Advancing the Use and Interpretation of Meaningful Within-Person Change Thresholds

13th Annual
Patient-Reported Outcome Consortium Workshop

April 13-14, 2022

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 Objectives

- Describe the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials – Innovative Medicines Initiative (SISAQOL-IMI) consortium and the work they are doing on clinically meaningful change
- Provide a case study from a PRO Consortium member's use of the FDA-qualified *Non-Small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ)* in a phase 3 clinical trial and the evaluation of measurement properties and meaningful within-person change thresholds
- Discuss a mixed-methods approach to ascertaining a meaningful within-person change threshold of the Worst Itching Intensity Numerical Rating Scale (WI-NRS) to support a labeling claim for the treatment of chronic kidney disease-associated pruritis

Session Participants

Moderator

- *Rebecca M. Speck, PhD, MPH* – Clinical Outcome Assessment Scientist, Clinical Outcome Assessment Program, C-Path

Presenters

- *Carla Mamolo, PhD* – Director, Health Economics & Outcomes Research, Pfizer, Inc.
- *Johannes Giesinger, PhD* – Assistant Professor, Medical University of Innsbruck
- *Josephine Norquist, MS* – Executive Director, Patient-Centered Endpoints & Strategy Lead, Merck & Co., Inc.
- *Margaret Vernon, PhD* – Senior Vice President, General Manager, Evidera, Inc.

Additional Panelists

- *Selena Daniels, PharmD, PhD* – Clinical Outcome Assessment Team Leader, Division of Clinical Outcome Assessment, U.S. Food and Drug Administration
- *Lili Garrard*, *PhD* – Lead Mathematical Statistician, Division of Biometrics III, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration

Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials – Innovative Medicines Initiative (SISAQOL-IMI) Work Package 6: Harmonization of the terminology and definitions of clinically meaningful change in cancer clinical trials

Carla Mamolo, PhD – Director, Health Economics & Outcomes Research, Pfizer

Johannes Giesinger, PhD – Assistant Professor, Medical University of Innsbruck

Outline

- Overview of the SISAQOL-IMI project
  - Overall objectives
  - The work package structure and how Work Package 6 (WP 6) fits in
  - Work of the SISAQOL consortium preceding the current project
- Overview of SISAQOL-IMI WP 6 *(recommendations related to clinically meaningful change)*
  - Objectives
  - Main tasks
    - Scoping reviews of methodological literature and of trial literature (protocols and publications)
    - Current status and upcoming activities

