Using Patient Input to Estimate Clinically Meaningful Within-Patient Change at the Scale Score Level

Tenth Annual
Patient-Reported Outcome Consortium Workshop

April 24 – 25, 2019 ■ Silver Spring, MD

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.
- These slides are the intellectual property of the individual presenters and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. All trademarks are the property of their respective owners.

Session Outline

- Introduction
- Using Patient Input to Estimate Thresholds: Examples
  - Making the Case with a Quantitative Approach to Defining Meaningful Change
  - Using Qualitative Methods to Explore Meaningful Change with Patients at the Scale Level
  - Leveraging Post-Study Interviews with Clinical Trial Subjects to Gain Insight into Meaningful Change on a Composite Score
- Panel Discussion and Q&A

Session Participants

- Moderator
  - *Lori McLeod, PhD* – Vice President, Patient-Centered Outcomes Assessment, RTI Health Solutions
- Presenters
  - *Jean Paty, PhD* – Vice President, VP and Head of Patient Centered Endpoints, IQVIA
  - *Kate Sully, PhD* – Senior Research Manager, Patient-Centered Outcomes, Adelphi Values
  - *Kelly McCarrier, PhD, MPH* – Director and Qualitative Research Lead, Pharmerit International
  - *Cheryl D. Coon, PhD* – Principal, Outcometrix
- Panelists
  - *Scott Komo, DrPH* – Team Leader, Division of Biometrics III, Office of Biostatistics, OTS, CDER, FDA
  - *Michelle Campbell, PhD* – Senior Clinical Analyst for Stakeholder Engagement and Clinical Outcomes, DNP, CDER, FDA

Focus is on Treatment Benefit

A treatment benefit is a favorable effect on a meaningful aspect of how a patient feels or functions in their life or on their survival

Source: Walton MK, et al. Clinical Outcome Assessments: Conceptual Foundation; Report of the ISPOR Clinical Outcomes Assessment; Emerging Good Practices for Outcomes Research Task Force. Value in Health. 2015;18:741-52.

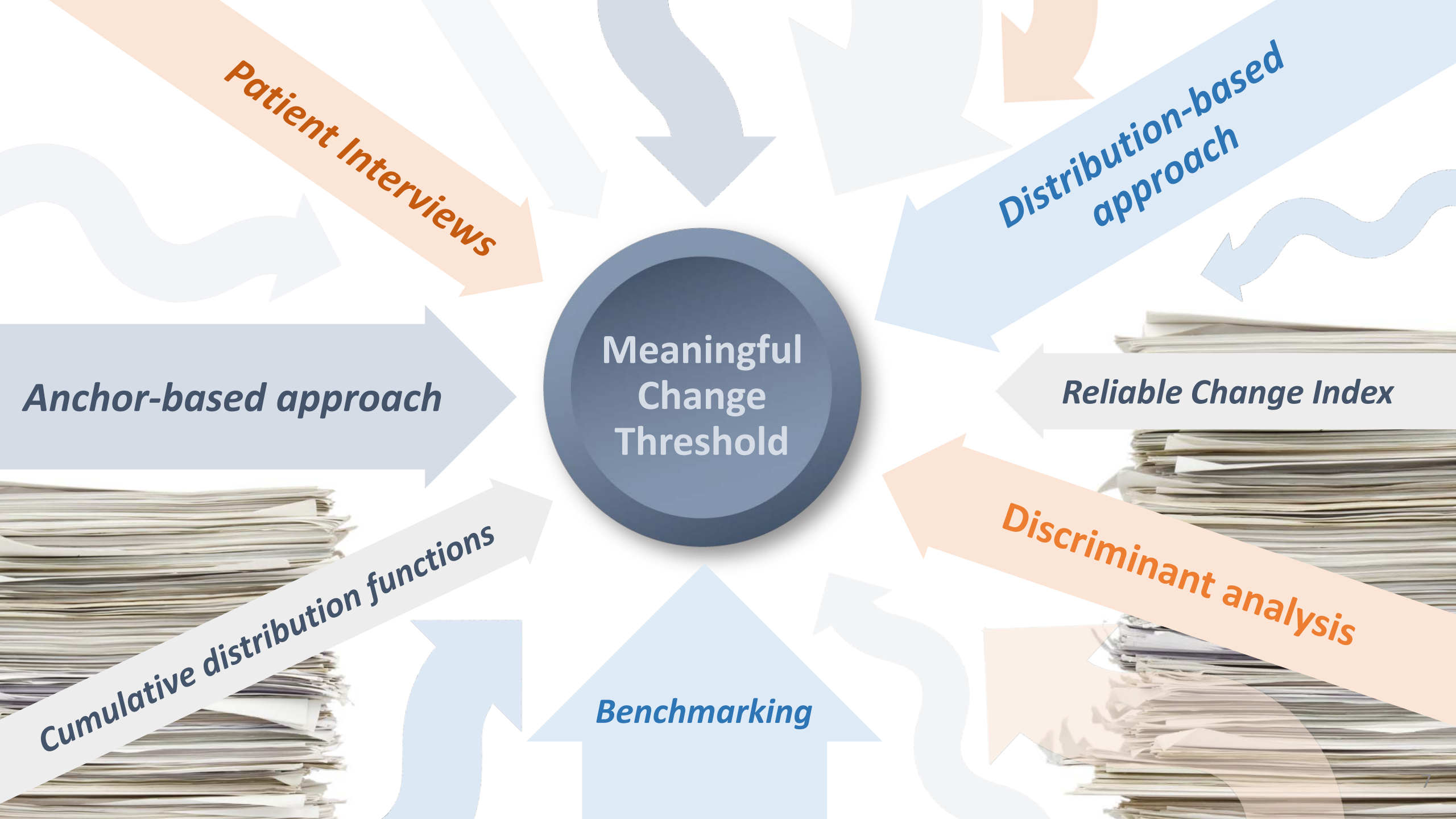
Is a 2-second reduction in a 100 meter freestyle time meaningful?

1:32 *Six-year-old swimmer…*

1:30 *unaware of 2 second difference*

0:49 *7th place 2005 World Aquatics Championship*

0:47 *Career best time in a 25-meter pool (Manchester 2009)*


Patient Interviews
Distribution-based approach
Anchor-based approach
Meaningful Change Threshold
Reliable Change Index
Cumulative distribution functions
Discriminant analysis
Benchmarking

Three examples that push us forward

Jean Paty:
Making the Case with a Quantitative Approach to Defining Meaningful Change

Kate Sully*:*
Using Qualitative Methods to Explore Meaningful Change with Patients at the Scale Level

Kelly McCarrier and Cheryl Coon:
Leveraging Post-Study Interviews with Clinical Trial Subjects to Gain Insight into Meaningful Change on a Composite Score

Making the Case with a Quantitative Approach to Defining Meaningful Change

Jean Paty, PhD, VP and Head of Patient Centered Endpoints, IQVIA

We pursued a fatigue label expansion for ruxolitinib

Jakafi Initial FDA Approval:
December 2011

Overall Goal

Our goal was to gain FDA approval for a fatigue-specific label expansion for ruxolitinib via PROMIS-Fatigue 7-item short form, showing:

- Instrument was fit-for-purpose for labeling
- Trial results were statistically significant and ***clinically meaningful*** (today's focus)

Regulatory Context

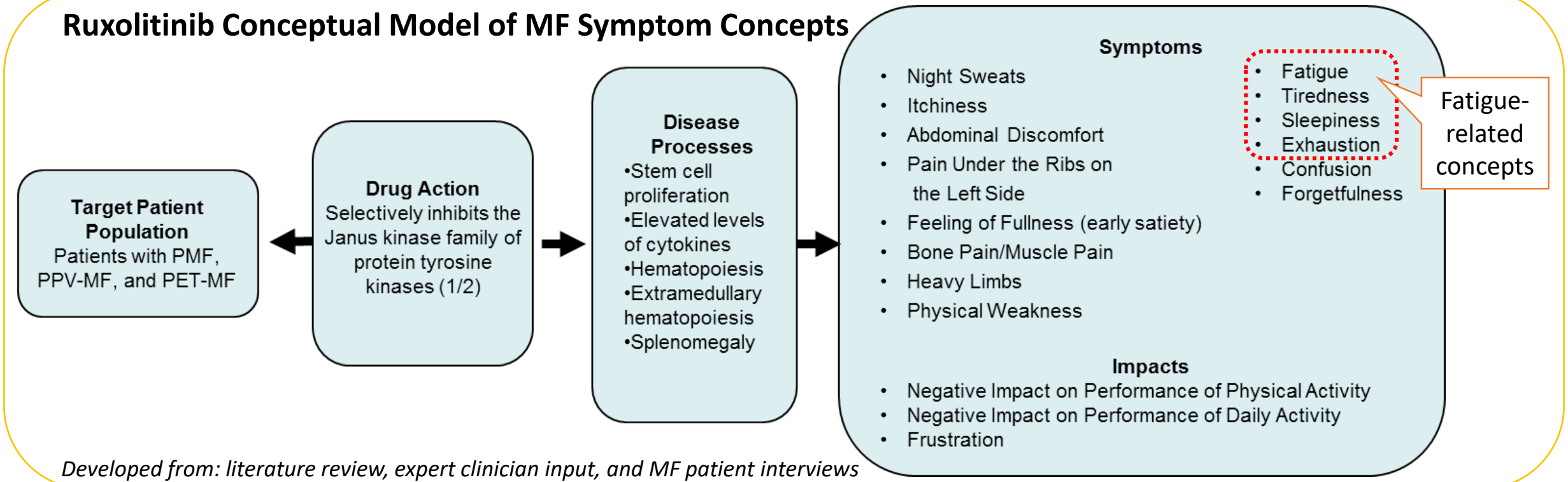
Ruxolitinib was already indicated for the treatment of patients with intermediate or high risk primary myelofibrosis (PMF), post-polycythemia vera-myelofibrosis (PPV-MF) or post-essential thrombocythemia-myelofibrosis (PET-MF).

Jakafi PRO Label Supplement:
December 2017

Fatigue is an important part of the myelofibrosis (MF) patient experience

PRO CONSORTIUM CRITICAL PATH INSTITUTE

Ruxolitinib Conceptual Model of MF Symptom Concepts



Developed from: literature review, expert clinician input, and MF patient interviews

*"The normal **fatigue** is just a feeling of **exhaustion**, um, not wishing to do my normal activities."*

*"When I feel **exhausted**, I feel **tired** and I feel like I must sit down and rest. If I am doing something, I will have to stop."*

*"My **energy** starts going down. I will start good but then as I keep going and progress, the energy starts weening from me."*

We sought to demonstrate Fatigue improvement with COMFORT-I trial data

PRO CONSORTIUM CRITICAL PATH INSTITUTE

COMFORT-I

- = objective clinical measure
- = patient-reported outcome

Patient Population	N = 309: PMF (50%), PPV-MF (31%), or PET-MF (18%)
Study Design	Randomized (1:1), double-blind, placebo-controlled trial comparing efficacy and safety of ruxolitinib to placebo
Primary Endpoint	Spleen volume (MRI/CT scan) response at Wk 24
Key Secondary Endpoints	Spleen volume (MRI/CT scan) duration of response MF symptoms (MFSAF v2.0 diary) response in TSS at Wk 24 MF symptoms (MFSAF v2.0 diary) change from BL in TSS Overall survival
Proposed Fatigue Endpoint	Fatigue (PROMIS-Fatigue Short Form 7a item symptom and impact scores) change from BL

- MFSAF v2.0 diary covers most key MF symptoms (from conceptual model), but not fatigue
- Fatigue was captured using the PROMIS-Short Form 7a
- New endpoint was proposed for fatigue in the label expansion submission package

PMF = primary myelofibrosis; **PPV-MF** = post-polycythemia vera-myelofibrosis; **PET-MF** = post-essential thrombocythemia-myelofibrosis; **MFSAF** = myelofibrosis symptom assessment form; **TSS** = total symptom score

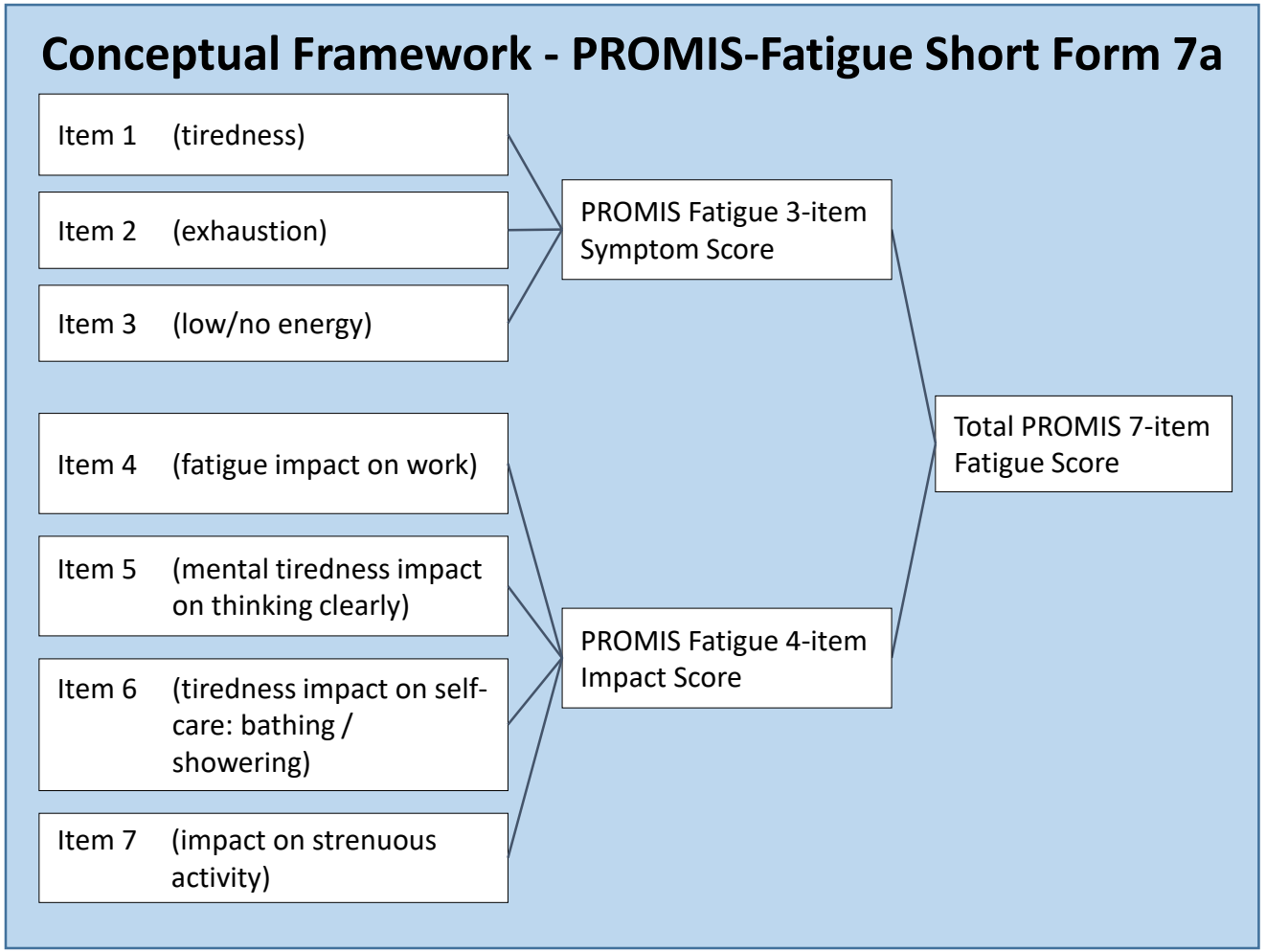
We used an FDA suggested structure for PROMIS-Fatigue scores

Scoring

- Traditional scoring is a total score of all 7 items (converted to T-score*)
- FDA suggested 2-scale approach; mean (vs. total) was used to better mirror patient response (1 = never to 5 = always)
  - **Symptoms**: Mean of items 1-3 *if* at least 2 items valid
  - **Impacts**: Mean of items 4-7 *if* at least 3 items valid

Minimally Important Difference (MID)

- Published definition of minimally important difference (MID) for the 7-item scale is raw score change of 2 to 3 points (*Yost, 2011*)
- MID definition is insufficient because:
  - Developed only for total 7-item score, not subscales
  - MID is not the same as clinically meaningful change



*T-score: Converted to mean = 50 and standard deviation of 10

Then rigorous methodology was used to define clinically meaningful change

Anchor-based quantitative approach

- "Responders" and non-responders" defined using patient-reported level of change
  - Irrespective of treatment assignment, intended to capture true *within* patient change
  - "Responders" chose PGIC response options '1' or '2' at week 24
- Calculated mean raw change scores for each scale
- Created 2x2 contingency tables
  - Treatment group by PGIC responder group
  - Tested with Fisher's Exact test
- Additional analyses were supportive and used to converge on responder definition

Patient Global Impression of Change (PGIC)

Since the start of the treatment you've received in this study, your myelofibrosis symptoms are:

1. Very much improved
2. Much improved
3. Minimally improved
4. No change
5. Minimally worse
6. Much worse
7. Very much worse

Results of this analysis were a crucial step and formed the basis for labeling submission

Results of meaningful change analyses:

- Resulting clinically meaningful change threshold values were:
  - **-2.35** for PROMIS-Fatigue 3-item Symptoms subscale
  - **-2.18** for PROMIS-Fatigue 4-item Impact subscale
  - **-4.53** for PROMIS-Fatigue Short Form 7a total scale
- PGIC group comparison ("responder" vs. "non-responder") contingency tables revealed clear statistically significant impact of Jakafi on all of the fatigue scales

Symptom Subscale

		Meaningful Change Group		
		Yes	No	Total
Jakafi	Count	55	87	142
	% within Ruxolitinib	38.70%	61.30%	100.00%
Placebo	Count	15	96	111
	% within Placebo	13.50%	86.50%	100.00%
Total	Count	70	183	253
	% across Trx Groups	27.70%	72.30%	100.00%

Impact Subscale

		Meaningful Change Group		
		Yes	No	Total
Jakafi	Count	55	87	142
	% within Ruxolitinib	38.70%	61.30%	100.00%
Placebo	Count	15	96	111
	% within Placebo	13.51%	86.49%	100.00%
Total	Count	70	183	253
	% across Trx Groups	27.67%	72.33%	100.00%

Total PROMIS-Fatigue SF 7a

		Meaningful Change Group		
		Yes	No	Total
Jakafi	Count	50	92	142
	% within Ruxolitinib	35.20%	64.80%	100.00%
Placebo	Count	16	95	111
	% within Placebo	14.40%	85.60%	100.00%
Total	Count	66	187	253
	% across Trx Groups	26.10%	73.80%	100.00%

Altogether, the submission made a case for clinically meaningful change in Fatigue

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Dossier evidence supported:

- Strong content validity for the PROMIS-Fatigue scales
- Understandable/meaningful measures to patients
- Relevance of content and scales to proposed labeling
- Valid/reliable scoring that is interpretable
- ***Defined change at conservative levels are clinically meaningful and can be used to support statements about treatment benefit***

Taken altogether, the dossier made the case that the instrument is fit-for-purpose to support labeling and that the results of the COMFORT-I trial reflect ***a statistically and clinically meaningful Fatigue improvement*** in patients treated with Jakafi, as compared to placebo.

FDA-approved label described a clinically meaningful effect of Jakafi on Fatigue

PRO CONSORTIUM CRITICAL PATH INSTITUTE

HIGHLIGHTS OF PRESCRIBING INFORMATION

These highlights do not include all the information needed to use JAKAFI safely and effectively. See full prescribing information for JAKAFI.

JAKAFI® (ruxolitinib) tablets, for oral use

Initial U.S. Approval: 2011

The label now highlights a ≥ 4.5-point reduction– our defined ***clinically meaningful change*** threshold– in 35% of patients on Jakafi vs. 14% on placebo

An exploratory analysis of patients receiving Jakafi also showed improvement in fatigue-related symptoms (i.e., tiredness, exhaustion, mental tiredness, and lack of energy) and associated impacts on daily activities (i.e., activity limitations related to work, self-care, and exercise) as measured by the PROMIS® Fatigue 7-item short form total score at Week 24. Patients who achieved a reduction of 4.5 points or more from baseline to Week 24 in the PROMIS® Fatigue total score were considered to have achieved a fatigue response. Fatigue response was reported in 35% of patients in the Jakafi group versus 14% of the patients in the placebo group.

Emphasis added

Source: Jakafi label https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/202192s015lbl.pdf

Successful labeling requires strong communication, methodology, and evidence

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Key success factors

Early and continuous communication throughout the process with the FDA culminated in an FDA submission via supplementary label extension package, after FDA drug approval.

Submission package provided necessary context and evidence to support proposed Fatigue labeling language to be added to the existing Jakafi label.

Storyline/submission included context of use justification and psychometrics, including a focus on derivation of clinically meaningful change definition with supporting quantitative analysis results.

Conclusion & future direction

Anchor-bases responder analysis for determining clinically meaningful change for Fatigue was acceptable to FDA, given strong supportive evidence and sound methodology

Until integration with novel qualitative methods is better defined and more widely used, there is still pathway to clinically meaningful change definitions via quantitative means only

Using Qualitative Methods to Explore Meaningful Change with Patients at the Scale Level

Kate Sully, PhD

Senior Research Manager, Patient-Centered Outcomes, Adelphi Values

Why is it important to explore meaningful change at the scale level with patients?

- Item-level estimates of clinically important change are sufficient for single item concepts e.g. a single item 0-10 pain score
- Although it is easier for patients to discuss meaningful change at the item level, interpretation is typically required at the scale level
- For multi-item scores, score-level estimates are required if they are to be of value for informing responder definitions to aid interpretation of the multi-item score
- Aggregating item-level estimates to provide scale level estimates is challenging and may result in inflated estimates, which represent unrealistic levels of change

Overview of the study and sample

Aim:

- To establish meaningful change thresholds for each scale of the EORTC QLQ-MY20 utilising anchor-based, distribution-based and qualitative methods*

Study design:

- The presentation will focus on the **qualitative methods only**
- Qualitative patient interview study, comprising the conduct of one-hour, semi-structured telephone interviews

Sample:

- 20 adult patients in the US and UK with a clinician confirmed diagnosis of relapsed/refractory (n=10) or newly diagnosed (n=10) multiple myeloma
- Good representation of gender and education level across the sample

*Findings from this qualitative study are being triangulated with anchor-based and distribution-based approaches to evaluate meaningful change thresholds for each scale of the EORTC QLQMY20

EORTC-QLQ MY20

- EORTC-QLQ MY20 four scales: Disease symptoms (6 items), Side effects of treatment (10 items), Body image (1 item) and Future perspectives (3 items)

EORTC MY20 Scale	Example items
Disease symptoms	Have you had bone aches or pain? Have you had pain in your back? Have you had pain in your hip?
Side effects from treatment	Did you feel drowsy? Did you feel thirsty? Did you have tingling hands or feet?

Scaffolding approach

Step 1

- Confirm patients' understanding of the underlying response scale

Step 2

- Introduce meaningful change at the item level

Step 3

- Build up to discussing meaningful change at the scale level

Step 1: Aim and approach

- Aim: To confirm patients' understanding of the underlying response scale.
- Approach: Patient completed the scale using a think aloud approach. Utilised cognitive debriefing questioning to confirm:

• What were you thinking of when you answered these questions?	• Of the locations you selected, which do you experience the worst pain in? Why?	• What does '1' or '4' mean to you on the response scale? How would that feel?
Concept(s) of interest assessed	Relevance of the concept(s)	Anchors and direction of response scale

Step 1: Results

Understanding

*"Well I'm thinking about the kind of—these kinds of **symptoms or, um, discomforts**, um, that I've had in the past week and really thinking back the last seven days of **how severe any of these pains are**, if I've had it."* (01-02)

Relevance

*"Um, **thirsty, uh, right now it's a four** because I'm on steroids. **Steroids make me—my mouth dry** and I feel like I've got to drink a lot."* (01-04)

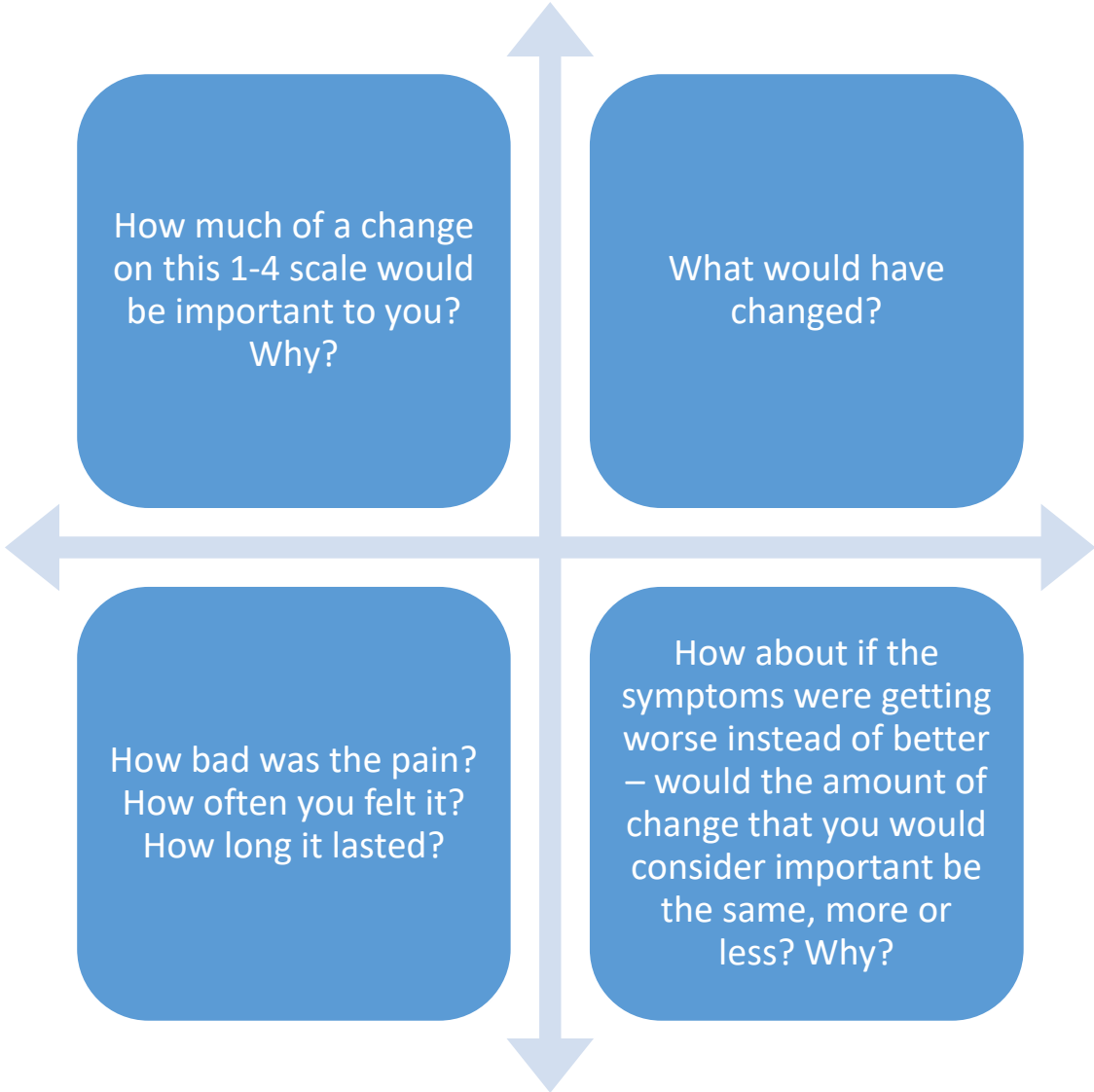
Response scale

*"Well a **one** to me would be not at all, **you don't feel any pain**."* (01-05)

*"I have been **drowsy all the time** when I was first diagnosed with myeloma because it was so bad **I couldn't stay awake**. Um, so I know what (a four) that feels like and it's awful."* (US-15)

Step 2: Aim and approach

- Aim: To help patients understand and articulate the concept of meaningful change at the item level before moving on to the scale level.
- Approach: Explored item level (1-4 scale) change for both improvement and worsening.
  - Discuss how a score change translates to changes in a patient's life
  - Discussed change in the context of a new treatment



Step 2: Results (improvement)

Score change	Impact of change on the patient	Worth taking a new treatment
1 point improvement	*"If I could **take my threes in pain down to a two**, which would be manageable, um, that would be—it's important to me to continue to actually **increase activity in my daily life without relying so much on other people or putting it off and finding other ways to do things**."* (01-18)	*"Yes, I would hope that a new treatment would like maybe **put you in remission** or something besides **getting rid of pain**."* (02-03)
2 point improvement	*"Hmm, I'd say **two points** would be important because it would mean **fewer medications and less aggressive therapy,** and certainly would—it would be **a visible demonstration of progress in removing symptoms.**"* (01-14)	*"**Yes because I definitely like the benefit of it, of the pain reduction**. I know a lot of medications, uh, in general really are about pain management, but especially for my myeloma."* (01-02)
3 point improvement	*"I could do—**I could be active again. I could walk**. Right now I can't even hardly walk to the barn. It's 200 feet and I can't hardly get there"* (01-03)	

Step 2: Results (worsening)

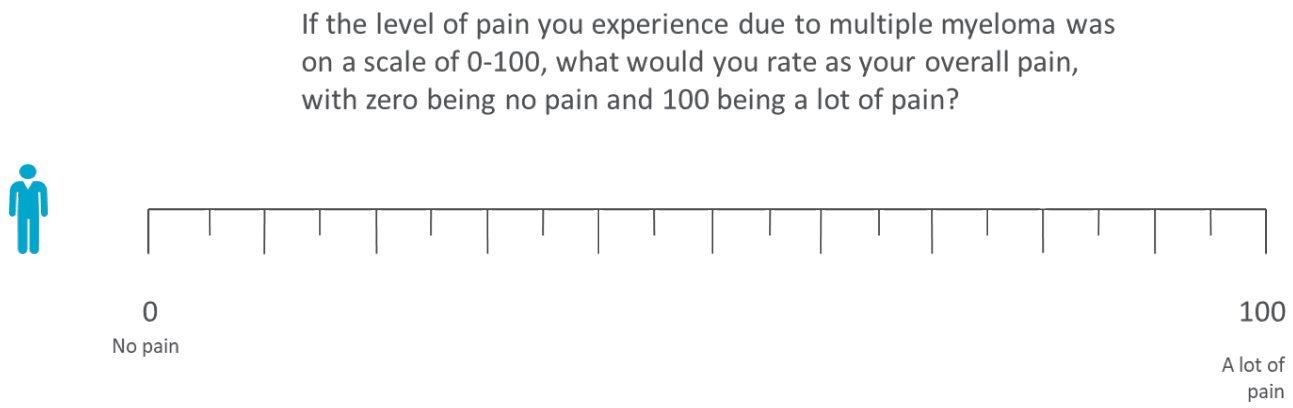
Score change	Impact of change on the patient	Worth taking a new treatment
1 point worsening	*"I wouldn't hold back on saying **a single point of, of reversion would be important** because rolling back to where I was before..that would be more important than demonstrating progress. Uh, because **losing ground to me would be symptomatic of, uh, of approaching mortality**."* (01-14)	*"I mean **if what I was taking—the side effects couldn't be controlled at three and we had to change the treatment plan, yes**. Then we change the treatment plan."* (01-04)
1 point worsening	*"Getting worse instead of incremental, uh, even a **one-point change** is noticeable. Um**, if I go, you know, from no dry mouth to a little bit of dry mouth, I know it**."* (01-01)	*"**Yes, it would probably be worth it**, because **these are things that are interrupting of the life**. Especially the feeling ill one which I think of as nausea. That is probably the very worst thing for me."* (01-07)
2 point worsening	*"I think a two-point change would be important, um, because that would mean **it's (side effects) becoming unmanageable**."* (01-19)	

Step 3: Aim and approach

- Aim: To confirm patients understand the move from the item level to the scale level and to explore meaningful change.
- Approach in interview: The scoring of the scale (0-100 scale) was explained and patients were asked to provide a scale score estimate, based on their responses to the items on the scale. Scale level change for improvement and worsening was explored.

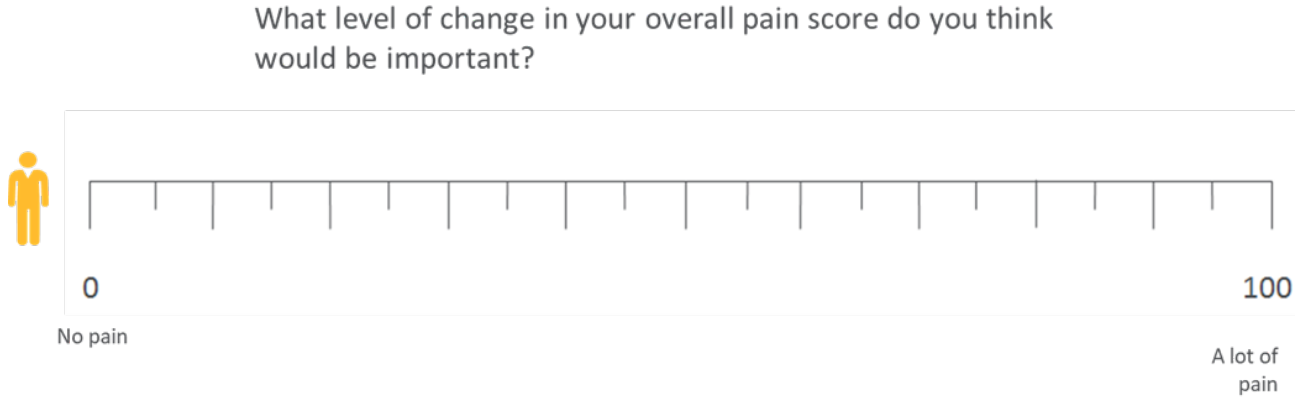
INTERVIEWER: What would you rate as your overall pain? So with zero being no pain and a hundred being a lot of pain.

SUBJECT: **60**



Step 3: Approach

INTERVIEWER: What were you thinking about when you answered that question and gave the answer 30?



SUBJECT: **Most people can live with pain of a level of three out of ten. And that I found that to be true for myself**. That it doesn't impact your quality of life until it starts to get over that. And so I kind of feel like right now I'm at like a 60 and that **if I could get back to 30 it's kind of back to functioning in a way where my quality of life feels higher**.

INTERVIEWER: You said your quality of life would be higher. What would have changed?

SUBJECT: I would be able to, **not have the spinal aches kind of full-time**, um, which then means that I don't worry about like **attending the museum** or, if I'm going to be **comfortable in a theater seat** if I—**I love to bake. If I can bake cakes and cookies**. If I can **cook a meal**. It **changes my energy level**. Um, and so when I can get it closer to a 30, **I can manage those things. I may have to plan a little bit, but I can get them all in**.

Step 3: Approach

INTERVIEWER: If we were to think about it the other way around and we were to think about a worsening, what level of change in your overall pain score do you think would be important?

SUBJECT: Oh, the one that—it would probably start to do me in at about 80 (20 point change).

INTERVIEWER: And again, um, if you could describe that to me. What would be happening?

SUBJECT: I probably **wouldn't be able to go out so much**. I probably **wouldn't ever find myself comfortable**. I **may have to consider pain meds during the day**. Um, I definitely **wouldn't be, you know, baking and cooking**. Um, I'd definitely be **more tired**. I hopefully wouldn't need **a walker** at that point, but I don't know.

Step 3: Results

EORTC MY20 Scale	Meaningful improvement	Example quote	Meaningful worsening	Example quote
Disease symptoms	20	*"Well then I could tolerate the pain instead at being at 70 or 90. I can* ***tolerate the pain and do all what I want to do****."* (US-09)	10	*"Because it's hard to—I mean it's* ***hard to handle a normal life*** *when it gets, when it gets to that kind of pain, you know."* (US-17)
Side effects from treatment	No clear pattern (evenly distributed 10/20/30)	*"If I could have it reduced to 60 or less, or less would be great. It would indicate* ***a significant decrease in the pain*** *I'm experiencing in my hands and feet, uh, my back, uh, and at times other parts of my body."* (US-11)	5-10	*"A change of 10 would be significant..because any side effects would cause me some concern, in that they had developed and that* ***could be because my myeloma would be deteriorating****."* (UK-02)

What worked well?

- Meaningful change estimates (ranges) were generated for all scales of the EORTC-QLQ MY20 for both improvement and worsening, based on the patient interviews
- Qualitative insights allowed us to explore why a change score was important and how it would translate into real world improvement or worsening for the patient
- The study highlighted which scales (disease symptoms and side effects of treatment) were most important to patients, as reflected in direct qualitative feedback from patients
- It is acknowledged that while it was still challenging for some patients to discuss meaningful change at the scale level the stepwise approach utilised led to more confidence in the estimates that were provided

Limitations, challenges and next steps

- Discussing meaningful change at the scale level with patients is cognitively complex
  - Should patients be provided with their calculated scale level or total score rather than estimating it themselves?
    - In the current study, patients' scale estimates were higher than the calculated scale scores.
    - However, any differences between patient estimates vs calculated scores typically fell below the meaningful change threshold.
- Variability in scale level estimates provided by patients
  - High degree of variability in qualitative patient estimates provided for the side effects from treatment and future perspectives scales
  - Makes it challenging to narrow to a single value or narrow range

Leveraging Post-Study Interviews with Clinical Trial Subjects to Gain Insight into Meaningful Change on a Composite Score

Kelly McCarrier, PhD, MPH, Director and Qualitative Lead, Pharmerit International

Cheryl D. Coon, PhD, Principal, Outcometrix

Study Context

- A PRO-based GI symptom index was developed for use in a clinical trial program evaluating a novel therapy for lactose intolerance (LI).
  - Includes four key symptoms of LI (i.e., pain, cramping, bloating, gas movement) assessed on an 11-point numerical response scale (NRS).
- Prior research established the symptom composite score derived from these items to be valid, reliable, and responsive.
- Anchor-based analyses were conducted on Phase 2 clinical trial data to identify meaningful change thresholds.
  - Additional evidence was requested during end-of-Phase 2 communications with the FDA.
- A qualitative interview study was planned to support the development and evaluation of additional PRO instruments for use in upcoming trials. This study provided an opportunity to further explore patients' perspective on the meaning of symptom change.

Qualitative Study Design/Methods

- Semi-structured interviews, approx. 90 min, and including both open-ended (concept elicitation) exercises and cognitive evaluation of PRO instrument content.
- Interviews conducted among adult patients with LI (n=23) who had previously participated in the P2 clinical trial.
  - Familiar with the symptom PRO instrument
  - Could reflect on actual symptom change experienced with study treatment
- To explore the experience of symptom change from the patient perspective and assess the magnitude of change deemed meaningful for each item, three key steps were utilized within the interview guide and process:

Qualitative Study Design/Methods

1. During cognitive interview evaluation of each PRO item, participants were probed to describe their interpretation of the response scale and provide examples of their experience with varying locations on the response scale (i.e., different levels of symptom severity)

You answered ____ for this item today. Can you describe what [symptom] at that level feels like?

- **What would be different about your experience if you had answered ___?**
- **How would your daily life change if you had moved to this level?**
- **What about ___?**
- **How would your daily life change if you had moved to this level?**

Qualitative Study Design/Methods

2. Participants were then probed to describe their location on the item's 0-10 response scale corresponding to the most severe (worst) experience with the symptom.

How would you have answered this question when this symptom was at its worst?

3. Participants were then asked to indicate the point they would need to reach to consider the improvement to be noticeable and meaningful to them.

From that point on the scale, how much improvement would you need to feel for it to be a noticeable and meaningful change in this symptom?

Qualitative Study Design/Methods

- Finally, for additional qualitative context, patients were provided with the PRO item scores they had reported at key timepoints during the study (from baseline to post-treatment), and were asked to characterize what that change meant to them:

You started the study at ____ and went to ____ 1 month after completing treatment.

Could you describe for me any ways in which your day-to-day life changed based on experiencing this level of change in this symptom?

Example Patient Quotes: Meaning of Change Experienced

- **2001: (Pain; 3 to 0)** *I was able to drink and eat a lot of more than what I was able to before.* (how change daily life?) *If it's bad, I'm going to be at home, and I know it. I'm not getting ready to go, because it's not going to allow me to go.*
- **2005: (Pain; 5 to 0)** *Well, that was a big relief for me. I was feeling wonderful. I was feeling less stressful. More relaxed. That allowed me to help more, concentrate on my job.*
- **1001: (Pain; 8 to 6)** *That's not exactly better.*
- **1007: (Pain; 5 to 3)** *By, with that change, I be able to drink a cup of milk, and have a bowl of cereal with milk, and I don't be able to used to do it before.*

Example Patient Quotes: Meaning of Change Experienced

- **2001: (Pain; 3 to 0)** *I was able to drink and eat a lot of more than what I was*

2001: (Pain; 3 to 0) *I was able to drink and eat a lot of more than what I was able to before. (how change daily life?) If it's bad, I'm going to be at home, and I know it. I'm not getting ready to go, because it's not going to allow me to go.*

- **1001: (Pain; 8 to 6)** *That's not exactly better.*
- **1007: (Pain; 5 to 3)** *By, with that change, I be able to drink a cup of milk, and have a bowl of cereal with milk, and I don't be able to used to do it before.*

Example Patient Quotes: Meaning of Change Experienced

2005: (Pain; 5 to 0) *Well, that was a big relief for me. I was feeling wonderful. I was feeling less stressful. More relaxed. That allowed me to help more, concentrate on my job.*

- **1001: (Pain; 8 to 6)** *That's not exactly better.*
- **1007: (Pain; 5 to 3)** *By, with that change, I be able to drink a cup of milk, and have a bowl of cereal with milk, and I don't be able to used to do it before.*

Example Patient Quotes: Meaning of Change Experienced

- **2001: (Pain; 3 to 0)** *I was able to drink and eat a lot of more than what I was able to before.* (how change daily life?) *If it's bad, I'm going to be at home, and I know it. I'm not getting ready to go, because it's not going to allow me to go*

1001: (Pain; 8 to 6) *That's not exactly better.*

- **1007: (Pain; 5 to 3)** *By, with that change, I be able to drink a cup of milk, and have a bowl of cereal with milk, and I don't be able to used to do it before.*

Example Patient Quotes: Meaning of Change Experienced

- **2001: (Pain; 3 to 0)** *I was able to drink and eat a lot of more than what I was able to before.* (how change daily life?) *If it's bad, I'm going to be at home, and I know it. I'm not getting ready to go, because it's not going to allow me*

1007: (Pain; 5 to 3) *By, with that change, I be able to drink a cup of milk, and have a bowl of cereal with milk, and I don't be able to used to do it before.*

have a bowl of cereal with milk, and I don't be able to used to do it before.

Example Patient Quotes: Change Needed From Worst to be Meaningful

- **2006:** (**worst pain?)** *I would say a 9. It was bad to me, for an abdominal pain. To the point where I curl, I couldn't stay up straight, because it just hurts.* **(how much change needed?)** *To a 2. I want to go that far down.*
- **1003: (worst pain?)** *Oh, about an 8. Oh yeah after like let's say I drank a milkshake.* **(how much change needed?)** *I would say [to] a 3, 4. To me that would be good. A 3, 4 …I would still be uncomfortable but at least I can have the milkshake without having the number 8 pain.*
- **1007: (worst bloating?)** *At an 8 yeah*. **(how much change needed?)** *Well when I feel like, what like a 4 it was a big difference.*
- **3004: (worst pain?)** 7 **(how much change needed?)** *In order for me to know that the medicine is taking effect and bringing it down, I'd say around 2! I should notice a difference if I'm coming down from a 7.*

Example Patient Quotes: Change Needed From Worst to be Meaningful

2006: (worst pain?) *I would say a 9. It was bad to me, for an abdominal pain. To the point where I curl, I couldn't stay up straight, because it just hurts.* **(how much change needed?)** *To a 2. I want to go that far down.*

- **1007: (worst bloating?)** *At an 8 yeah*. **(how much change needed?)** *Well when I feel like, what like a 4 it was a big difference.*
- **3004: (worst pain?)** 7 **(how much change needed?)** *In order for me to know that the medicine is taking effect and bringing it down, I'd say around 2!I should notice a difference if I'm coming down from a 7.*

Example Patient Quotes: Change Needed From Worst to be Meaningful

1003: (worst pain?) *Oh, about an 8. Oh yeah after like let's say I drank a milkshake.* **(how much change needed?)** *I would say [to] a 3, 4. To me that would be good. A 3, 4 …I would still be uncomfortable but at least I can have the milkshake without having the number 8 pain.*

- **1007: (worst bloating?)** *At an 8 yeah*. **(how much change needed?)** *Well when I feel like, what like a 4 it was a big difference.*
- **3004: (worst pain?)** 7 **(how much change needed?)** *In order for me to know that the medicine is taking effect and bringing it down, I'd say around 2!I should notice a difference if I'm coming down from a 7.*

Example Patient Quotes: Change Needed From Worst to be Meaningful

1007: (worst bloating?) *At an 8 yeah*. **(how much change needed?)** *Well when I feel like, what like a 4 it was a big difference.*

- **1007: (worst bloating?)** *At an 8 yeah*. **(how much change needed?)** *Well when I feel like, what like a 4 it was a big difference.*
- **3004: (worst pain?)** 7 **(how much change needed?)** *In order for me to know that the medicine is taking effect and bringing it down, I'd say around 2!I should notice a difference if I'm coming down from a 7.*

Example Patient Quotes: Change Needed From Worst to be Meaningful

3004: (worst pain?) 7 **(how much change needed?)** *In order for me to know that the medicine is taking effect and bringing it down, I'd say around 2! I should notice a difference if I'm coming down from a 7.*

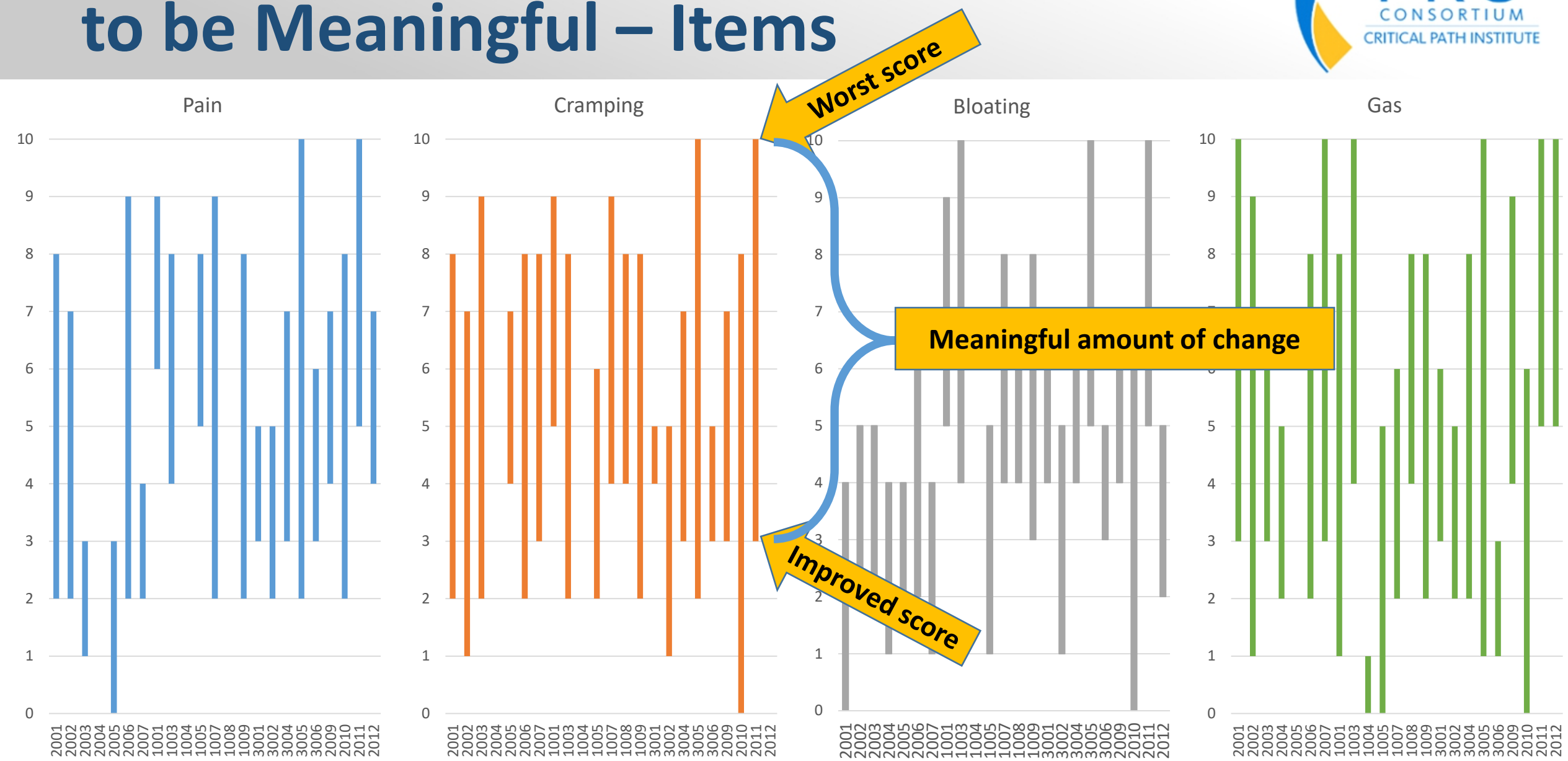
- **1007: (worst bloating?)** *At an 8 yeah*. **(how much change needed?)** *Well when I feel like, what like a 4 it was a big difference.*
- **3004: (worst pain?)** 7 **(how much change needed?)** *In order for me to know that the medicine is taking effect and bringing it down, I'd say around 2!I should notice a difference if I'm coming down from a 7.*

Quantifying the Qualitative Data

- Numbers corresponding to each subject's worst score experienced and the score that would be considered a meaningful improvement from that worst score were extracted and tabulated.

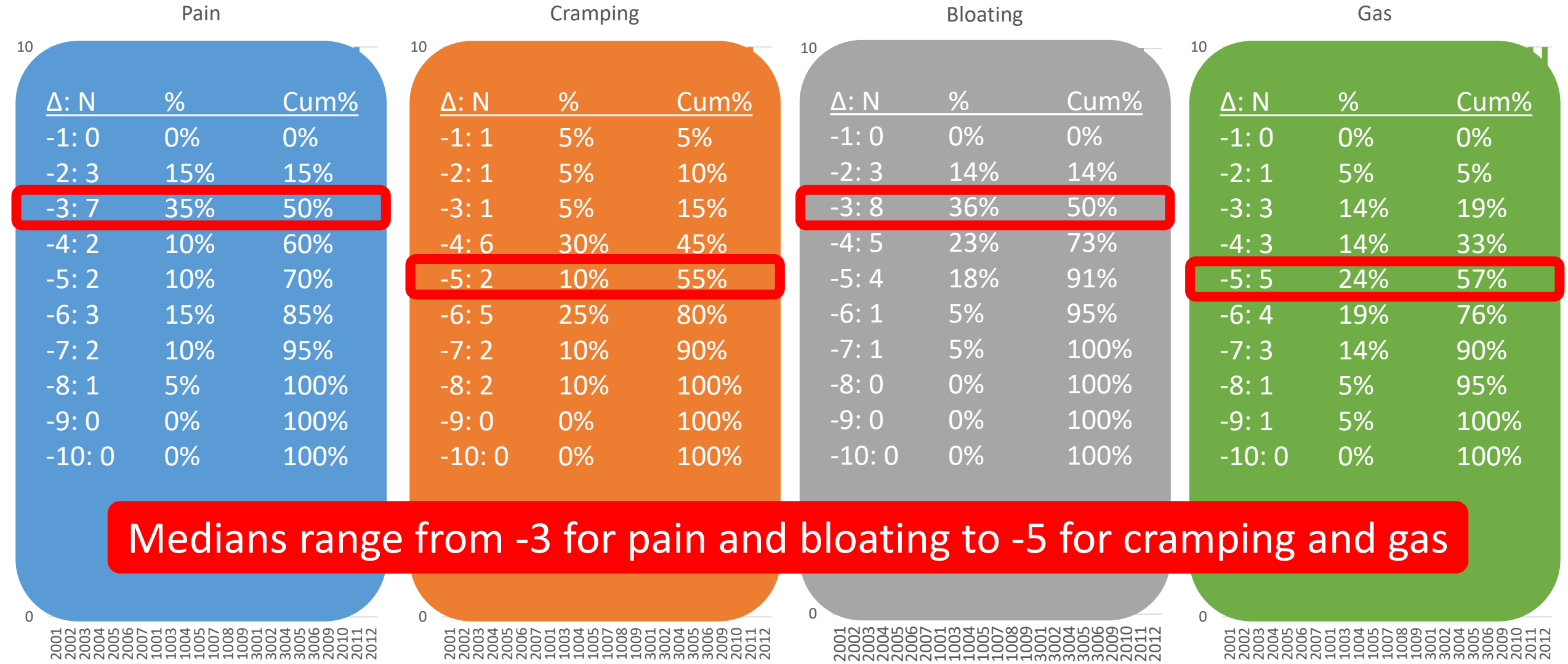
- Ranges of score changes were considered for common trends across subjects and across items.

Change Needed from Worst Score to be Meaningful – Items



Note. Top of bar is score for subject's worst experience. Bottom of bar is score needed for meaningful improvement from worst.

Quantifying Change Needed from Worst Score to be Meaningful – Items



Δ: N	%	Cum%
-1: 0	0%	0%
-2: 3	15%	15%
-3: 7	35%	50%
-4: 2	10%	60%
-5: 2	10%	70%
-6: 3	15%	85%
-7: 2	10%	95%
-8: 1	5%	100%
-9: 0	0%	100%
-10: 0	0%	100%

Δ: N	%	Cum%
-1: 1	5%	5%
-2: 1	5%	10%
-3: 1	5%	15%
-4: 6	30%	45%
-5: 2	10%	55%
-6: 5	25%	80%
-7: 2	10%	90%
-8: 2	10%	100%
-9: 0	0%	100%
-10: 0	0%	100%

Δ: N	%	Cum%
-1: 0	0%	0%
-2: 3	14%	14%
-3: 8	36%	50%
-4: 5	23%	73%
-5: 4	18%	91%
-6: 1	5%	95%
-7: 1	5%	100%
-8: 0	0%	100%
-9: 0	0%	100%
-10: 0	0%	100%

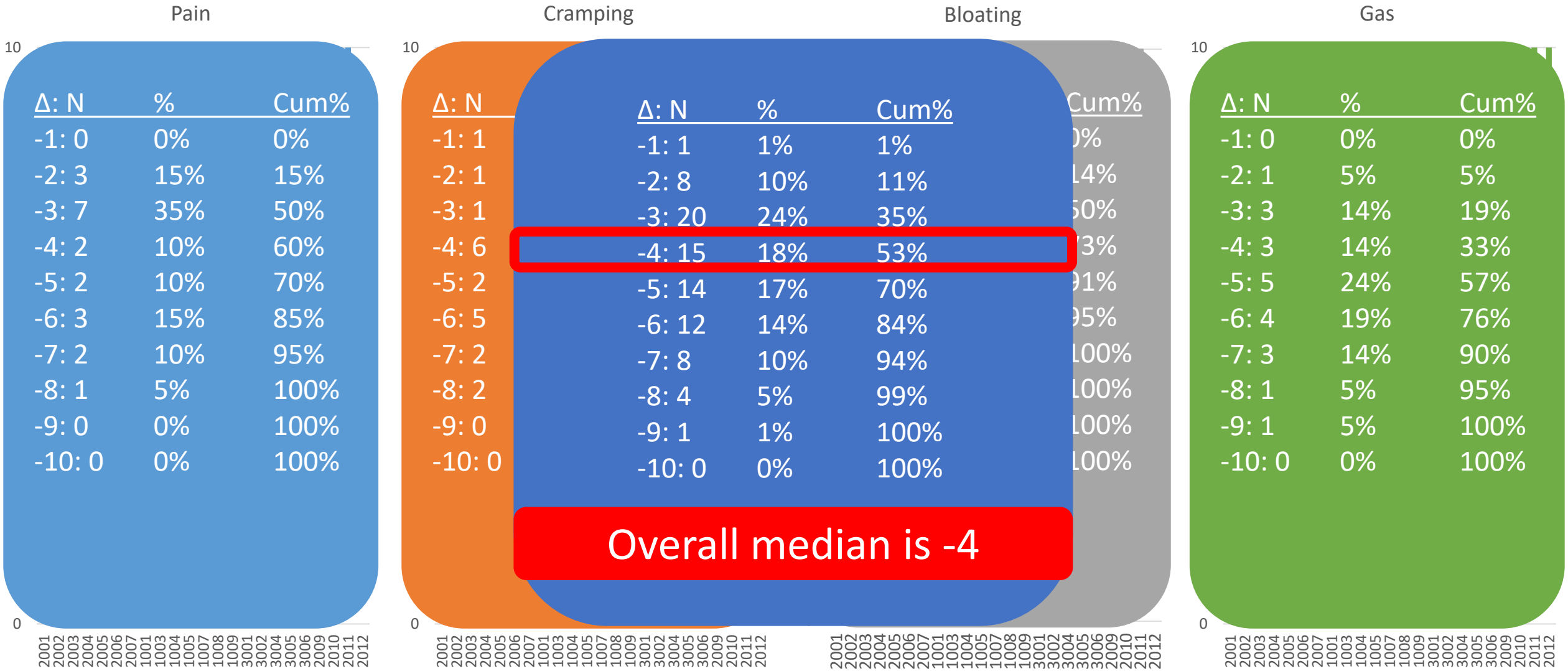
Δ: N	%	Cum%
-1: 0	0%	0%
-2: 1	5%	5%
-3: 3	14%	19%
-4: 3	14%	33%
-5: 5	24%	57%
-6: 4	19%	76%
-7: 3	14%	90%
-8: 1	5%	95%
-9: 1	5%	100%
-10: 0	0%	100%

2001 2002 2003 2004 2005 2006 2007 1001 1003 1004 1005 1007 1008 1009 3001 3002 3004 3005 3006 2009 2010 2011 2012

Medians range from -3 for pain and bloating to -5 for cramping and gas

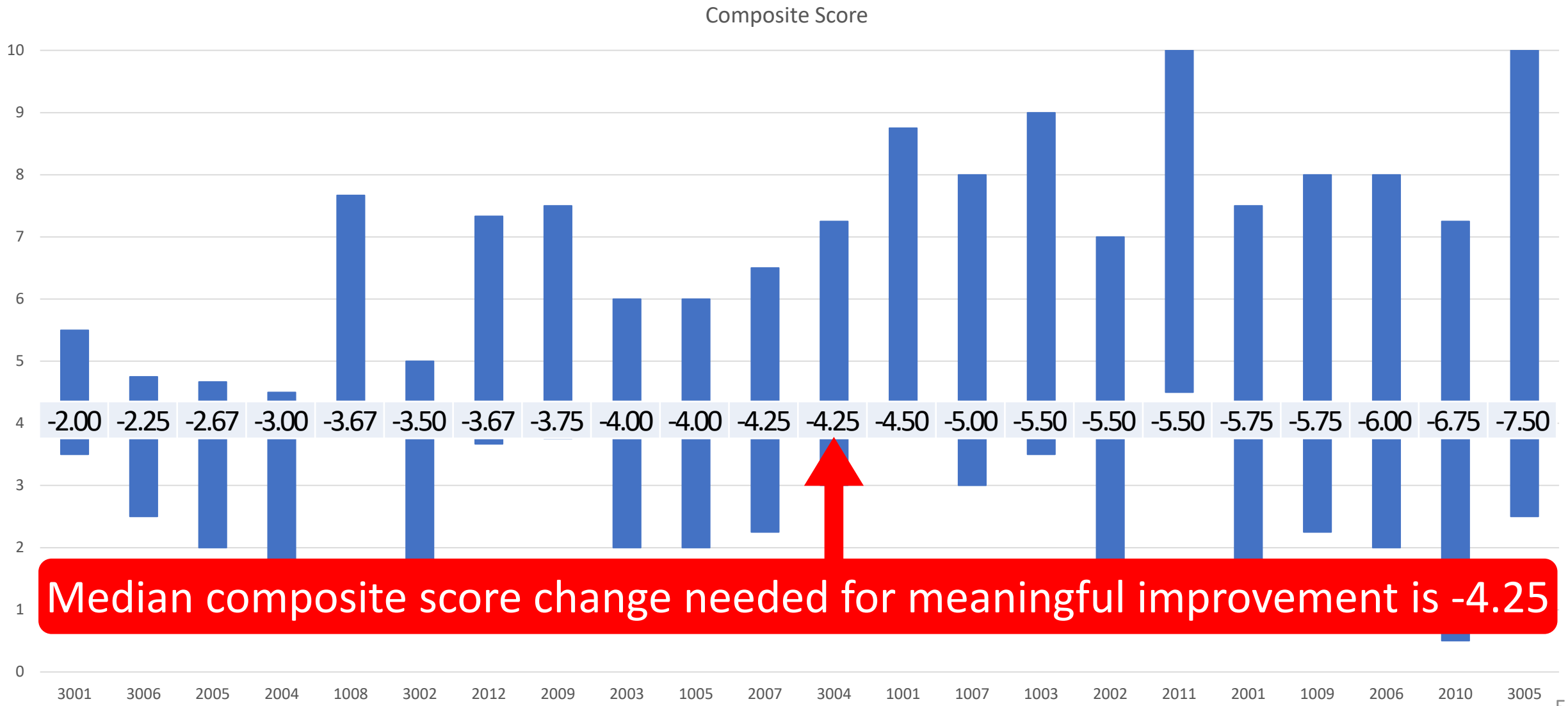
Note. Top of bar is score for subject's worst experience. Bottom of bar is score needed for meaningful improvement from worst.

Quantifying Change Needed from Worst Score to be Meaningful – Items



Note. Top of bar is score for subject's worst experience. Bottom of bar is score needed for meaningful improvement from worst.

Quantifying Change Needed from Worst Score to be Meaningful – Composite



Importance of Change Experienced During the Clinical Trial

- This process of going from qualitative item responses to the composite score assumes that change must be experienced on all symptoms simultaneously to be considered meaningful

- To evaluate this, we compared the composite score needed for meaningful change to the actual changes experienced on the composite score in the clinical trial for the subset of interview patients who reported that they had experienced a meaningful change overall when probed about their PGIS scores

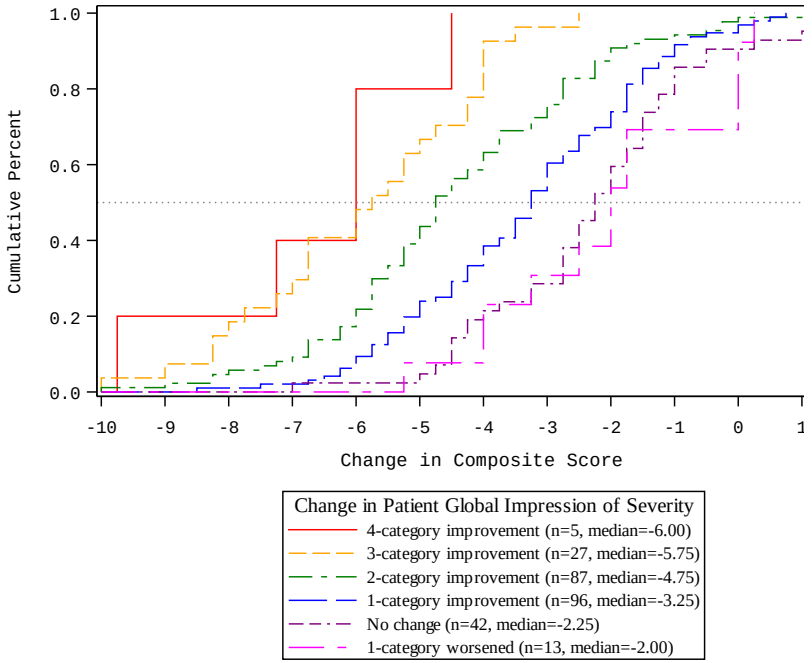
Importance of Change Experienced During the Clinical Trial

- The points should fall along the diagonal if meaningful change on overall symptom severity is equal to the sum of its parts
- However, 5 out of the 8 subjects who improved overall reported less change on the composite score than would've been expected by their individual item probing
  - Thus, the composite score meaningful change estimate of -4.25 may be considered on the high end, and lower amounts of change may also be perceived as meaningful when considering overall change

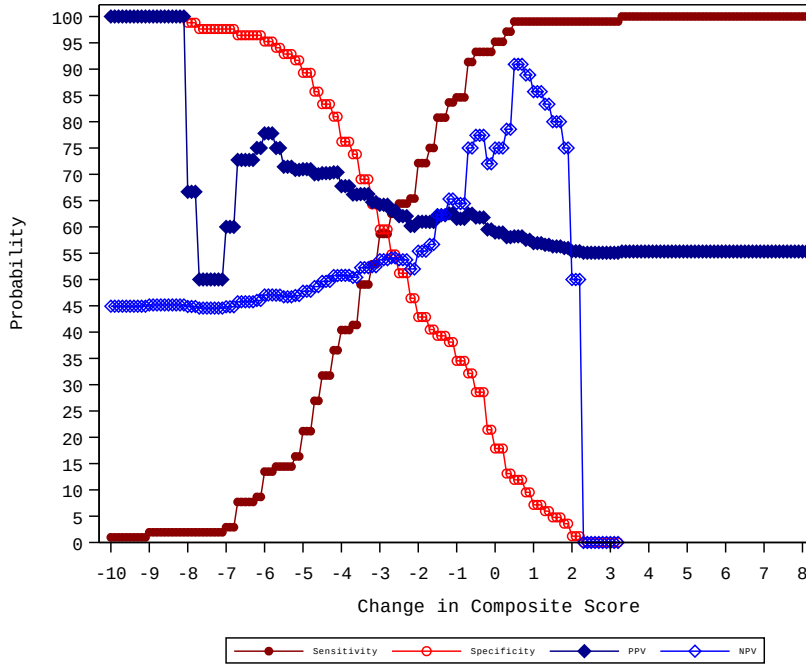
Psychometric Results

- Psychometric evidence was available from two clinical trials each with multiple anchors and multiple anchor-based methods

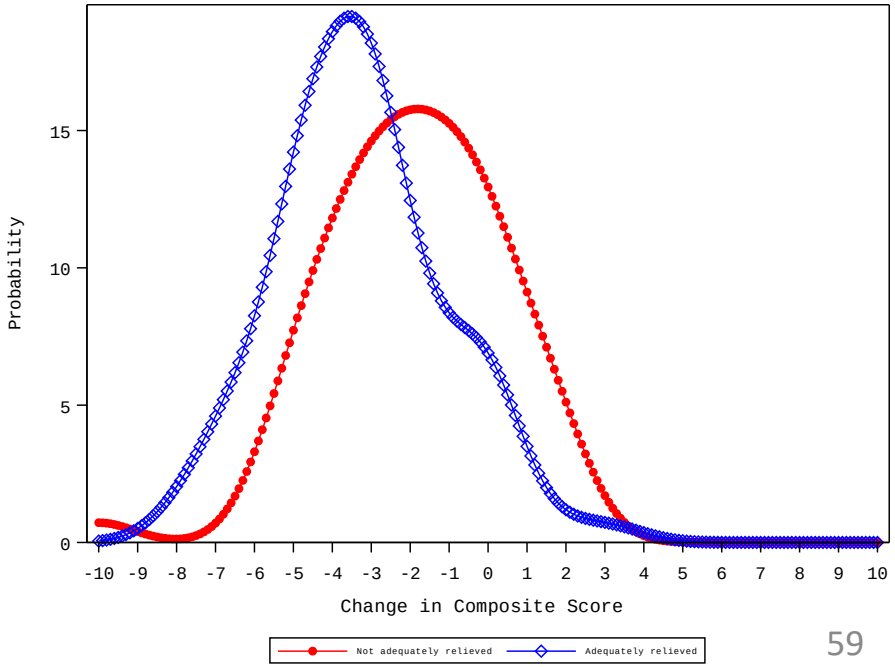
Phase 2b eCDF of composite score change by PGIS change



Phase 3 classification statistics for composite score change on PGIC moderate improvement or greater



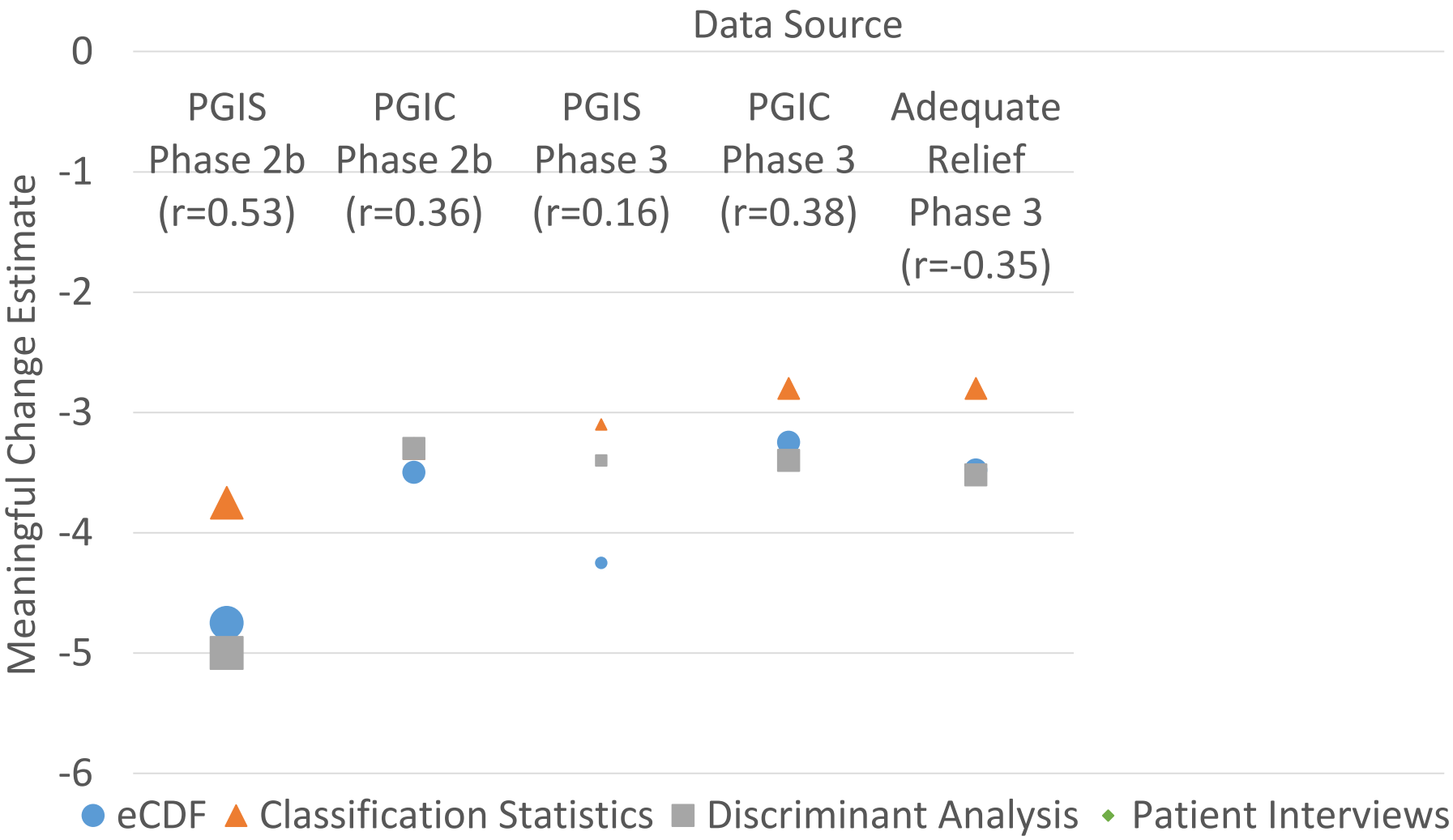
Phase 3 discriminant analysis of composite score change using assessment of adequate relief



Triangulation Process

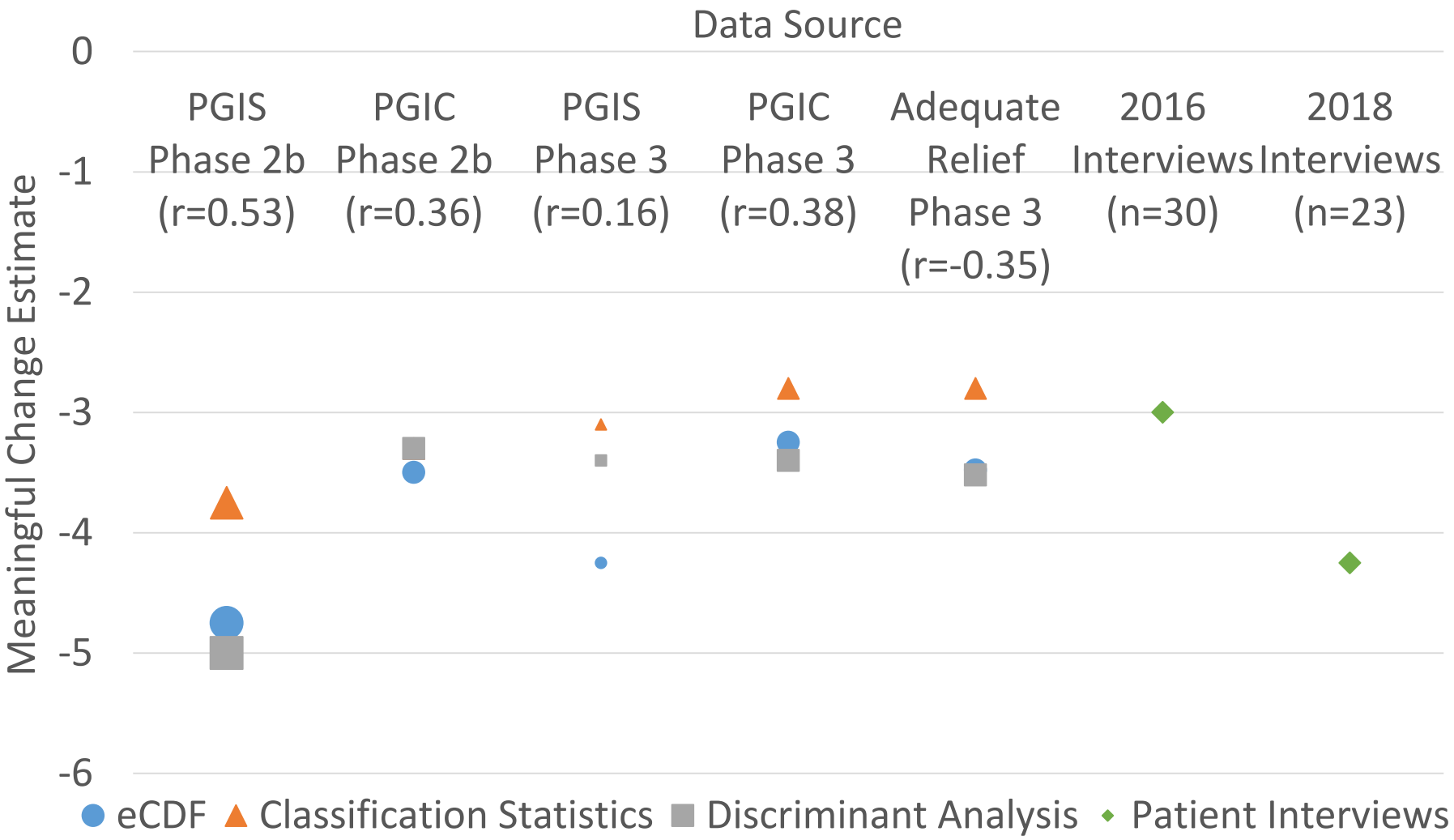
- The anchor with the strongest correlation points to a meaningful change estimate between −3.75 and −5

Triangulation Process



- The other anchors point to somewhat lower estimates between −2.8 and −4.25
- While there is a range of estimates across all anchors, they all appear to center around a meaningful change estimate of −4

Triangulation Process



- The interview results were consistent with the estimates from the other anchors, indicating that patients generally interpret change scores between −3 and −4.25 as meaningful

Conclusions

1. The psychometric data suggested that a 4-point change is generally considered a meaningful improvement on the composite symptom score, and qualitative data support this proposed location.

2. Hearing about patient-perceived changes in symptoms due to treatment in the patient's own words provides context for a 4-point change that could not be gained with psychometric data alone.

Insights for Future Research

1. Asking patients to reflect on actual treatment experience in the metric of a COA score scale is a cognitively complex process. Selection of appropriate question framing is important, as is the need to allow for enough time in the interview to carefully probe on this topic.
2. Qualitative data are inherently messy, as patients may offer seemingly inconsistent or variable responses. Look for patterns but don't expect consensus. If attempts are made to quantify qualitative data, don't lose sight of the context of those values within the patient quotes.
3. Post-study interviews also offer the opportunity to debrief the study anchor to be confident that it is being applied in anchor-based methods according to what is truly meaningful to patients.

Discussion and Q&A

- Moderator
  - *Lori McLeod, PhD* – Vice President, Patient-Centered Outcomes Assessment, RTI Health Solutions
- Presenters
  - *Jean Paty, PhD* – Vice President, VP and Head of Patient Centered Endpoints, IQVIA
  - *Kate Sully, PhD* – Senior Research Manager, Patient-Centered Outcomes, Adelphi Values
  - *Kelly McCarrier, PhD, MPH* – Director and Qualitative Research Lead, Pharmerit International
  - *Cheryl D. Coon, PhD* – Principal, Outcometrix
- Panelists
  - *Scott Komo, DrPH* – Team Leader, Division of Biometrics III, Office of Biostatistics, OTS, CDER, FDA
  - *Michelle Campbell, PhD* – Senior Clinical Analyst for Stakeholder Engagement and Clinical Outcomes, DNP, CDER, FDA