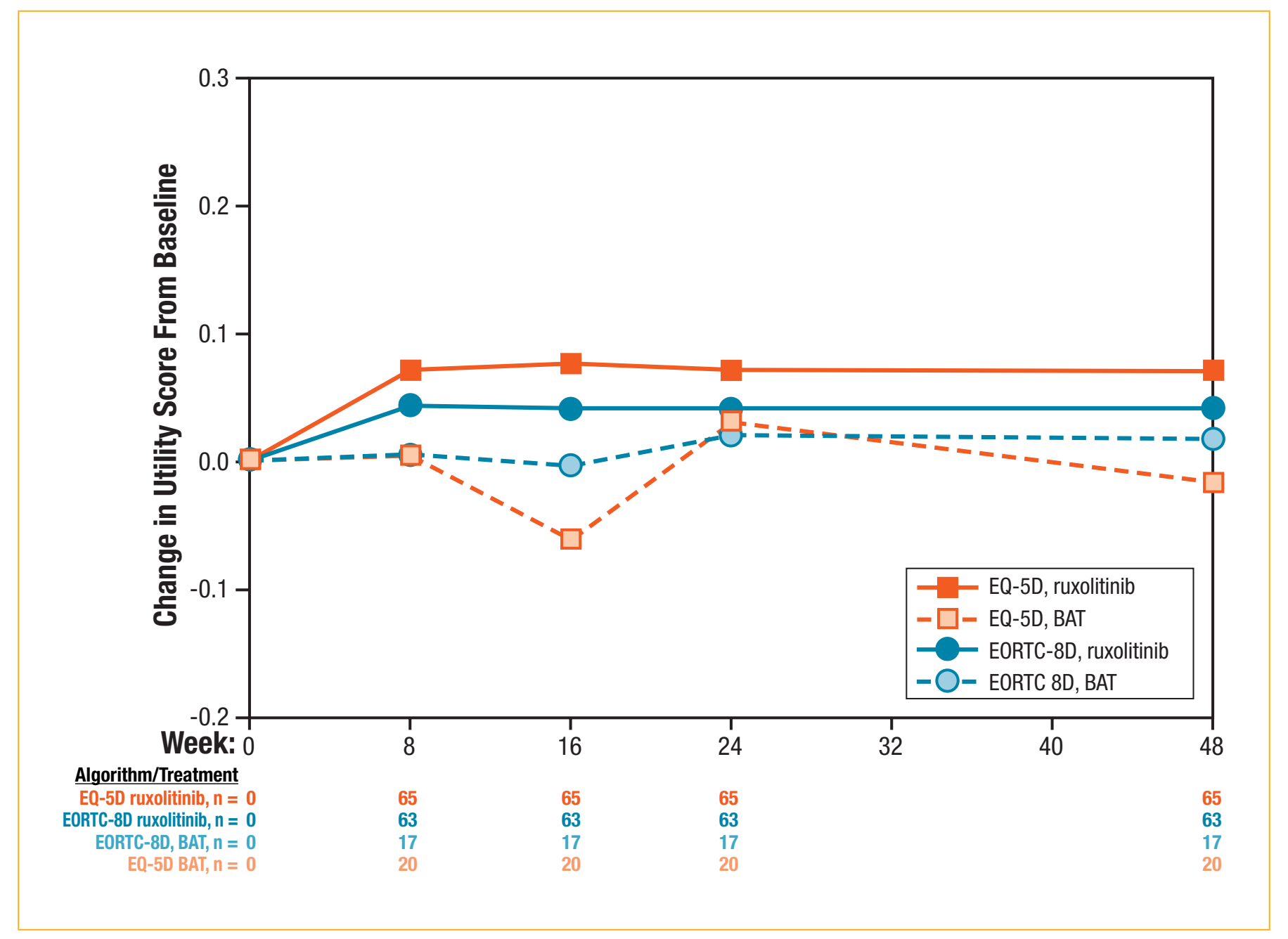
- Given the dropout rate in both treatment arms over time, we repeated the analyses summarizing utility scores by treatment over time for patients with data available throughout the 48 weeks (ie, the "complete dataset"). **Figure 3** and **Figure 4** present mean utilities and changes from baseline in utility by treatment and week for this subgroup of patients

Figure 3. Mean Utility Values by Treatment and Assessment Week for the Group of Patients Followed Through 48 Weeks