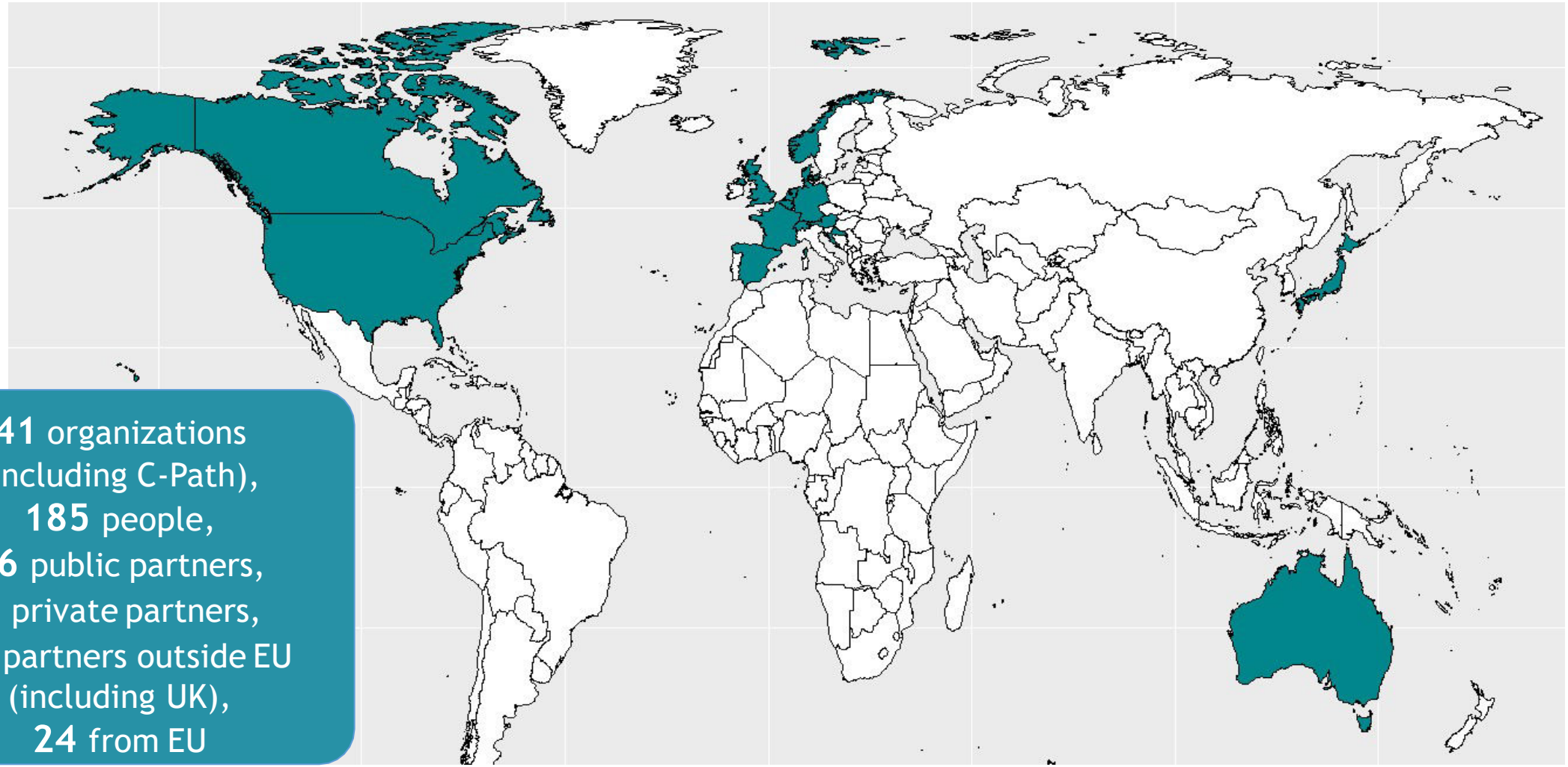
Overview of SISAQOL-IMI

This project has received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking (JU) under grant agreement No 945052. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations.

SISAQOL-IMI in numbers

Includes regulators (FDA, EMA, Health Canada, MHRA), HTAs (NOMA, IQWiG), patient advocacy groups, industry, academic groups, and SMEs

PRO CONSORTIUM
CRITICAL PATH INSTITUTE



41 organizations (including C-Path),
185 people,
36 public partners,
5 private partners,
17 partners outside EU (including UK),
24 from EU

EMA = European Medicines Agency; FDA = Food and Drug Administration; HTA = Health Technology Assessment; IQWiG = Institute for Quality and Efficiency in Health Care; MHRA = Medicines and Healthcare products Regulatory Agency (UK); NOMA = Norwegian Medicines Agency; SME = subject matter experts

SISAQOL was formed to address the lack of consensus on standards and how to interpret and report PRO endpoints

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium

Corneel Coens, Madeline Pe*, Amylou C Dueck, Jeff Sloan, Ethan Basch, Melanie Calvert, Alicyn Campbell, Charles Cleeland, Kim Cocks, Laurence Collette, Nancy Devlin, Lien Dorme, Hans-Henning Flechtner, Carolyn Gotay, Ingolf Griebsch, Mogens Groenvold, Madeleine King, Paul G Kluetz, Michael Koller, Daniel C Malone, Francesca Martinelli, Sandra A Mitchell, Jammbe Z Musoro, Daniel O'Connor, Kathy Oliver, Elisabeth Piault-Louis, Martine Piccart, Chantal Quinten, Jaap C Reijneveld, Christoph Schürmann, Ashley Wilder Smith, Katherine M Soltys, Martin J B Taphoorn, