Second Annual Workshop on Clinical Outcome Assessments in Cancer clinical trials

April 25, 2017 ■ Bethesda, MD

Co-sponsored by

FDA U.S. FOOD & DRUG ADMINISTRATION

CRITICAL PATH INSTITUTE
collaborate · innovate · accelerate

Session 3
Analysis and Display of PRO-Based Tolerability Data – Metrics and Paths Forward

SECOND ANNUAL WORKSHOP ON

CLINICAL OUTCOME ASSESSMENTS IN CANCER CLINICAL TRIALS

April 25, 2017 ■ Bethesda, MD

Co-sponsored by

FDA U.S. FOOD & DRUG ADMINISTRATION

CRITICAL PATH INSTITUTE
collaborate · innovate · accelerate

Disclaimer

- The views and opinions expressed in the following slides are those of the individual presenters and should not be attributed to their respective organizations/companies, the U.S. Food and Drug Administration or the Critical Path Institute.
-

Session 2 was about Assessment
Session 3 is about Analysis and Display

- Analyzing and describing symptomatic (or not) adverse events and tolerability data
  - Descriptive Analyses (not hypothesis testing, not dose determination)
- Systematically obtained PRO data
- Communicate while minimizing misinterpretation

(Future) Objectives

- Define the analysis population and discuss missing data/completion rates
- Longitudinal PRO analysis methods
  - Are there differences between safety, tolerability, and efficacy objectives and how impact analysis and communication
- Discuss various analysis and presentation examples using a common dataset

- Ultimately…Description of how patients experience symptomatic side effects of a therapy

This Session is a Little (Lot) Different

- Standard dataset given to all speakers and panelists to try ideas on
  - What clinical questions can we answer with different analyses?
  - What is the clinical interpretability and utility of each visualization?
  - What assumptions must be made for each analysis?
  - What are statistical strengths and limitations of each analysis?

Single Data File Given to All Speakers

- Compare methods and visualizations
- But what data to use?
  - Ask a statistician, get simulated data
  - Inherent properties and ties between variables and time points?
- We know what variables we tend to see, we know some data structure, we can manufacture data
  - Caution: we did not have time to make a huge great dataset (not datapalooza quality)
  - In fact one question we asked speakers was what other data do you wish you had to do or make the analyses and pictures you want (like CONMED)
  - Did not try to make data to prove our analysis points (although wish we had time to)
  - Generated from study data but altered to avoid identification
  - Data do NOT truly reflect real patient experience

PRO-CTCAE Example Data: Study X

- Randomized trial
- Evaluated efficacy and safety of Drug Y in 200 patients with metastatic/advanced cancer
  - N = 100 in Arm 0
  - N = 100 in Arm 1
- PRO-CTCAE

Study X: NCI- PRO-CTCAE Items (2 items, one with branching…how to score/interpret that?)

1. **PRO-CTCAE™ Symptom Term:** Diarrhea				
In the last 7 days, how OFTEN did you have LOOSE OR WATERY STOOLS (DIARRHEA)?				
☐ Never	☐ Rarely	☐ Occasionally	☐ Frequently	☐ Almost constantly

2. **PRO-CTCAE™ Symptom Term:** Fatigue				
In the last 7 days, what was the SEVERITY of your FATIGUE, TIREDNESS, OR LACK OF ENERGY at its WORST?				
☐ None	☐ Mild	☐ Moderate	☐ Severe	☐ Very severe
In the last 7 days, how much did FATIGUE, TIREDNESS, OR LACK OF ENERGY INTERFERE with your usual or daily activities?				
☐ Not at all	☐ A little bit	☐ Somewhat	☐ Quite a bit	☐ Very much

PRO-CTCAE Example "Data": Study X

- PRO-CTCAE
  - Two symptoms
  - "Collected" every week for the first 8 weeks of the study, and every 4 weeks thereafter
    - Weeks 1-8, 12, 16, 20, 24
- Additional "information"
  - Week of death
  - Week of discontinuation from treatment
  - Week of progression
  - If speakers wanted to make up more "information" on their own they could

But now we can compare analyses and data displays a bit easier

- Concentrate discussion around the different analyses and design considerations with respect to:
  - What clinical questions can we answer with different analyses?
  - What assumptions must be made for each analysis?
  - What are statistical strengths and limitations of each analysis?
  - What is the clinical interpretability and utility of each visualization?

Session Participants

- **Chair**
  - *Laura Lee Johnson, PhD* – Deputy Director, Division of Biometrics III, Office of Biostatistics (OB), Office of Translational Sciences (OTS) , CDER, FDA
- **Presenters**
  - *Yolanda Barbachano, PhD* – Senior Statistical Assessor, Licensing Division, Medicines and Healthcare Products Regulatory Agency
  - *Mallorie H. Fiero, PhD* – Mathematical Statistician, Division of Biometrics V, OB, OTS, CDER, FDA
  - *Diane L. Fairclough, DrPH* – Professor, Biostatistics, Colorado School of Public Health
- **Panelists**
  - *Corneel Coens, MSc* – Lead Statistician, QoL Department, EORTC
  - *Joseph C. Cappelleri, PhD, MPH, MS* – Senior Director of Biostatistics, Pfizer Inc
  - *Sandra A. Mitchell, PhD, CRNP* – Research Scientist and Program Director, Outcomes Research Branch, NCI, NIH
  - *Paul G. Kluetz, MD* – Acting Associate Director of Patient Outcomes, OCE, FDA

Regulatory considerations in Europe for trials collecting PRO data on tolerability

Yolanda Barbachano, PhD

Senior Statistical Assessor

MHRA

FDA U.S. FOOD & DRUG ADMINISTRATION

CRITICAL PATH INSTITUTE

collaborate · innovate · accelerate

Background from an EU regulatory perspective – PRO guidance

- Current version of the EMA's Reflection Paper on Health Related Quality of Life measures came into effect in January 2006 and focuses on HRQL only.

- At an EMA level there is currently no regulatory guidance on the use of patient reported outcomes in general.
- Apart from an appendix to the oncology guideline

- But there is interest within the system to try and produce such guidance

PRO measures – problems so far

- Mostly focused on QoL
- Lack of rationale for instrument selection
- Lack of up front hypothesis and prospective objectives
- Data collection not prioritised
- Amount of missing data made results uninterpretable
- Analysis not adequately planned

Suggested new PRO guideline

- Would outline broad principles of scientific best practice and methodology
- Provide guidance on the value of PRO data in the development of medicinal products
- With the aim of better supporting data generation for specific label claims

- Would not explain how to validate PRO instruments

Background from an EU regulatory perspective – Tolerability assessment

- EMA anticancer guideline is currently being revised to expand safety section
- So far tolerability has been based on
  - Frequency and severity of AEs
  - Withdrawals from treatment
  - Dose reductions
  - Treatment interruptions
  - Hospitalisations

- Not based on any PRO instruments (up to now)

However…

- PRO appendix to oncology guideline (2016) mentions:

At the time of this appendix, there is no EMA/CHMP experience from the use of the NCI's PRO version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) and more experience is needed before these tools can be used reliably. However, it is envisioned that the PRO-CTCAE could enhance the precision and patient centeredness of adverse event reporting in cancer clinical research and ultimately provide a more representative account of patients' treatment experiences.

Tolerability assessment

- EU medical assessors used to CTCAE data presented in tables (e.g. proportion of patients with the event, or limited to grade 3/4)
- Could potentially assess PRO CTCAE similarly
- OR for assessment of tolerability could combine old and new ideas:

  - What happened at the time of dose reduction/ interruption?
  - Or withdrawal from treatment?

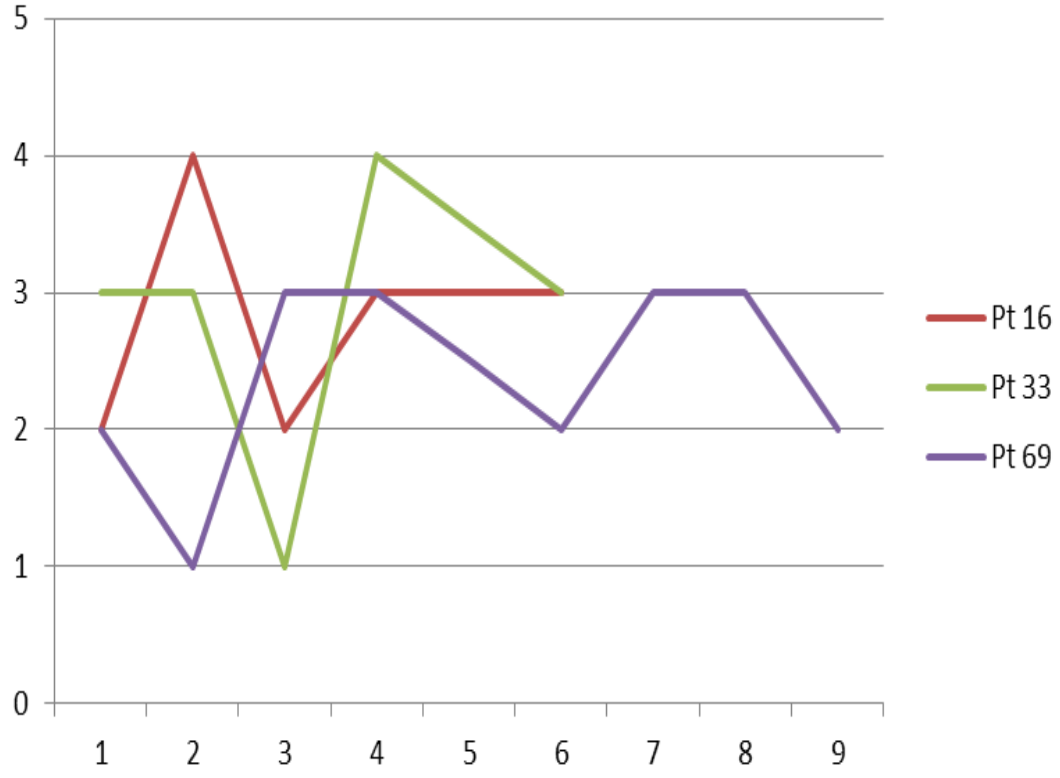
  - Look at PRO-CTCAE at the time of event (or leading to it)
  - Trend across patients?
  - How does it compare to the rest of the trial population?

Discontinued Patients – Fatigue severity

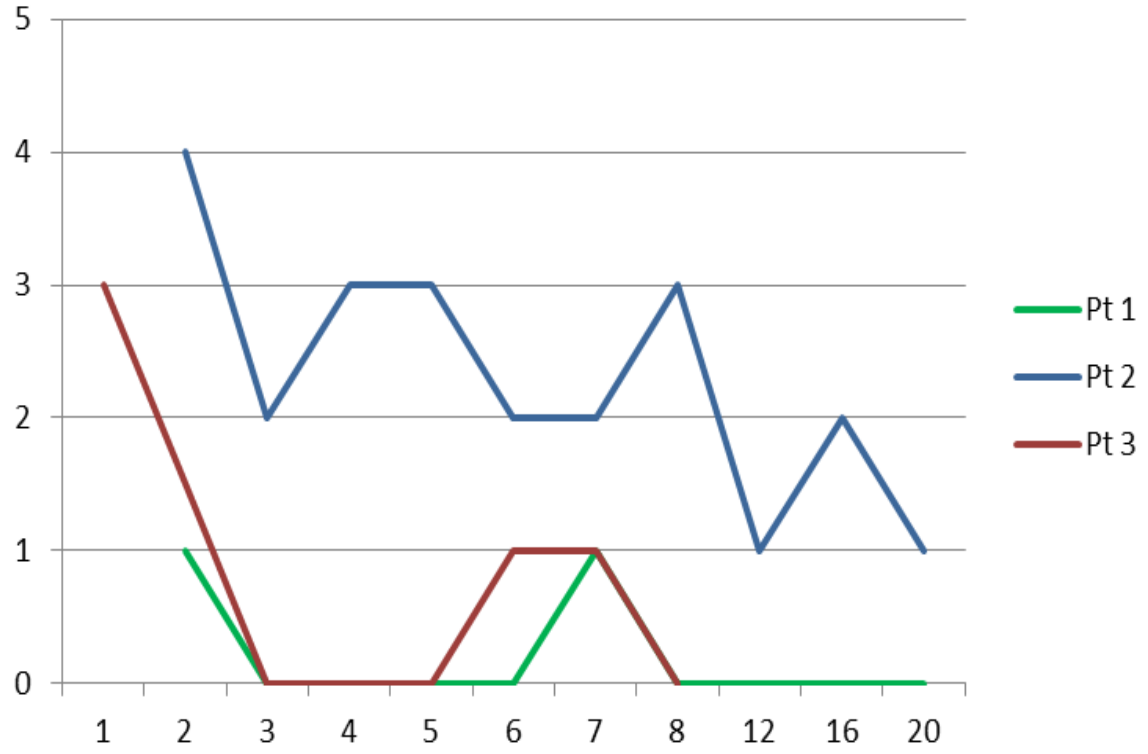
Patient No	Week 1	2	3	4	5	6	7	8	12	16	20
1		1	0	0	0	0	1	0	0	0	0
2		4	2	3	3	2	2	3	1	2	1
3	3		0	0	0	1	1	0			
16	2	4	2	3	3	3	DISC				
33	3	3	1	4		3	DISC				
69	2	1	3	3		2	3	3	DISC		

Discontinued Patients – Fatigue severity

Discontinued patients

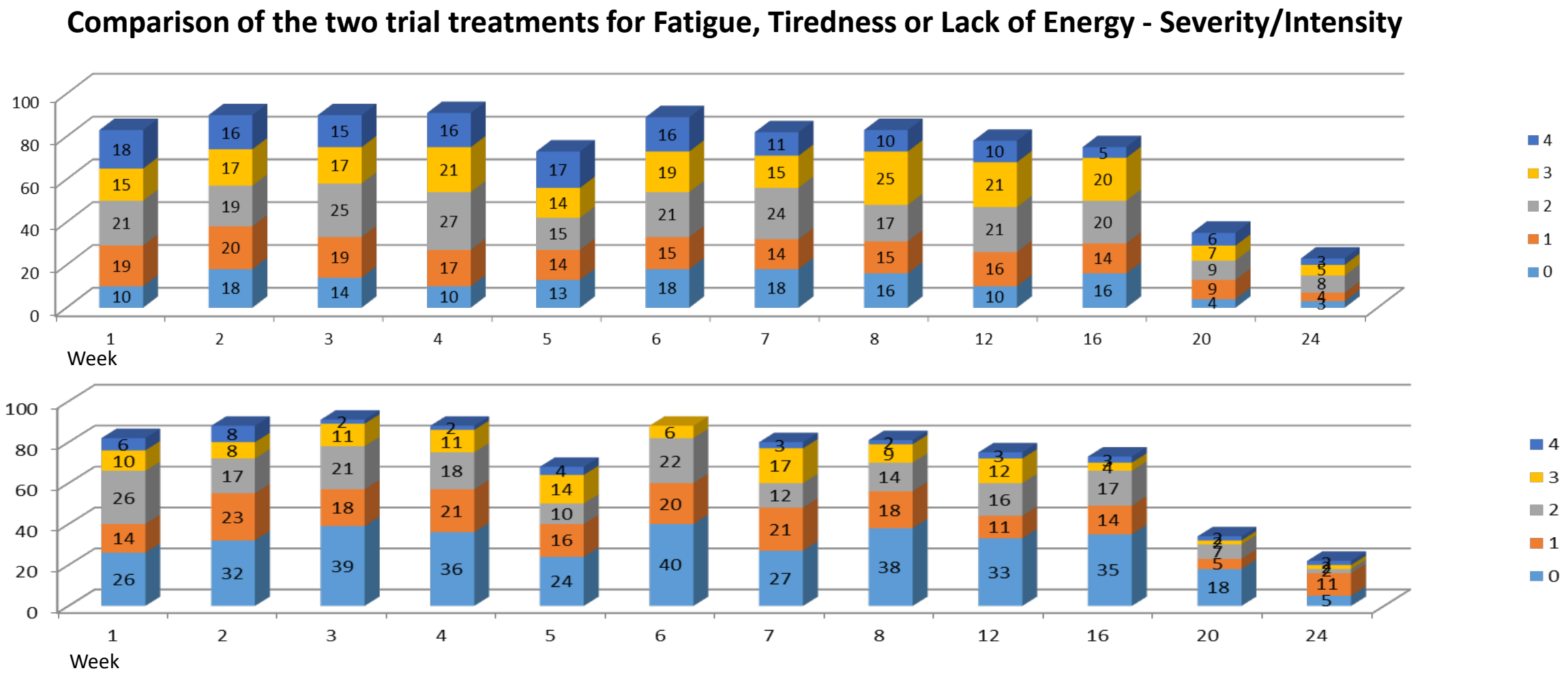


Patients completing treatment



Interestingly patients 1 and 3 were on a different treatment to the others

All patients - visualise each AE over time



Estimands, PRO-CTCAE and tolerability

- What are we trying to estimate?
- Generally more than one symptom needed to assess tolerability
- But if we try to collect too much info… are we at risk of going backwards to uninterpretable data?
- Consider patient burden and missing data
- Use phase II to inform what symptomatic AEs to focus on phase III
- Furthermore use phase II to decide how to analyse phase III
- Different symptomatic AEs may require different analyses

References

- Guideline on evaluation of anticancer medicinal products in man (EMA/CHMP/205/95 Rev.5) – revised draft

- Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man - the use of patient-reported outcome (PRO) measures in oncology studies (EMA/CHMP/292464/2014)

- Reflection paper on the regulatory guidance for the use of Health Related Quality of Life (HRQL) measures in the evaluation of medicinal products (EMEA/CHMP/EWP/139391/2004)

Defining the population, missing data vs. completion rate, and presentation methods

Mallorie H. Fiero, PhD

Mathematical Statistician

FDA, CDER, Office of Biostatistics, DBV

FDA U.S. FOOD & DRUG ADMINISTRATION

CRITICAL PATH INSTITUTE

collaborate · innovate · accelerate

Outline

What is the best way to present PRO-based tolerability data over time in a way that is not misleading?

- Analysis population
  - Are the data truly missing? A case study
- Visualizing patient experience while on therapy
  - Missing data
  - Accounting for baseline response

Objective: Describe tolerability while patients are on treatment

- May not be traditional ITT population
- Generalizability?

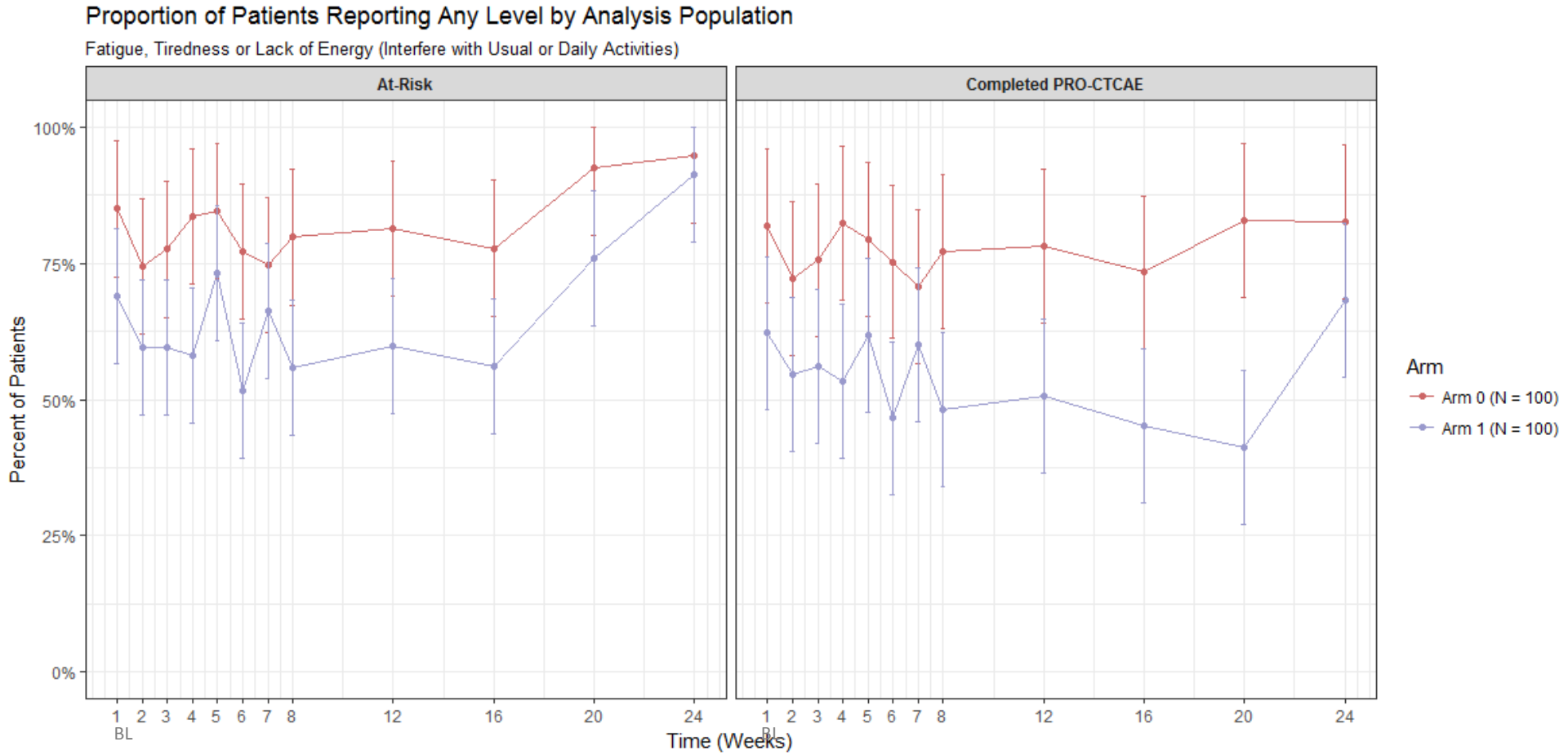
Are the data truly missing?

Case	Death	Progression	Discontinuation	Intermittent Missing
Example	By week 7, 5 (2.5%) patients died	By week 7, 3 (1.5%) patients progressed and have moved on, in hospice, etc.	By week 7, 2 (1%) patients discontinued treatment	At week 7, 28 (14%) patients had intermittent missing data
Denominator	Exclude	Exclude (complicated)	Exclude (complicated)	Include
Reason	Not at-risk for PRO-CTCAE event after death	Moved on to new therapy No change in therapy → Include	Withdrew consent or moved on to new therapy No change in therapy → Include	On study and at-risk for PRO-CTCAE event

Analysis population: at-risk patients

- Death/progression/discontinuation
  - Cannot assume that patients who dropped out are similar to the patients observed

- Denominator = Number of patients on study and considered at-risk



Choice of analysis population affects interpretation

- The proportion of patients reporting any level of fatigue over time differs by the analysis population (denominator)
- Perform sensitivity analysis for missing data

Overall response

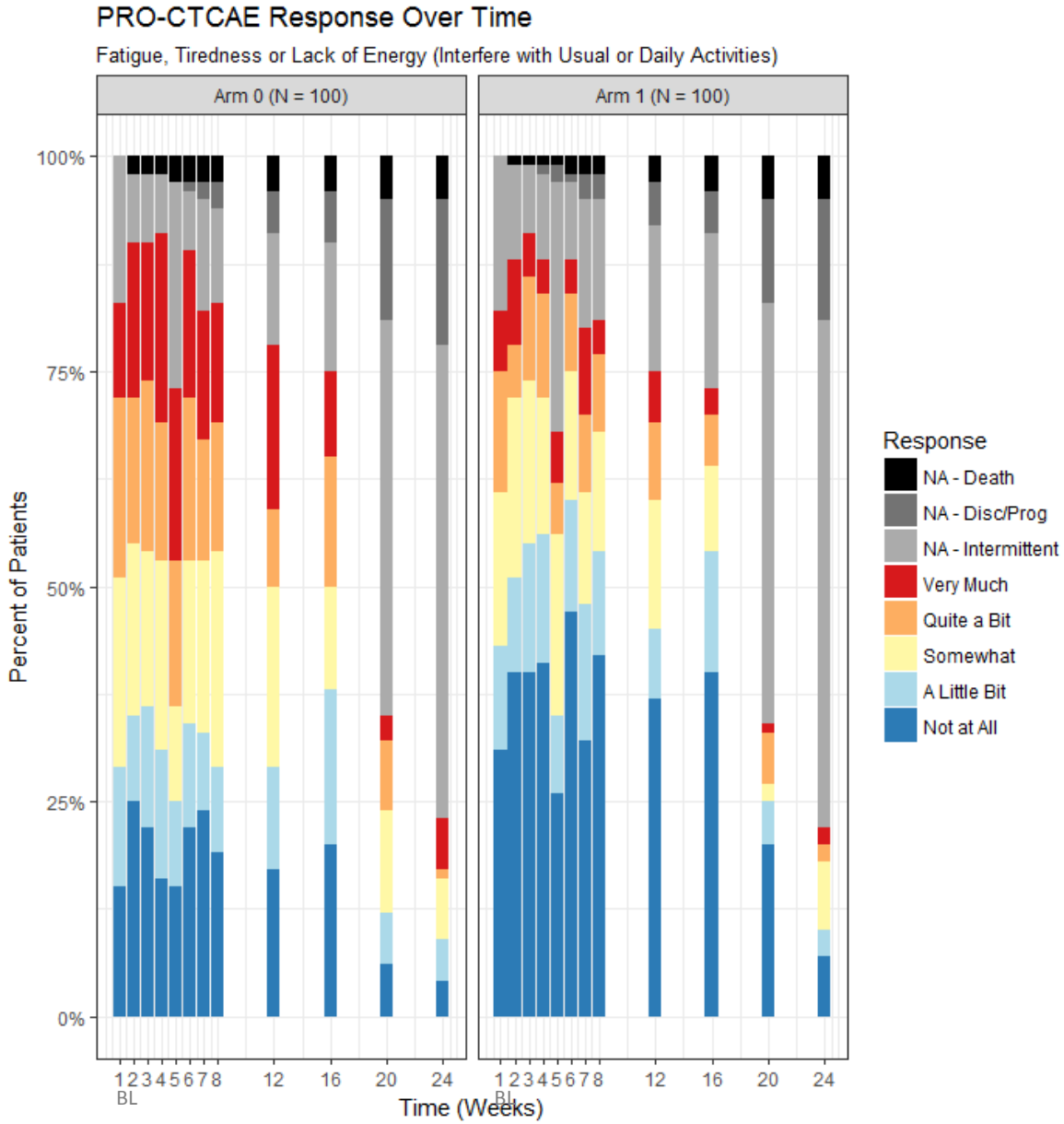
Positives

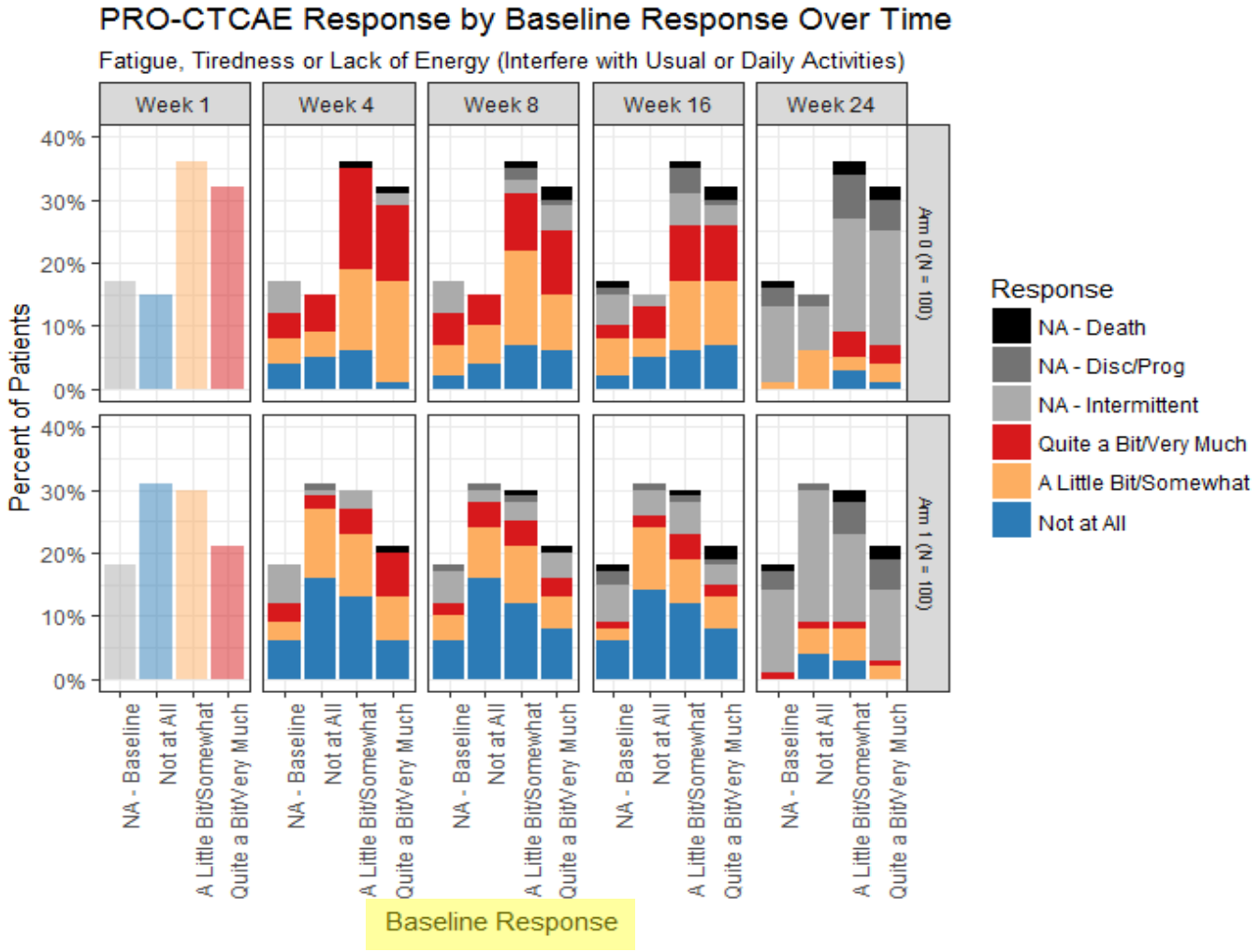
- Overall snapshot
- NA due to death, discontinuation/progression, intermittent missing are differentiated

Negatives

- Were the patients who reported “Very Much” for fatigue at Week 1 the same patients who reported “Very Much” at subsequent time points?
  - Does not take baseline into account
  - Cannot see where patients are moving over time
- Difficult to do a direct comparison of all responses between treatment arms

Not a good idea…





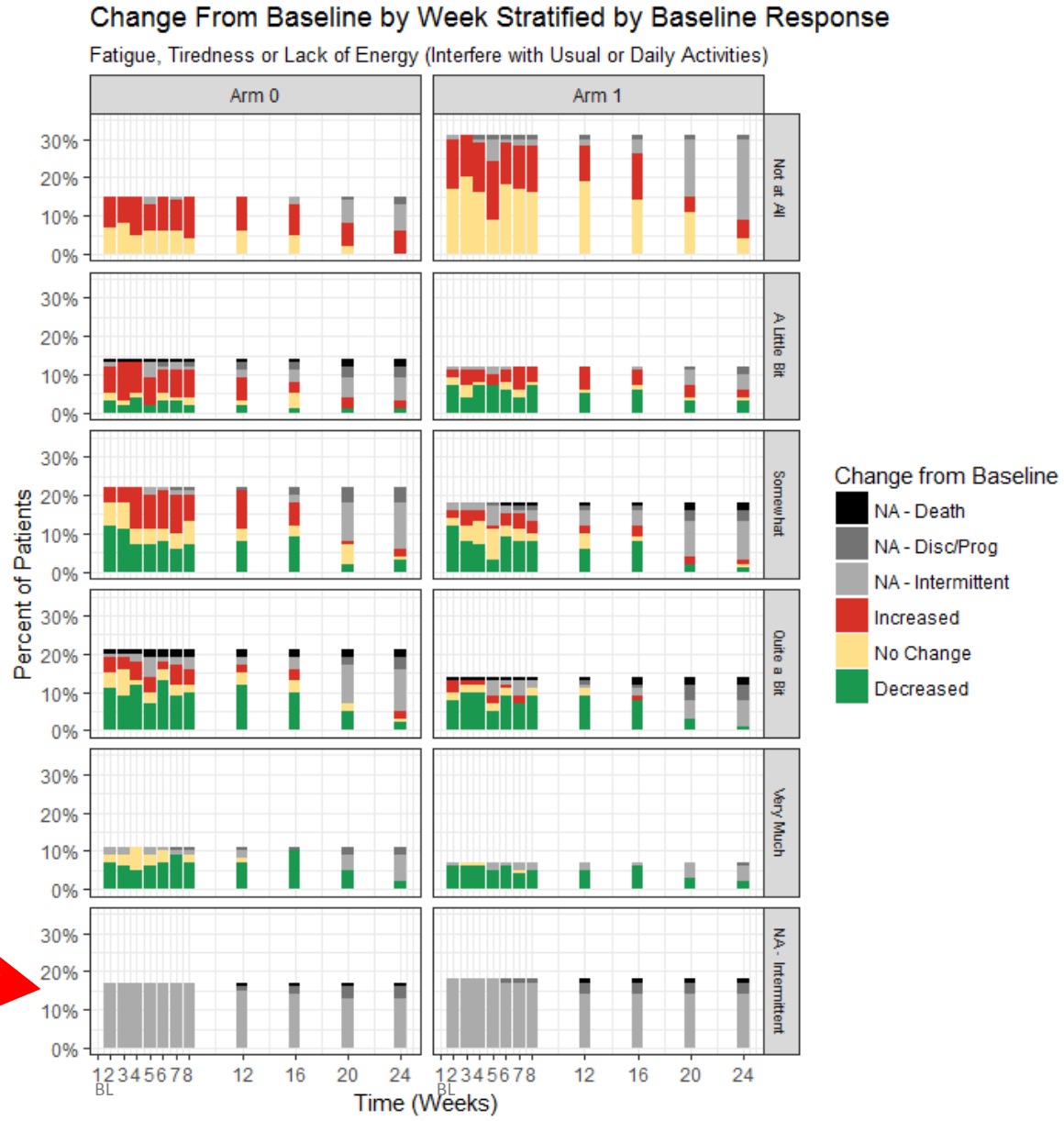
Overall response by baseline

- More informative way to visualize patient experience while on therapy
- Shows movement of response from baseline
- Still has some of the same issues as overall response stacked bar chart

Change from baseline

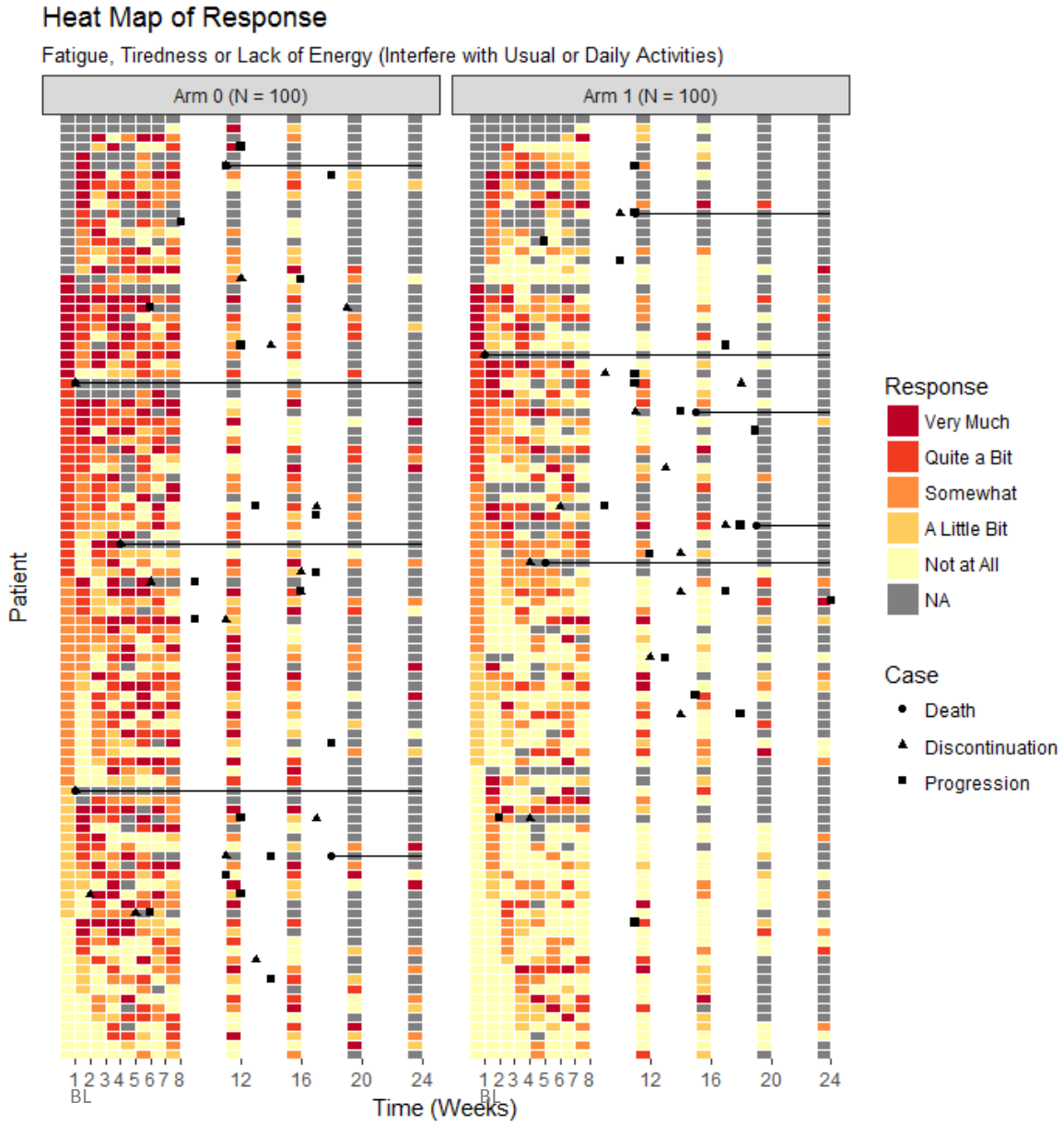
- Still has some of the same issues as overall response stacked bar plot
- Response decreases over time due to dwindling numbers

- Missing values are important to include in data visualization



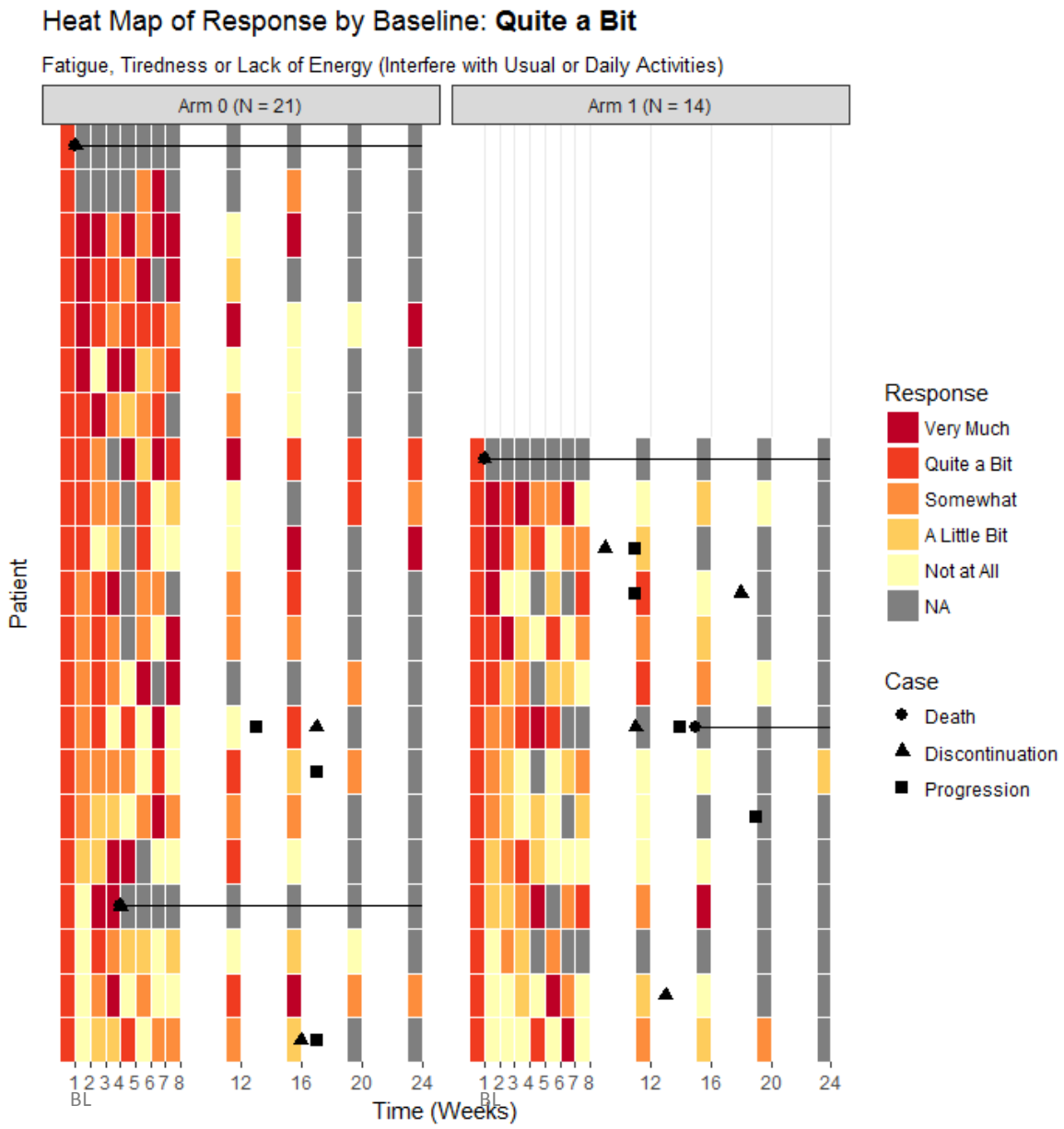
Individual response over time

- Visual representation of individual-level PRO-CTCAE response over time
- More difficult to interpret
- Cannot summarize which treatment was more tolerable with respect to fatigue



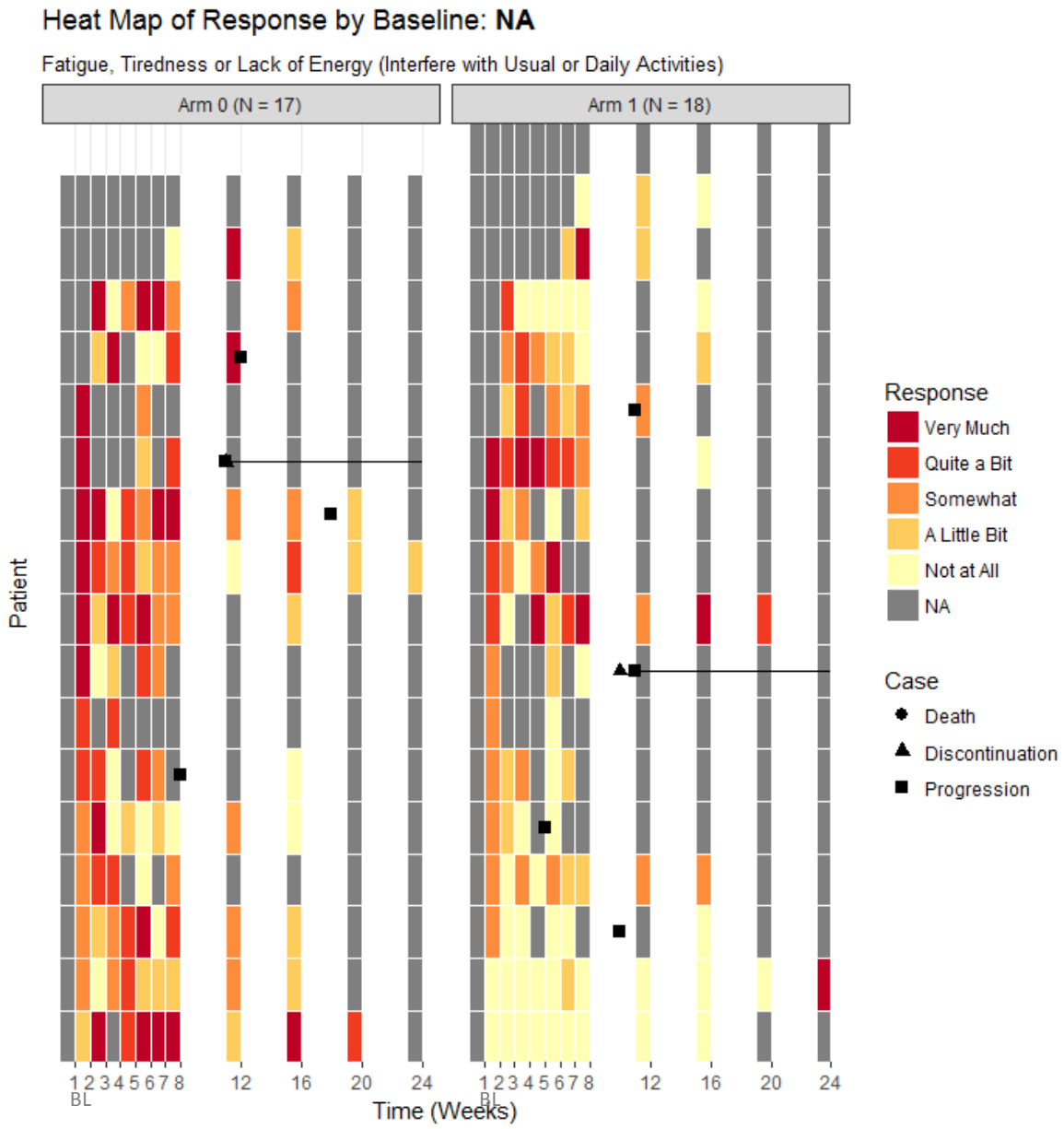
A closer look at individual response by baseline

- Patients who died, discontinued, or progressed generally had more severe fatigue



Patients with missing baseline data

- Patients with missing response at Week 1 had more severe fatigue at subsequent weeks.
- Incorporate patients who have missing baseline data
- Otherwise, lose important information
- Sensitivity analysis for missing data



Conclusions

What is the best way to present PRO-based tolerability data over time in a way that is not misleading?

- Choice of analysis population impacts interpretation of patient experience
  - At-risk population
- Take baseline into account in data visualization and analysis
  - Informative of patient experience while on therapy
- Deal with missing data appropriately
  - Include missing data
  - Incorporate patients with missing baseline data
  - Perform sensitivity analysis for missing data
- There is no best way. Further collaboration is needed for best practices in analysis and presentation.
- Only looked at one symptom. Patients' experience includes more than one symptom. Have not discussed presenting more than one symptom.

References

- National Research Council. (2011). *The prevention and treatment of missing data in clinical trials*. National Academies Press.
- US Department of Health and Human Services (USDHHS). Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. www.fda.gov/downloads/Drugs/Guidances/UCM19328
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.

Acknowledgments

- Laura Lee Johnson
- Jonathon Vallejo
- Paul Kluetz
- Rajeshwari Sridhara

DBV R Working Group

Analysis Strategies for PRO-CTCAE

Diane L. Fairclough, DrPH

Professor, Biostatistics

Colorado School of Public Health

Measuring Tolerability in Cancer Treatments

- Traditional Non-PRO indicators
  - Drop out from treatment
  - Significant delays in cycles
  - Dose reductions
- Limitations
  - Indirect; have other causes
  - Does not always identify source of problem
  - Borderline problems – pts hanging-in in context of trial, but might not in practice

PRO-CTCAE as a measure of tolerability

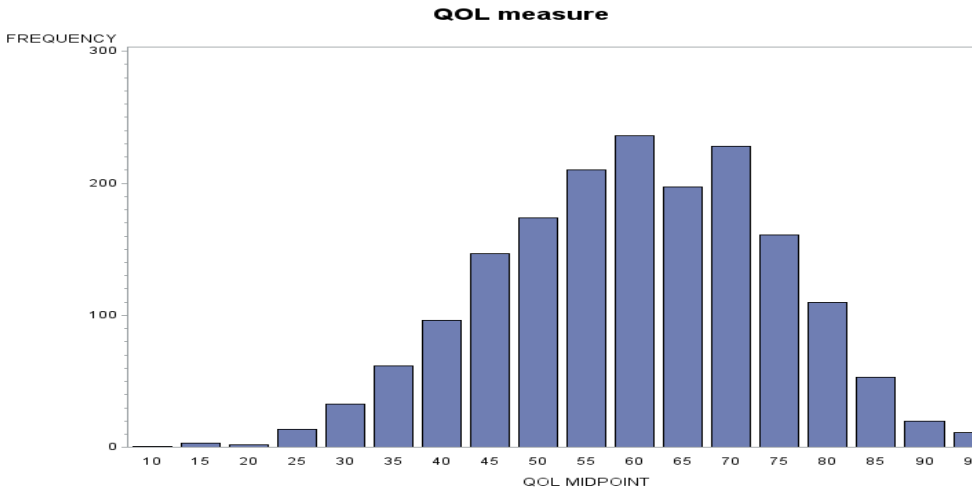
- Strengths
  - Patient reported
  - Covers symptoms/treatment toxicities likely to lead to dropout
- Weaknesses
  - Single item measures => low precision
  - Prevalence of problems may be low for some items
  - Single symptom vs multiple symptoms

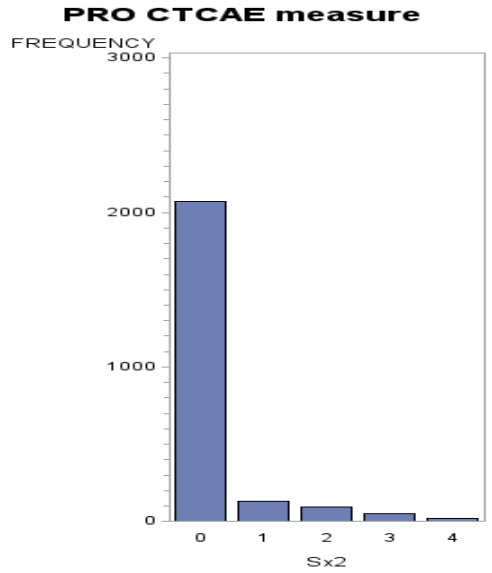
=> Analytic issues

How would the analysis of PRO-CTCAE differ from other longitudinal measures (e.g. HRQOL)?

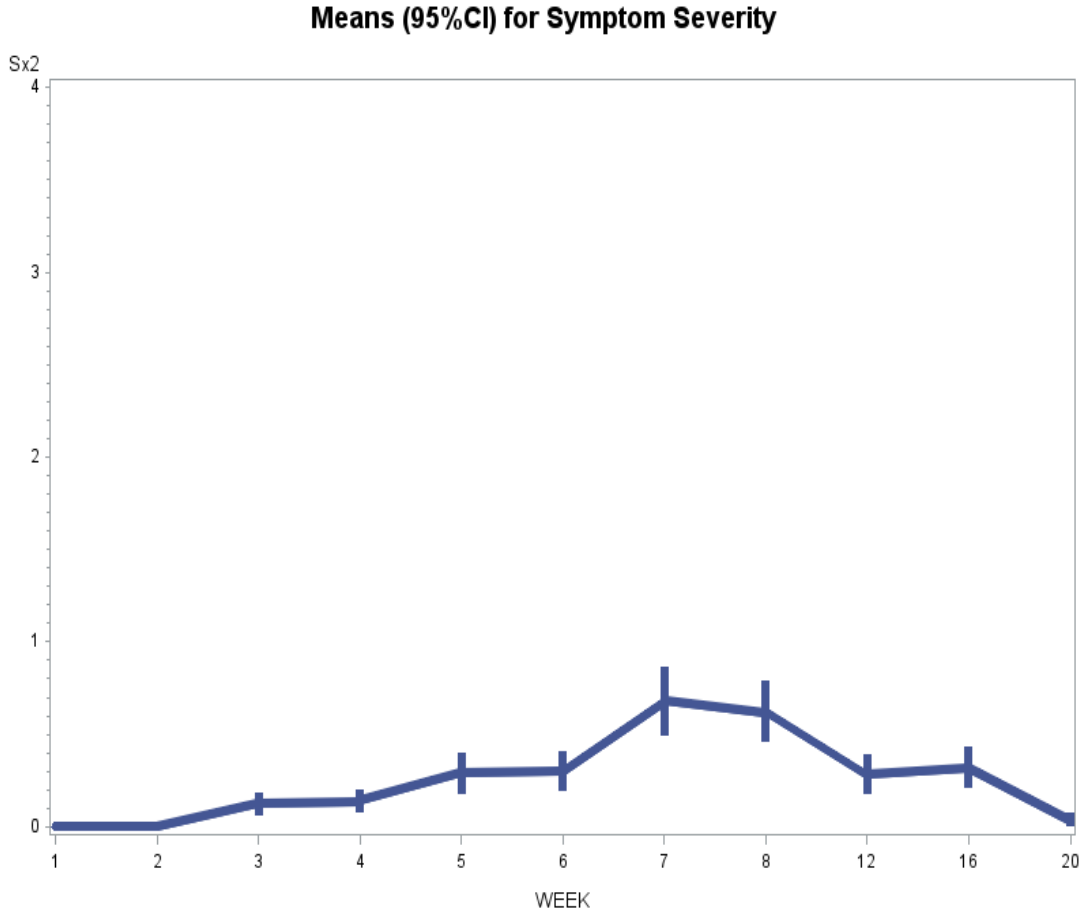
- HRQOL
  - Distribution roughly symmetric
  - Minimal floor/ceiling problems
  - Means are a reasonable measure of central tendency

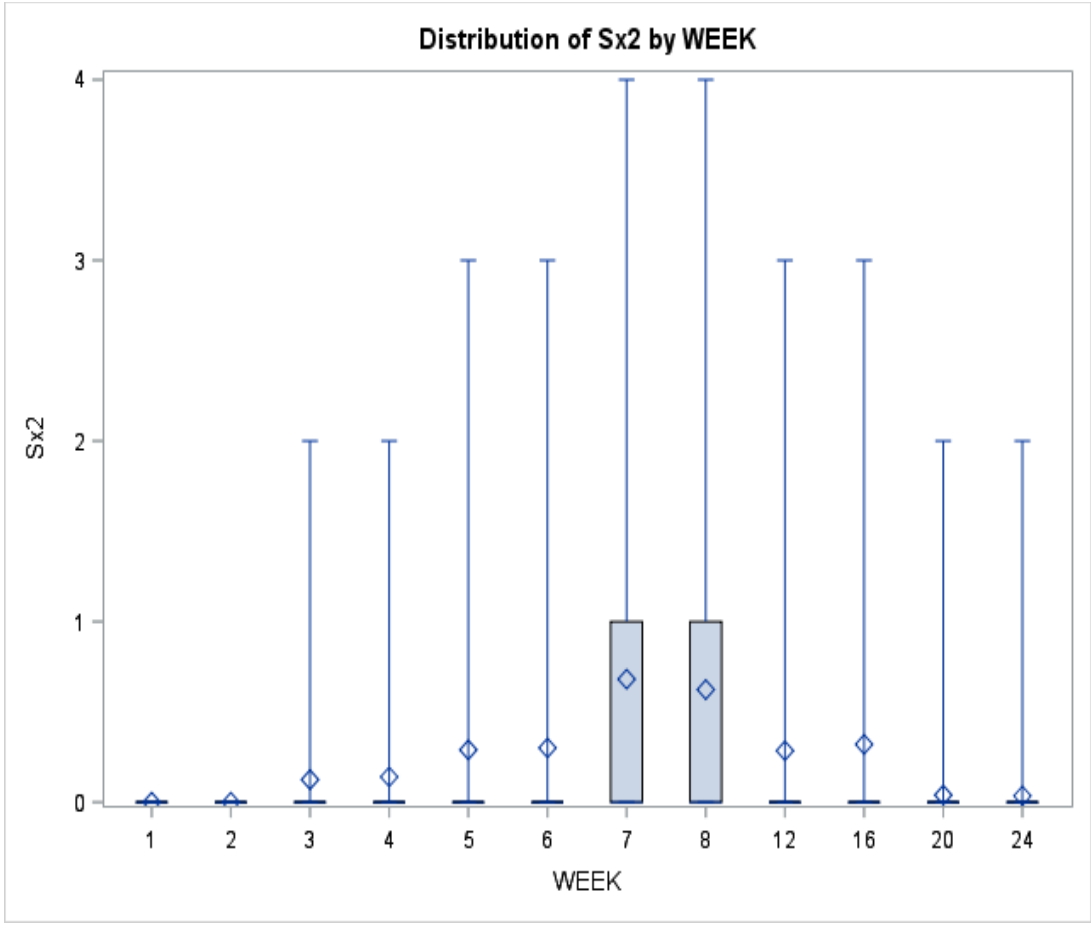
- PRO-CTCAE
  - Distribution very skewed for most Sx
  - High proportion of ‘none’
  - Means not a good measure of central tendency; median worse