  - The mean utilities at week 48 were larger for ruxolitinib (EQ 5D = 0.730 [SE 0.028], EORTC-8D = 0.823 [SE 0.015]) than for BAT (EQ-5D = 0.647 [SE 0.056], EORTC-8D = 0.807 [SE0.027]) (**Figure 3**)
  - However, the differences between treatment groups at baseline seen in the observed dataset are not evident in this complete dataset (EQ-5D: ruxolitinib = 0.660 [SE 0.031], BAT = 0.665 [SE 0.057]; EORTC-8D: ruxolitinib = 0.782 [SE 0.016], BAT = 0.790 [SE 0.031])
  - The mean changes from baseline in utility at week 48 were: EQ-5D = 0.070 [SE 0.026] and EORTC-8D = 0.041 [SE 0.014] for ruxolitinib, and EQ 5D = –0.017 [SE 0.052] and EORTC-8D = 0.017 [SE 0.031] for BAT (**Figure 4**)
  - These results suggest greater improvements over time with ruxolitinib than with BAT; again, the differences between treatments were more notable when using the EQ-5D algorithm

Figure 4. Mean Utility Change From Baseline by Treatment and Assessment Week for the Group of Patients Followed Through 48 Weeks



Utility Scores by Response Outcome

- The following figures provide utility summaries for the different response outcome groups in terms of spleen volume reduction categories (**Figure 5**), and 35% spleen volume response, constitutional symptoms, and grade 3 or 4 adverse events (**all Figure 6**)

Figure 5. Mean Utility Values (95% CIs) Presented for Spleen Volume Response, by Level of Response and Algorithm

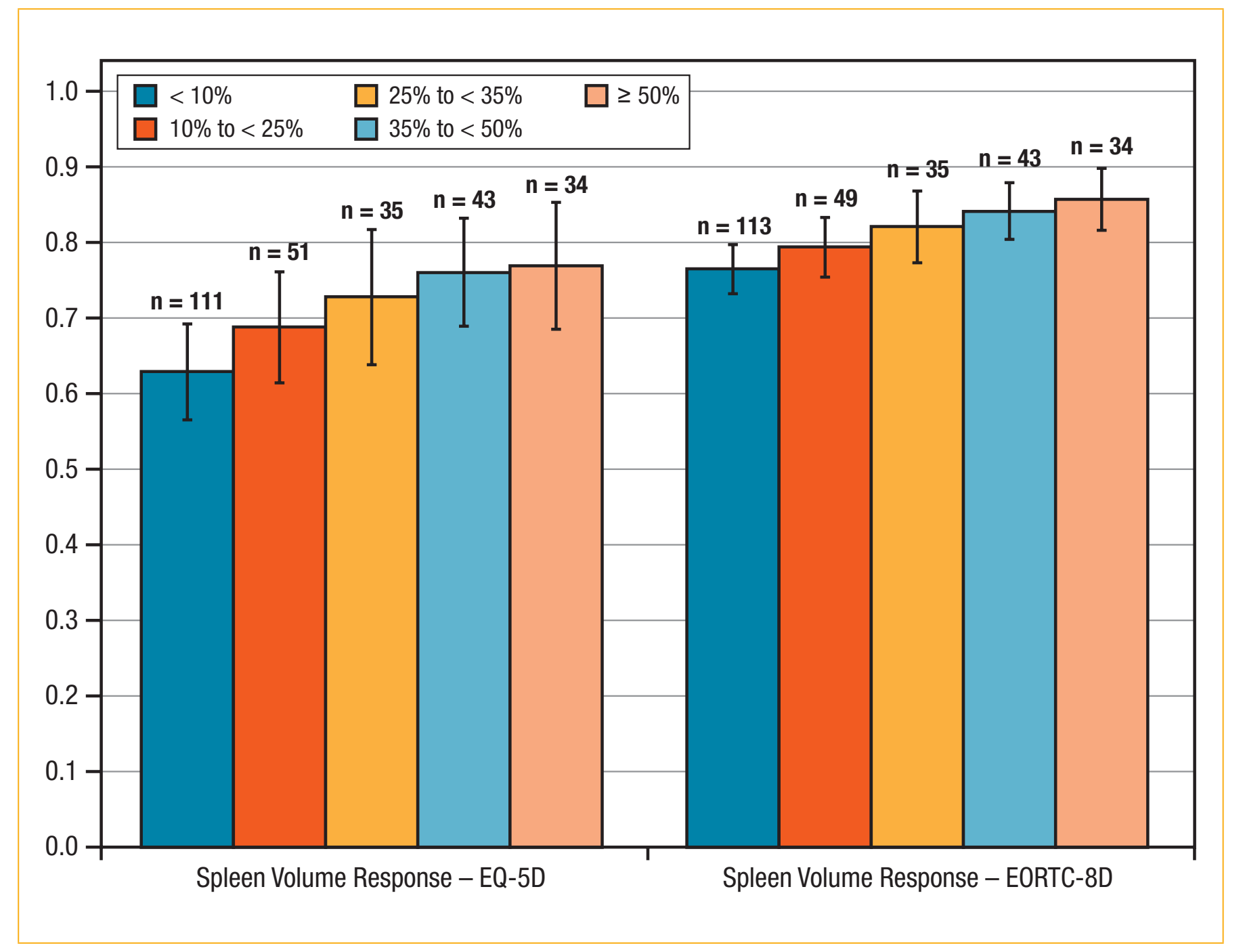
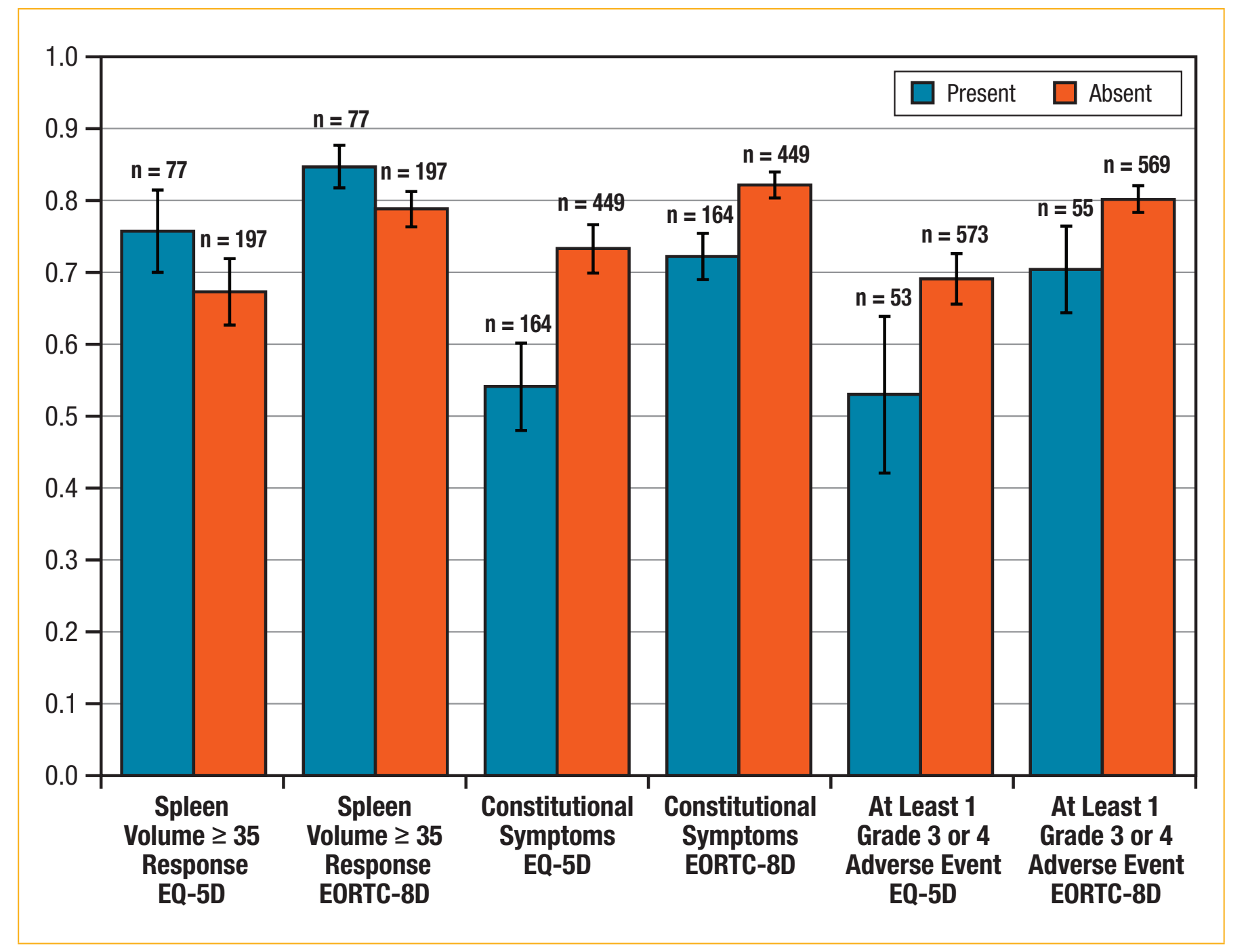