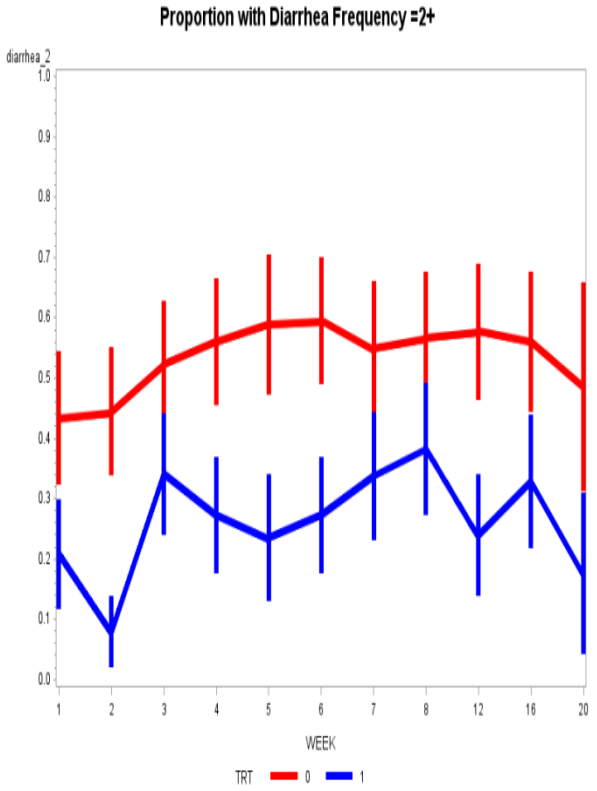
Presenting PRO-CTCAE results

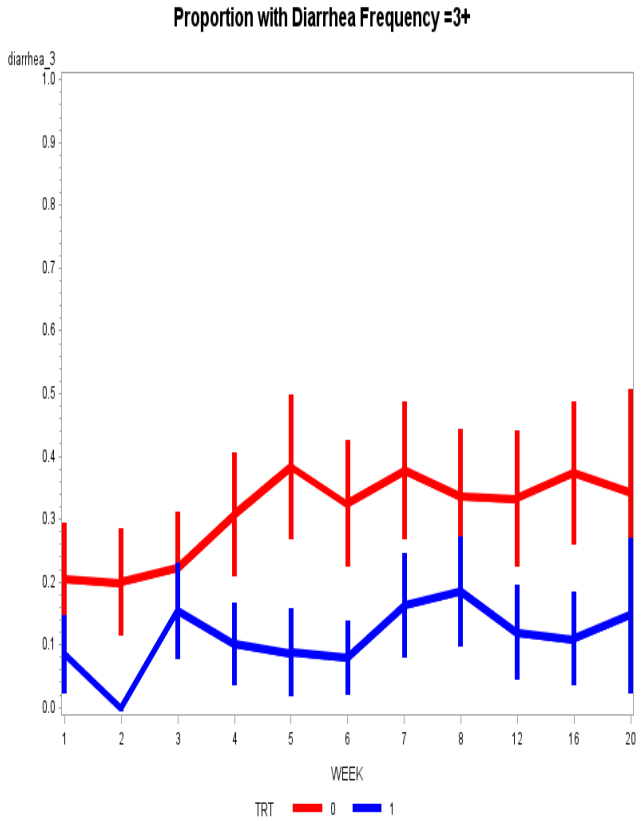


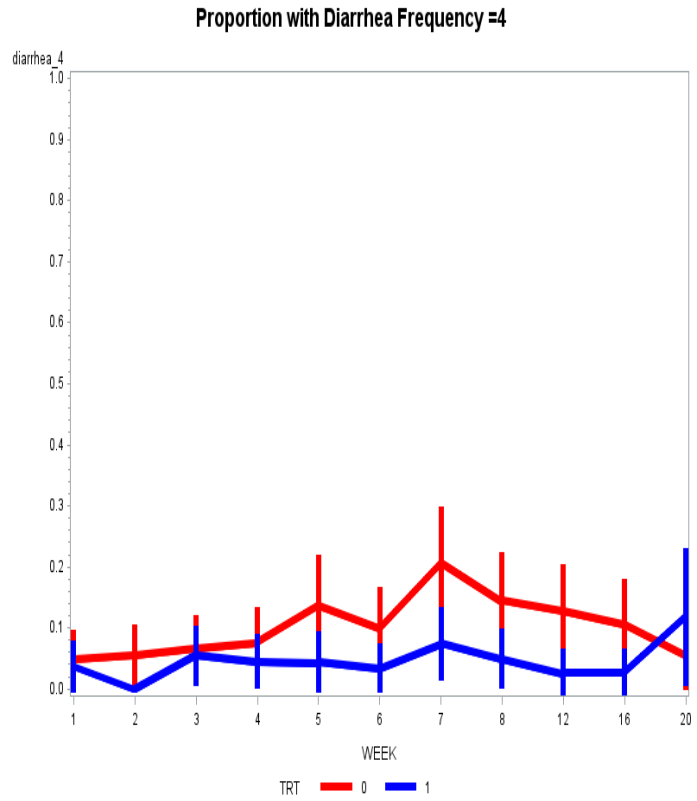


Proportions with specific levels of severity by time

- What cut off is relevant?/ How does dropout effect estimates?

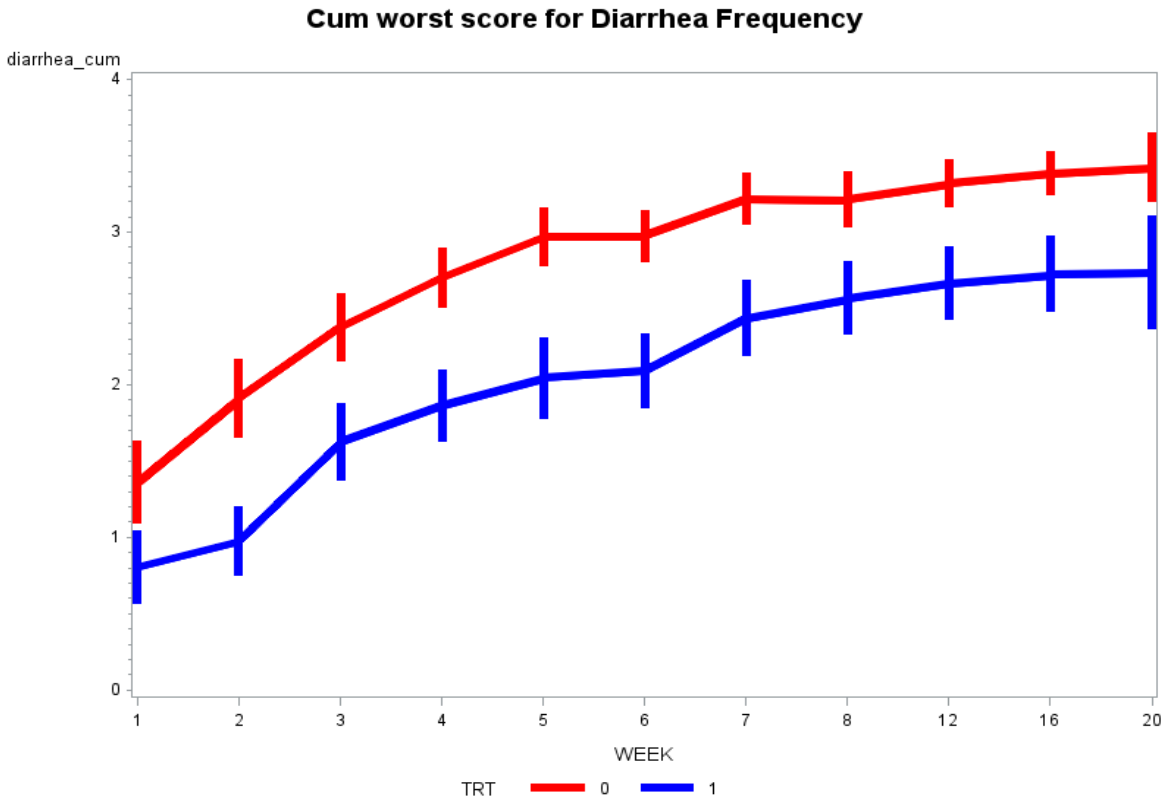


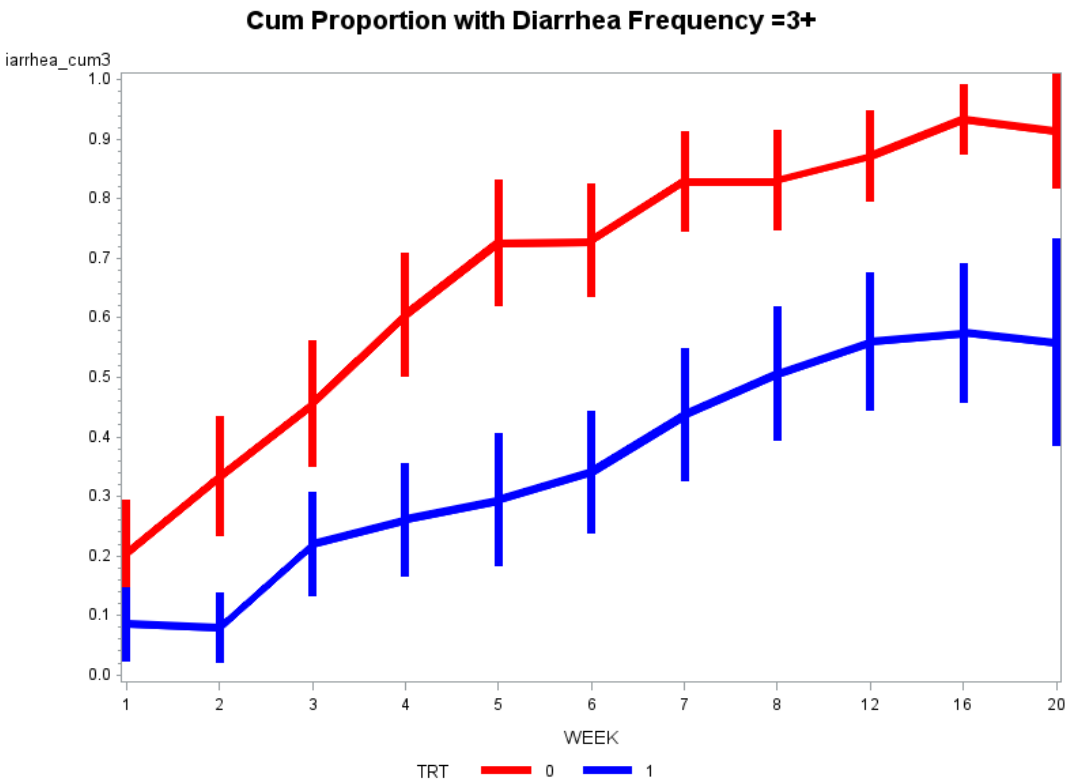




Cumulative worst score or proportions over time

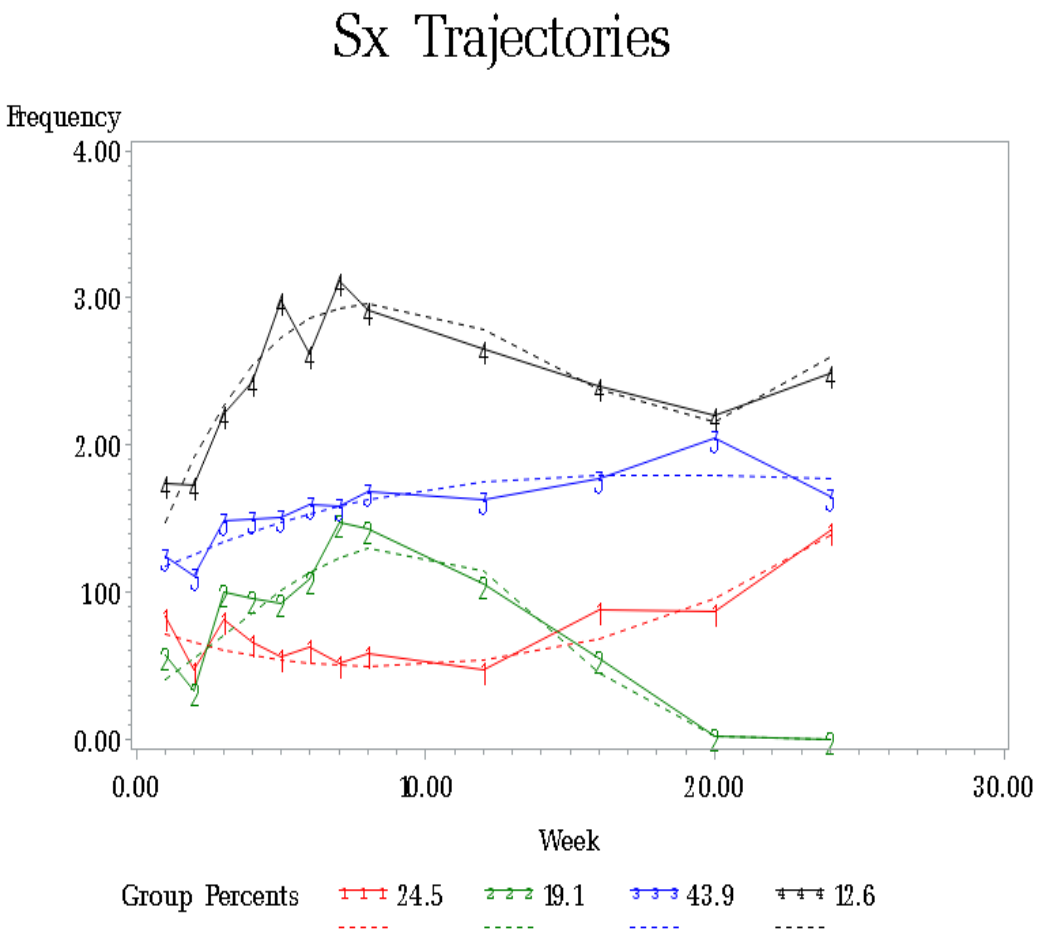
- Reflects cumulative issues/ Does not reflect transient symptoms





Individual vs group trajectories

- Latent class trajectory analysis
  - Identifies latent classes (groups of individuals) with similar trajectories
  - Individuals may be ‘classified’ into the latent class based on their probability of belonging



- Jones and Nagin (2007) Soc Methods and Research 35: 542-71
- Nagin and Odgers (2010) Annu. Rev. Clin. Psychol.. 6:109–38

Challenges

- Effect of ‘missing’ data due to discontinuing treatment ➔
  - Changing ‘population’
  - Patients experiencing serious Sx will drop out
  - May appear that *prevalence/severity* is decreasing over time
- Impact of adjuvant meds
  - Interpretation when meds such as anti-emetics are used
- Others???

Going forward

- Analysis strategies will depend on the GOALS

- Design of studies
  - Identifying the analytic strategy that fits with the goal
  - Selecting the ‘right’ Sx to measure
    - Known toxicities
    - Unknown toxicities

Panelist Comments

- **Chair**
  - *Laura Lee Johnson, PhD* – Deputy Director, Division of Biometrics III, Office of Biostatistics (OB), Office of Translational Sciences (OTS) , CDER, FDA
- **Presenters**
  - *Yolanda Barbachano, PhD* – Senior Statistical Assessor, Licensing Division, Medicines and Healthcare Products Regulatory Agency
  - *Mallorie H. Fiero, PhD* – Mathematical Statistician, Division of Biometrics V, OB, OTS, CDER, FDA
  - *Diane L. Fairclough, DrPH* – Professor, Biostatistics, Colorado School of Public Health
- **Panelists**
  - *Corneel Coens, MSc* – Lead Statistician, QoL Department, EORTC
  - *Joseph C. Cappelleri, PhD, MPH, MS* – Senior Director of Biostatistics, Pfizer Inc
  - *Sandra A. Mitchell, PhD, CRNP* – Research Scientist and Program Director, Outcomes Research Branch, NCI, NIH
  - *Paul G. Kluetz, MD* – Acting Associate Director of Patient Outcomes, OCE, FDA

Example graphical presentation of dummy COA data

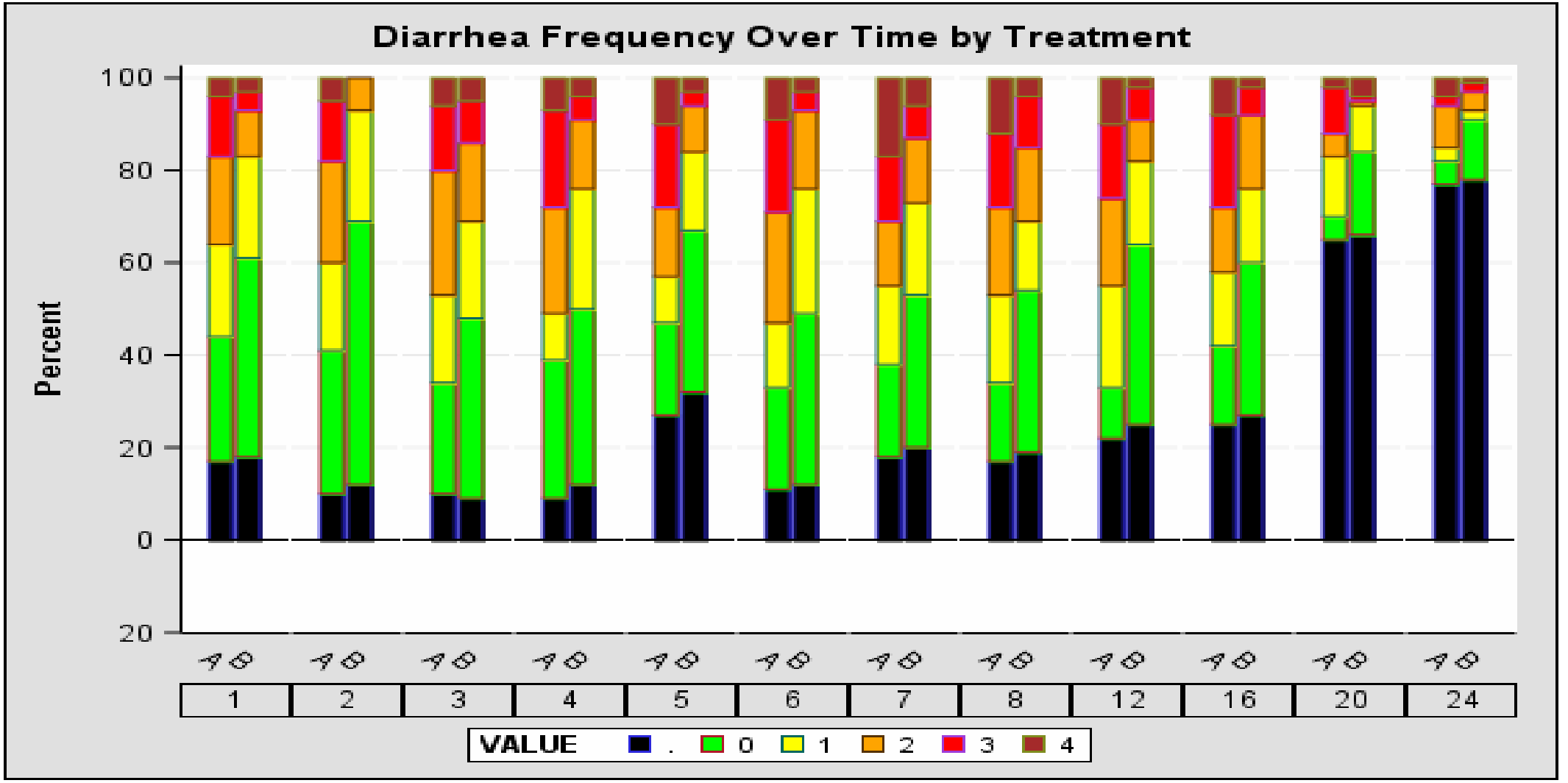
Corneel Coens

EORTC QoL Department

FDA U.S. FOOD & DRUG ADMINISTRATION

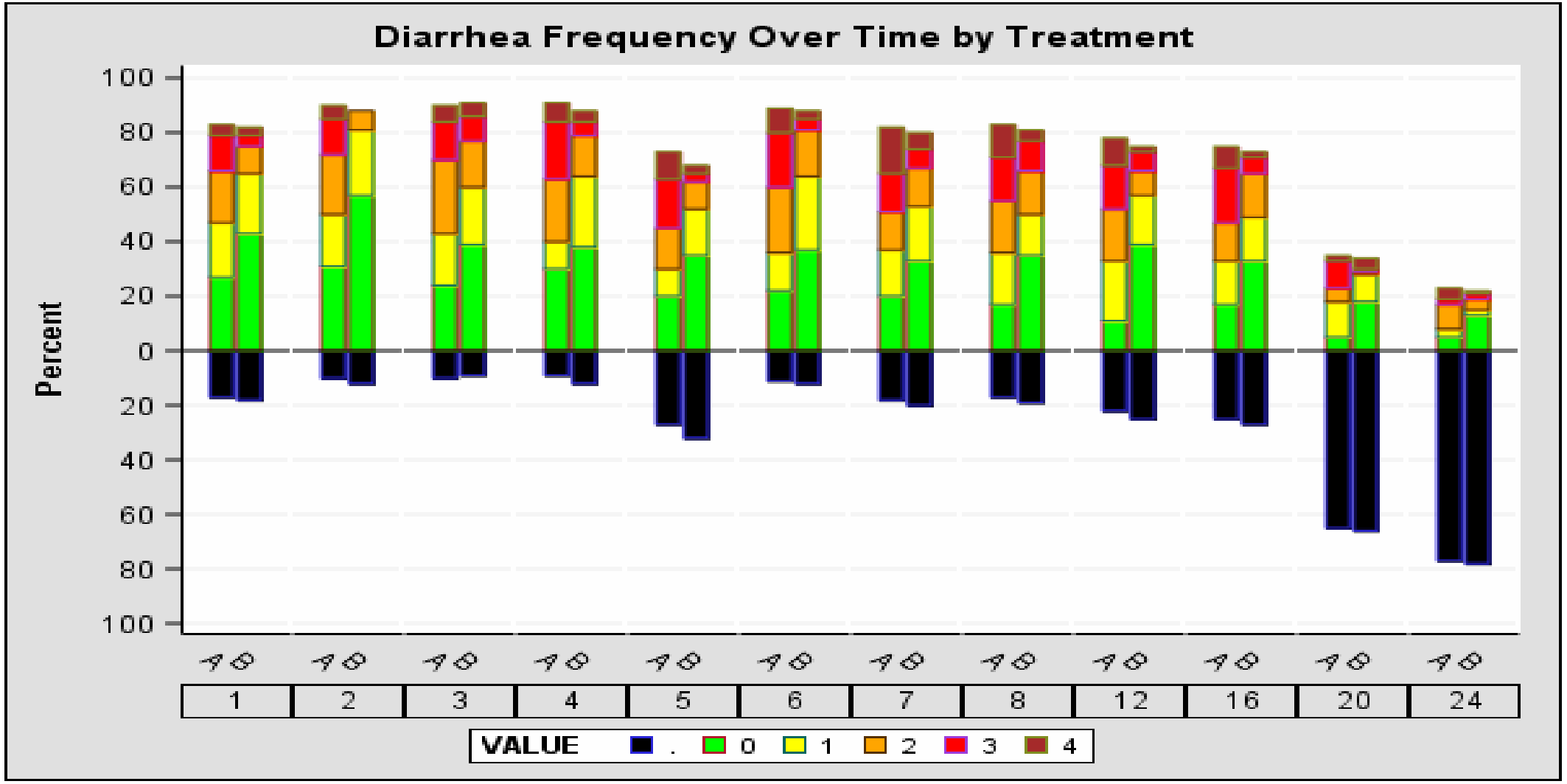
CRITICAL PATH INSTITUTE

collaborate · innovate · accelerate



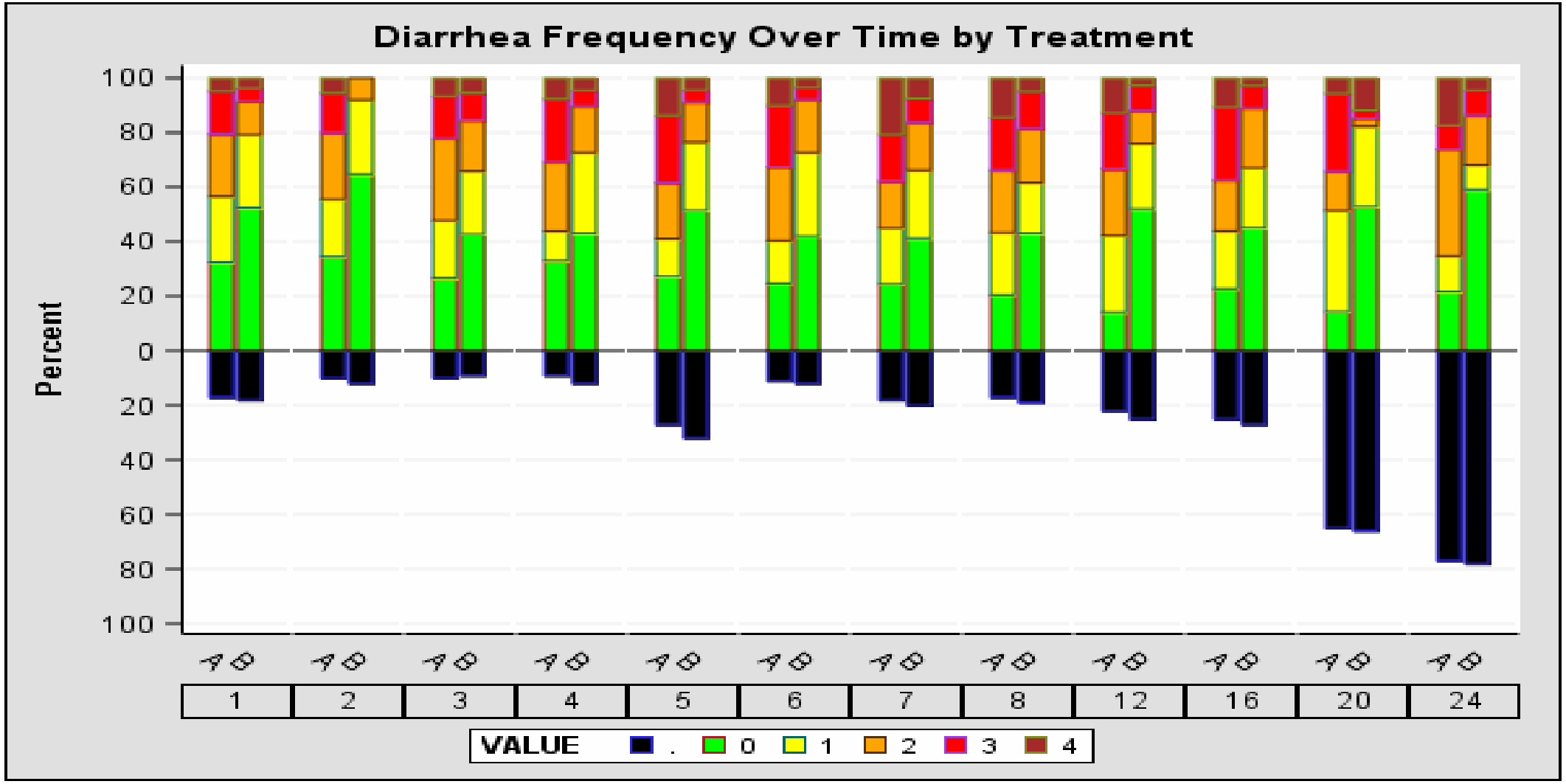
Classical % bar chart

- **Always summing to 100%**
- **Missing data as separate category**



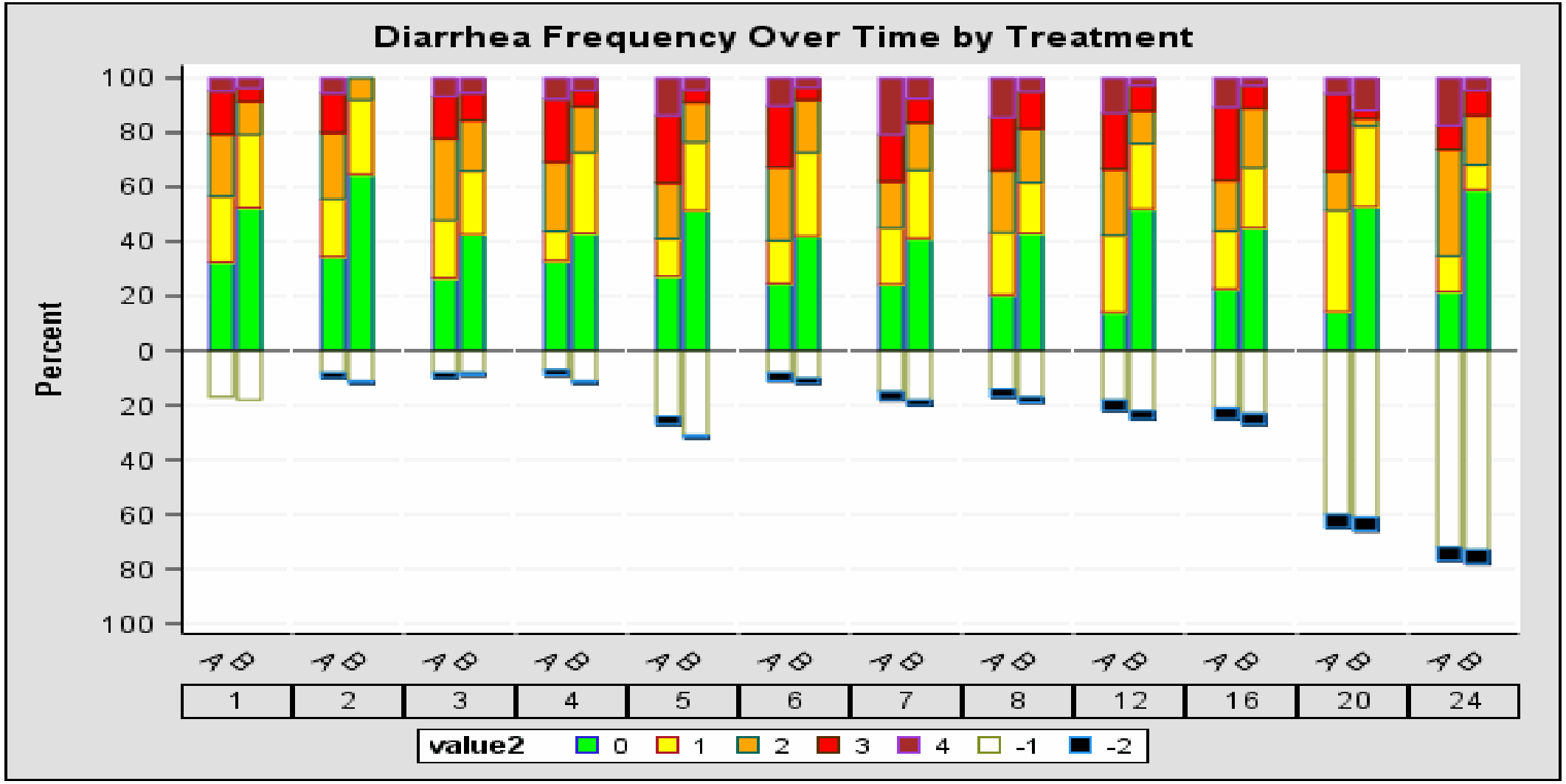
Modification:

- **Move missing category to lower end to denote attrition**
- **Comparison of actual outcome less obvious**