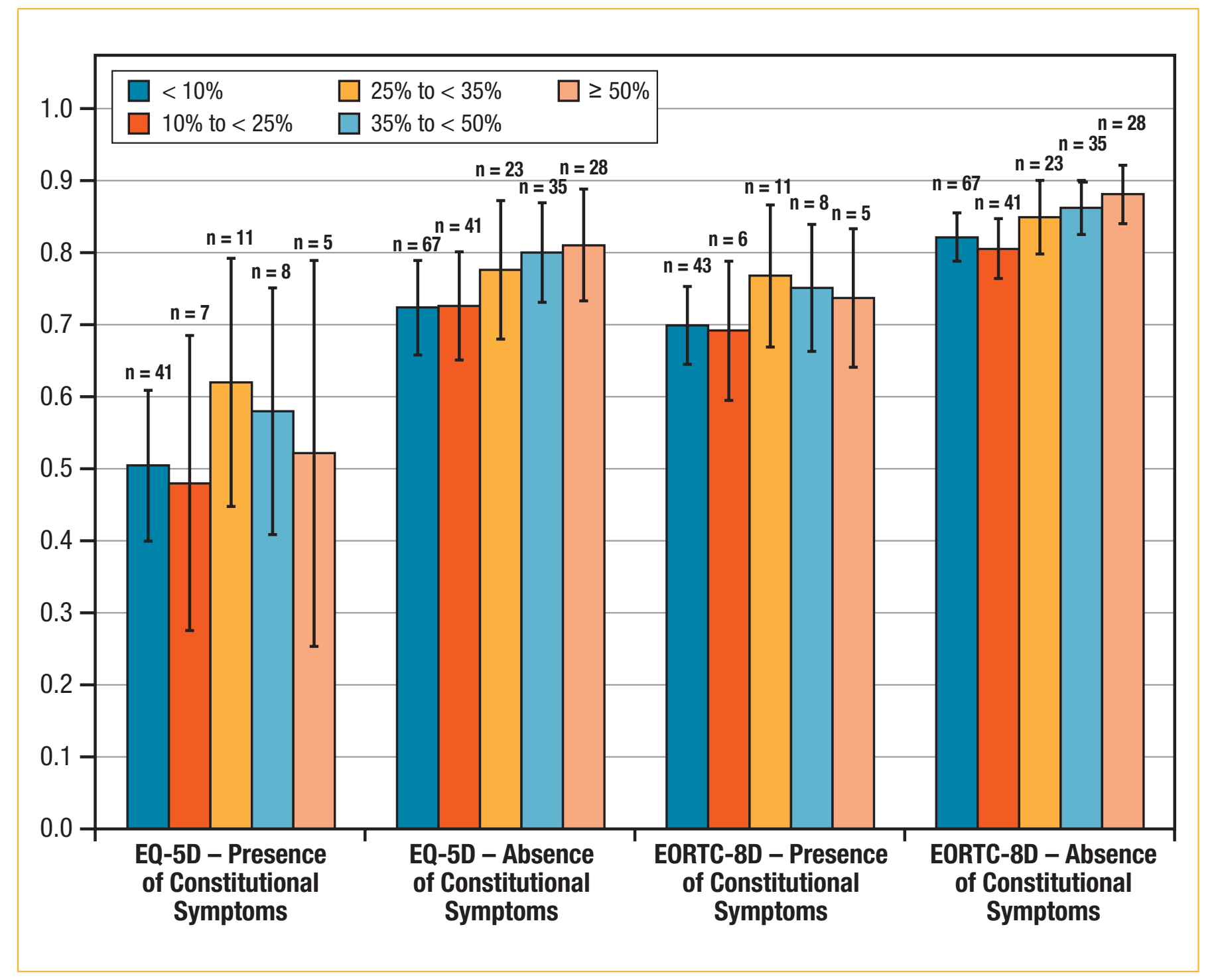
Figure 6. Mean Utility Values (95% CIs) Presented for Presence or Absence of 35% Spleen Volume Response, Constitutional Symptoms, and Grade 3 or 4 Adverse Events by Algorithm



[a] Note that the blue bars indicate presence of each outcome, and that presence of spleen volume response is a positive outcome while presence of constitutional symptoms and adverse events are negative outcomes.

- The results from each algorithm showed that a greater response to treatment with respect to reduction in spleen volume was associated with higher utility values
  - There appeared to be a large negative impact on utility relating to the presence of constitutional symptoms, which was more pronounced in the EQ-5D algorithm
- The presence of a grade 3 or 4 adverse event also was associated with a lower utility than the absence of one

Figure 7. Mean Utility Values (95% CIs) Presented for Presence or Absence of Constitutional Symptoms and Level of Spleen Volume Response, by Algorithm



- **Figure 7** provides utility summaries for the composite endpoint of spleen volume reduction and constitutional symptoms
  - There was a marked difference in mean utility between the categories representing the greatest response and those representing the least response for both algorithms
  - For example, using the EQ-5D algorithm, a patient without constitutional symptoms and spleen volume reduction of 50% or more had a mean (CI) utility of 0.809 (0.732, 0.887) compared with only 0.504 (0.399, 0.608) for a patient with constitutional symptoms and a spleen volume reduction of less than 10%
  - Similarly, using the EORTC-8D algorithm, a patient without constitutional symptoms and spleen volume reduction of 50% or more had a mean (CI) utility of 0.880 (0.839, 0.920), compared with 0.698 (0.644, 0.752) for a patient with constitutional symptoms and spleen volume reduction of less than 10%

LIMITATIONS

- There were notable differences in observed baseline utility scores between treatments. It is unclear whether these differences reflect natural variation in the measure or whether there are some underlying patient differences in the treatment groups that resulted in the BAT group having lower utility values, in spite of the baseline demographics being similar between groups
- Applying 2 different algorithms in the analysis provided an assessment of the sensitivity of the results to the method of deriving utilities from the condition-specific QLQ-C30. However, the analysis did not inform on the quality of the algorithms when used in myelofibrosis clinical trials
- Although the analysis of change scores helps to account for differences at baseline, it assumes that patients are missing or drop out at random, which is unlikely to have been the case
  - Patients lacking treatment response are more likely to drop out; therefore, the overall utility estimates at the end of the trial are likely to be derived from a higher percentage of responders than would be in the starting population
  - This in turn will overestimate the mean utilities at the end of the trial. Given the larger dropout in the BAT arm, it can be argued that this overestimation will be greater in the BAT arm than in the ruxolitinib arm. In such a case, the observed treatment differences between ruxolitinib and BAT would be underestimated

CONCLUSIONS

- This study derived utility values for myelofibrosis patients from a condition-specific measure, using 2 different algorithms
- The improvement in utility in patients treated with ruxolitinib was greater than that seen in patients treated with BAT using both algorithms
- The analyses further suggest that splenomegaly and the presence of the constitutional symptoms associated with myelofibrosis have a detrimental impact on patients' utility

REFERENCES

1. Verstovsek S, Kantarjian H, Mesa RA, Pardanani AD, Cortes-Franco J, Thomas DA, et al. Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N Engl J Med*. 2010;363(12):1117-27.
2. McKenzie L, van der Pol M. Mapping the EORTC QLQ C-30 onto the EQ 5D instrument: the potential to estimate QALYs without generic preference data. *Value Health*. 2009;12(1):167-71.
3. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health*. 2011;14(5):721-31.
4. Kaneko J, Sugawara Y, Matsui Y, et al. Normal splenic volume in adults by computed tomography. *Hepatogastroenterology*. 2002;49(48):1726-1727.

CONFLICT OF INTEREST DISCLOSURE

RTI Health Solutions was a paid consultant to Novartis for this project. Neil S Roskell, Diane Whalley, and Christopher J Knight are paid employees of RTI Health Solutions. Estella Mendelson is a paid employee of Novartis. Novartis licensed ruxolitinib from Incyte Corporation for development and potential commercialization outside the United States. Incyte has retained rights for the United States and has received FDA approval for ruxolitinib for the treatment of intermediate- or high-risk myelofibrosis.

Poster Presentation at the ISPOR 17th Annual International Meeting; June 2-6, 2012; Washington, DC, USA