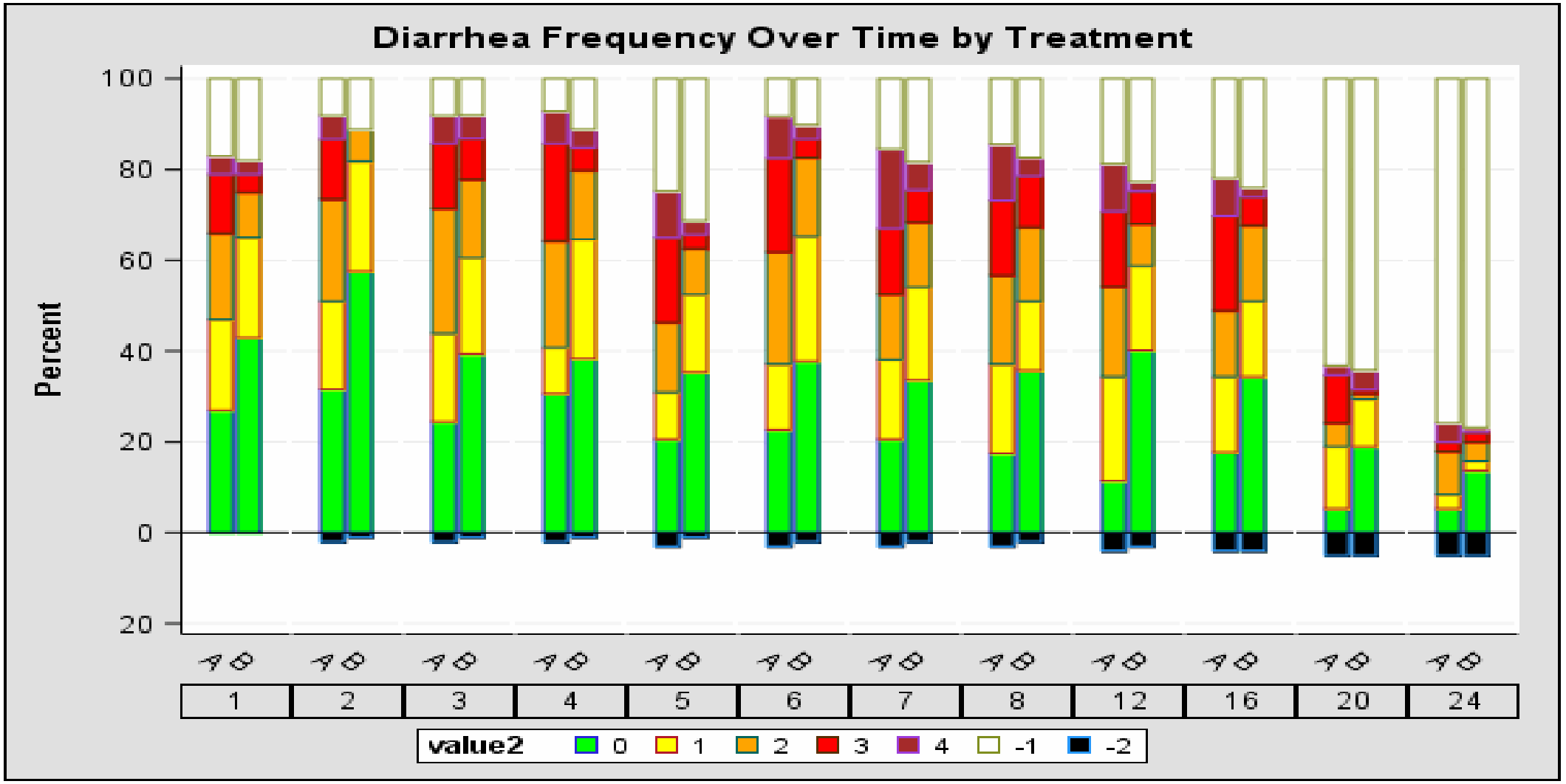
Modification:

- **Rescale actual outcomes to 100% - conditional on observation (ideally actual numbers added on graph)**
- **Overview of outcome and attrition over time.**



Modification:

- **Separate ‘missing’ in ‘non-compliance’ and ‘attrition (death)’.**
- **WHITE = non-compliance / BLACK = death**



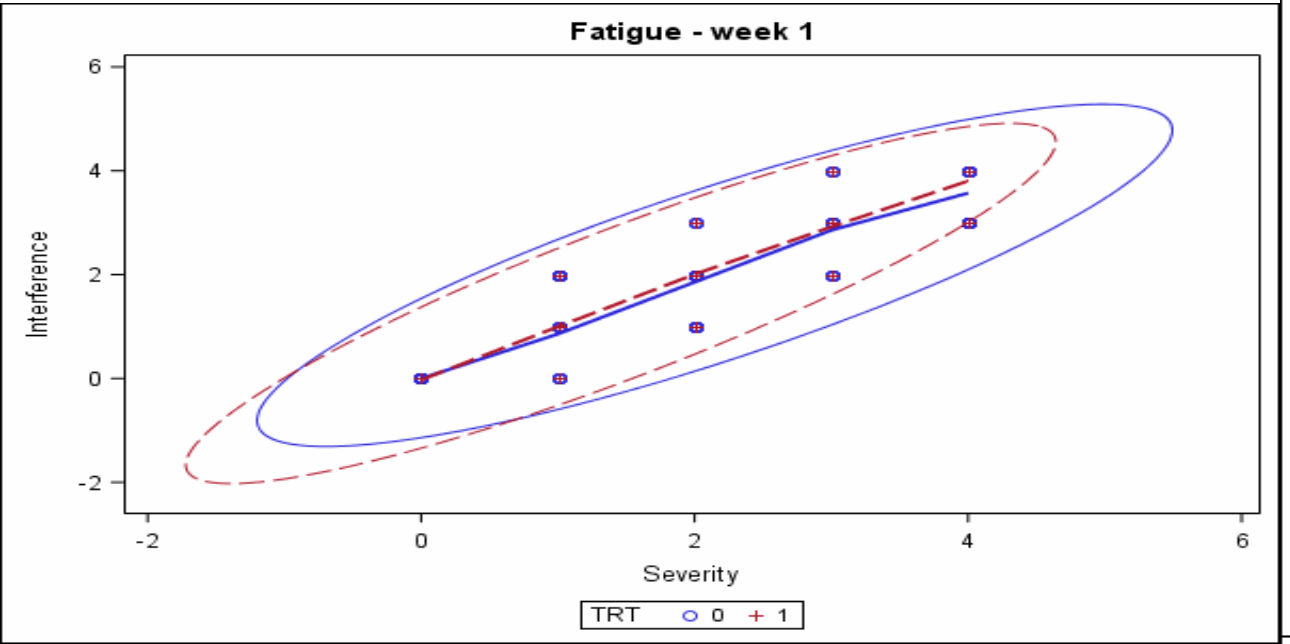
Final version:

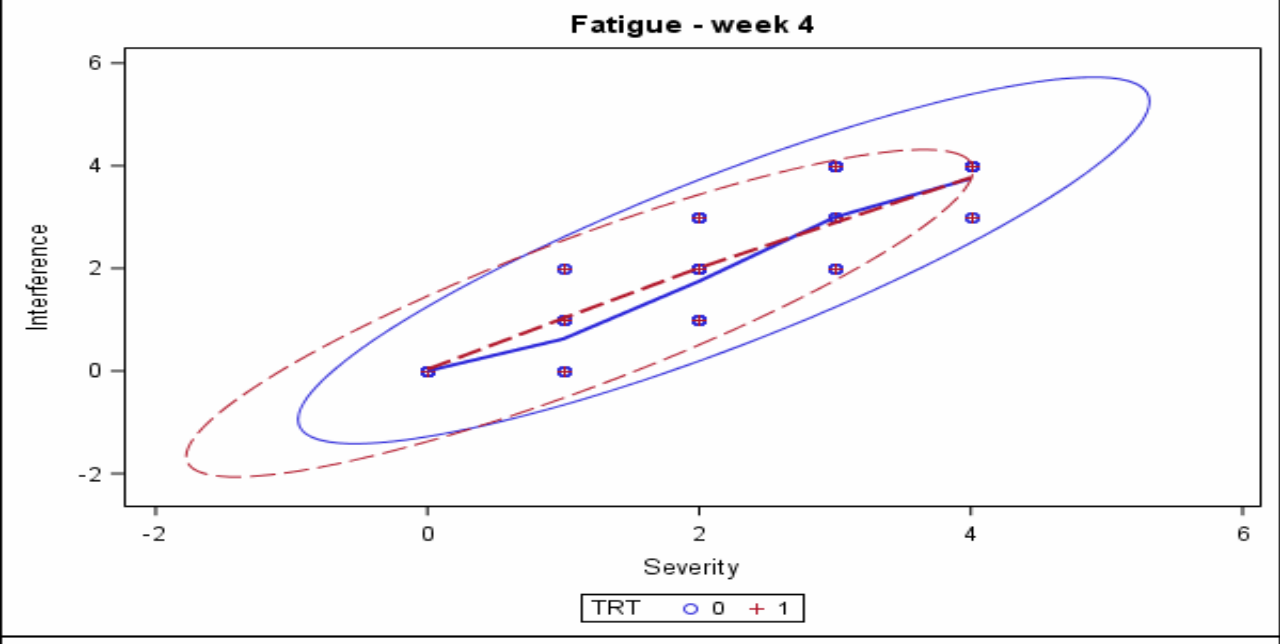
- **'non-compliance' considered outcome category.**
- **Outcome % conditional on expected nbr. Attrition % (death) moved to negative side.**

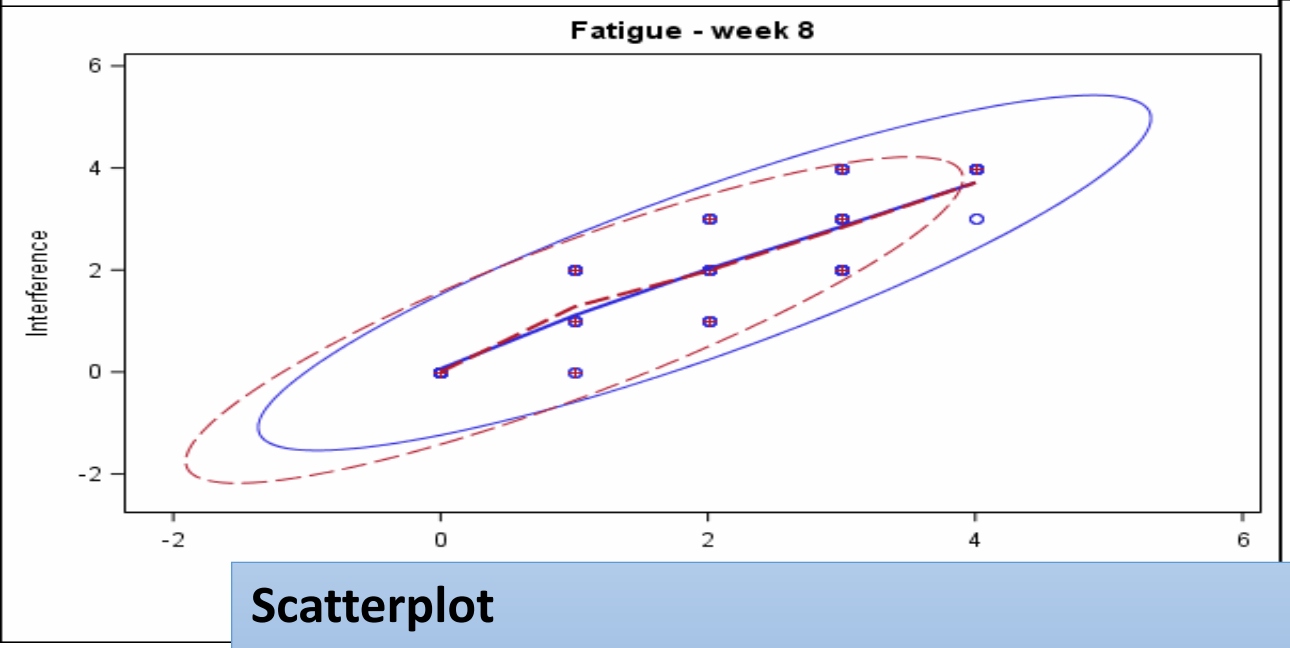
Extra graphs on fatigue

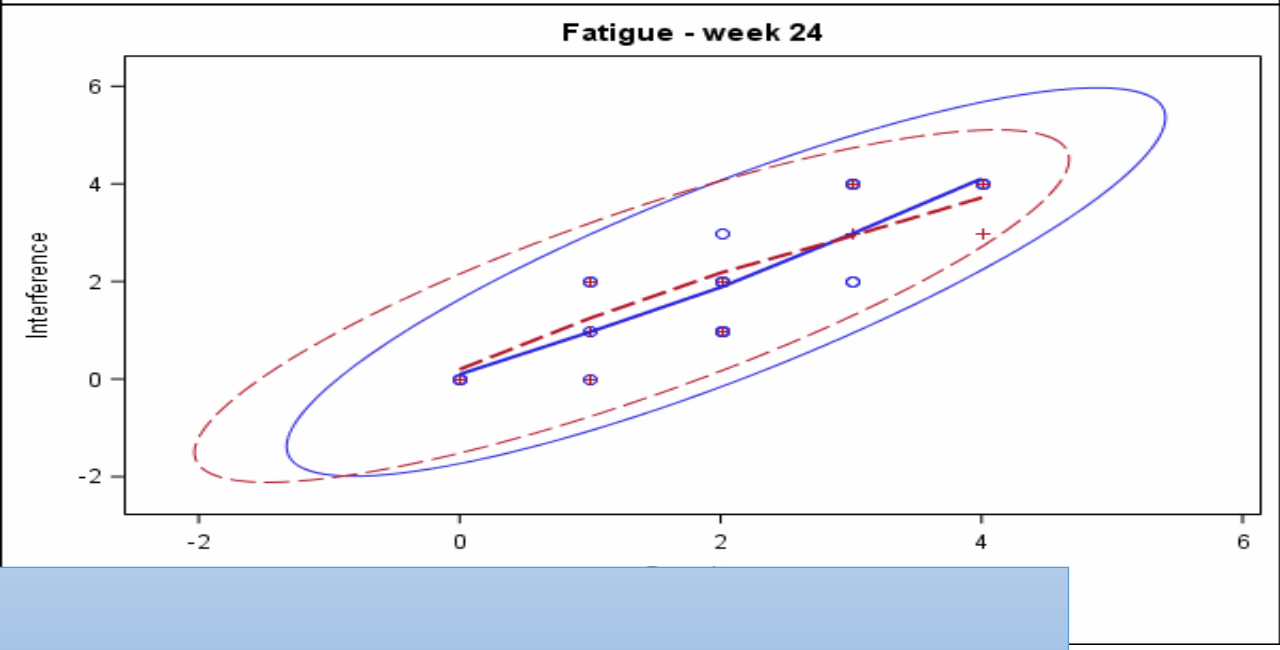
FDA U.S. FOOD & DRUG ADMINISTRATION

CRITICAL PATH INSTITUTE
collaborate · innovate · accelerate



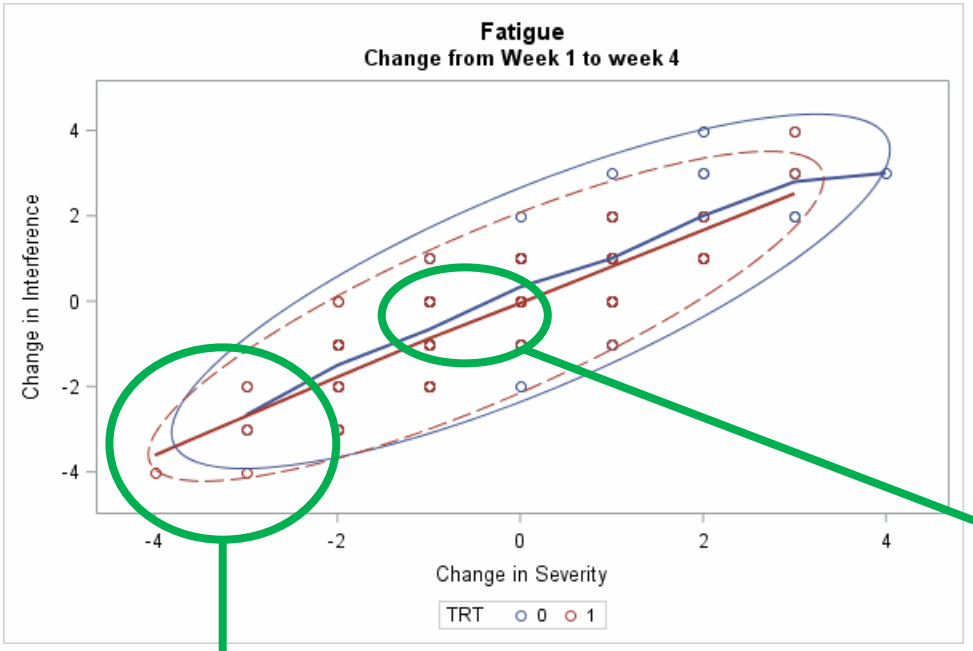






Scatterplot

- **+ Loess curve**
- **+ 95% confidence ellipse**



Narrow ellipses denotes strong correlation between severity and interference.

For the same severity, Treatment 1 gives less interference.

Confidence area closer to 0 = less overall problems

Confidence ellipse can be adjusted to account for missing data.

Issues: - more suited for continuous scales
- confidence ellipse assume normality and can exceed boundaries
- Loess curves susceptible to outliers

Further remarks

- Graphs to be tailored to study objectives
  - Eg. worst reported outcome during treatment.
  - Cumulative scores
- Ancillary data on missingness can be incorporated by subdividing missingness into relevant categories:
  - Eg: missing due to death, toxicity, lost to follow up, administrative
- Presenting mean scores requires assumptions – suitable for continuous outcomes, not so much for categorical.
- Study design properties:
  - Can it be assumed complete observation period is covered by COA?

A Descriptive Look at the PRO-CTCAE

Joseph C. Cappelleri, PhD, MPH, MS

Senior Director of Biostatistics

Pfizer Inc

Proportions (Counts): By Assessment and Across Assessments

- Scores for each item attribute (presence/absence/amount, frequency, severity, interference) can be presented descriptively with summary statistics
  - Item-level statistics by visit and across visits (typically, counts and proportions)
  - Portrayed in tabular or graphical format
  - CTCAE may or may not accompany PRO-CTCAE in same table or figure

- **One set of tables** can be based on proportions (counts) at the baseline assessment and at each post-baseline assessment of each visit
  - Grouped by treatment arm and, within each treatment arm, by item attribute

- **A second set of tables** can be portrayed similarly but based instead on cumulative incidence at any time across the post-baseline assessments
  - Proportion of individuals who had a particular event, regardless of grade (score > 0)
  - Proportion with severe event (score of 3 or 4)

Illustration of Descriptive Layout: Cumulative Incidence (Across Post-Baseline)

Adverse Event (AE)	Measures	Any Level AE (>0), %		Severe AE, %	
		Arm A	Arm B	Arm A	Arm B
Anxiety	CTCAE	31	37	--	--
	PRO-CTCAE: *Frequency*	59	72	8	12
	Severity	57	70	5	8
	Interference	40	51	6	9
Depression	CTCAE	18	12	1	--
	PRO-CTCAE: *Frequency*	30	48	4	7
	Severity	27	46	3	6
	Interference	23	40	2	9
Diarrhea	CTCAE	75	35	9	1
	PRO-CTCAE: *Frequency*	90	80	60	20

Baseline-Adjustment Method
(E Basch, LJ Rogak, AC Dueck. *Clinical Therapeutics* 2016; 38:821-830)

- **A third set of tables** can be based on the "baseline-adjustment method"
  - Sweeps across time (rather than by time)
  - Does NOT directly subtract baseline score
  - Incorporates only the maximum (worst) post-baseline score at the patient level
  - Then this worst level of an adverse event (AE) during treatment is tabulated only for AEs that are worse than the baseline score

- *Two versions*
  - Any AE: Maximum post-baseline score > baseline score
  - Severe AE: Maximum post-baseline score 3 or 4 (severe or very severe; frequently or almost constantly, quite a bit or very much) > baseline score

Illustration of Baseline Adjustment

Patient ID	W1 (Baseline)	W2	W3	W4	Maximum Score at Post-Baseline	Baseline Adjustment on Any Adverse Event? Yes or No (Score)	Baseline Adjustment on Severe Adverse Event? Yes or No (Score)
1	0	0	1	3	3	Yes (3)	Yes (3)
2	1	1	4	3	4	Yes (4)	Yes (4)
3	0	1	2	1	2	Yes (2)	No (2)
4	2	1	0	1	1	No (0)	No (0)
5	3	2	3	3	3	No (0)	No (0)

Means By Assessment and Rates Across Assessments

- If deemed useful, *means* (arithmetic average number of events) and *rates* (number of events per person-time) can be considered, along with their standard deviations

- **One set of tables** can be based on the *mean* number of each AE at the baseline assessment and at each post-baseline assessment, grouped by treatment arm
  - Within each treatment arm, mean numbers for the relevant attributes of interest (amount, frequency, severity, interference) can be ascertained

- **A second set of tables** can be based on the *rate* of each AE across post-baseline assessments, grouped by treatment arm and relevant attributes

Figures

- In addition to tables, figures can be considered to accompany and complement a given set of tables

- May be useful for interpretation purposes

National Cancer Institute

Analyzing Patient-Reported Outcomes to Draw Conclusions About Tolerability

Sandra A. Mitchell, PhD, CRNP

Research Scientist and Program Director

NCI Scientific Director, PRO-CTCAE

Outcomes Research Branch,

Division of Cancer Control and Population Sciences

National Cancer Institute

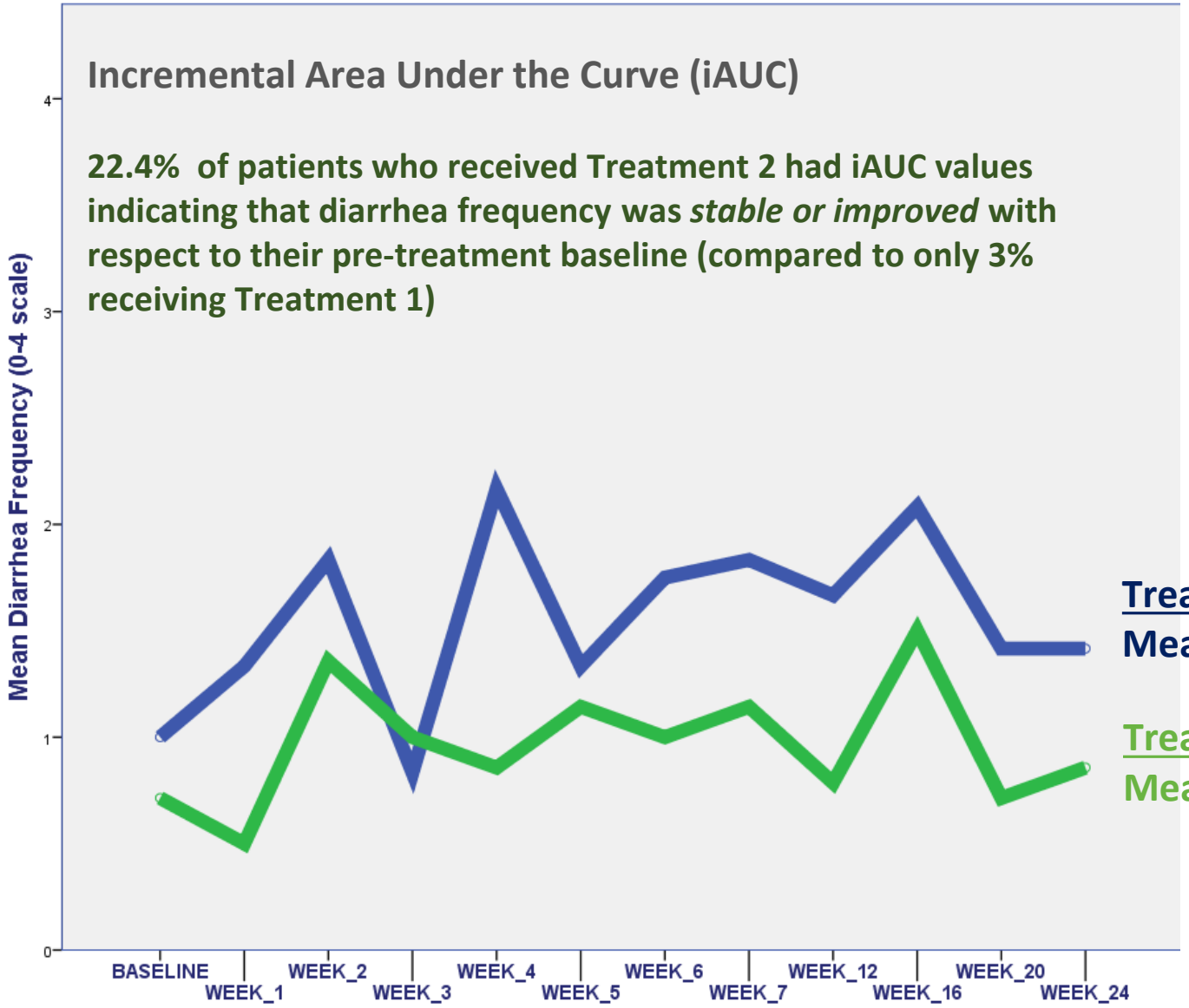
Rockville, MD

mitchlls@mail.nih.gov

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Area Under the Curve (AUC) to Summarize Individual Trajectories of Symptomatic AEs



Data Quality Metrics	Treatment 1 (n=100	Treatment 2 (n=100)
Missing at Baseline	17%	18%
Proportion within-patient missingness)† (12 timepoints of measurement)		
0 timepoints	14%	16%
1-2 timepoints	44%	39%
3-5 timepoints	26%	26%
6 or more timepoints	16%	19%

†No significant difference in proportion of missingness by treatment group χ^2=2.52 (n.s.)

Treatment 1
Mean AUC= 2.85 ±1.33*

*p<.001)

Treatment 2
Mean AUC=1.51± 0.98*

Modeling Symptomatic Adverse Events: Some Reflections

- Caveat: represents an oversimplification of a complex data set
  - Distinguish between missing and not expected to have been available (e.g. serious CTCAE event)
  - CTCAE data not provided to aid interpretation; PRO-CTCAE designed to be complementary to CTCAE
  - Reason for off-treatment (toxicity vs. progression vs. elective discontinuation) is important to know
  - Aims of trial and anticipated on- and off-target effects of regimen not known
- Metrics of data quality must reflect both within- and between-case missingness
  - Must distinguish missingness due to administrative issues, high-grade CTCAE event, off-treatment, withdrawal from study
- Examples of descriptive, graphical, and multivariate statistical approaches
  - Analysis summarizing the group/subgroup level (means across time) vs. analysis at the level of the individual (Growth Mixture Modeling, AUC)
- Data about symptomatic AEs will have distributional challenges (non-normal distribution and preponderance of zeros)
- Analytic approaches accommodate the fact that symptomatic AEs attributable to prior therapy may be present at baseline, and may *resolve, persist or worsen* due to regimen under study

Modeling Symptomatic Adverse Events: Some Reflections

- Cross-sectional analytic approaches (worst severity, with or without adjustment for baseline score) are easy to perform but do not fully leverage the available data and may be challenging to interpret
- Growth mixture models are intriguing and powerful approach:
  - Computationally demanding to perform and may be challenging interpret, and require fairly large sample sizes (>200), which may not be typical of many trials
- Selection of analytic strategies should consider:
  - Underlying conceptual assumptions
  - Distributional properties of the data; strengths and limitations of the analytic technique
  - Inferences we wish to make, potential sources of bias, and interpretability
- Sensitivity analyses and multiple analytic approaches converging on similar conclusions increase confidence in the results
- Current methodologic work to develop approaches to combine the attributes and identify thresholds of meaningful change will enhance interpretability and support additional analytic methods

We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.

Otto Neurath

Panel Discussion and Q & A

- **Chair**
  - *Laura Lee Johnson, PhD* – Deputy Director, Division of Biometrics III, Office of Biostatistics (OB), Office of Translational Sciences (OTS) , CDER, FDA
- **Presenters**
  - *Yolanda Barbachano, PhD* – Senior Statistical Assessor, Licensing Division, Medicines and Healthcare Products Regulatory Agency
  - *Mallorie H. Fiero, PhD* – Mathematical Statistician, Division of Biometrics V, OB, OTS, CDER, FDA
  - *Diane L. Fairclough, DrPH* – Professor, Biostatistics, Colorado School of Public Health
- **Panelists**
  - *Corneel Coens, MSc* – Lead Statistician, QoL Department, EORTC
  - *Joseph C. Cappelleri, PhD, MPH, MS* – Senior Director of Biostatistics, Pfizer Inc
  - *Sandra A. Mitchell, PhD, CRNP* – Research Scientist and Program Director, Outcomes Research Branch, NCI, NIH
  - *Paul G. Kluetz, MD* – Acting Associate Director of Patient Outcomes, OCE, FDA

Discussion (and reminder disclaimer)

- What questions can we answer with different longitudinal data and analyses?
- Are the visualizations interpretable?
- What assumptions must be made for each analysis?
  - Which assumptions make an analysis or visualization unuseful
- What are strengths and limitations of each analysis?

- What data do we need?
- How do we approach longitudinal analysis of PRO tolerability data?
- Which analysis population(s) should we consider?
- How can misinterpretation of analyses be minimized?
- What is the most appropriate way to deal with missing data for descriptive presentation (if at all)? How much does it matter given how safety data is currently collected and analyzed?

Not determining if an analysis "works" because data is FAKE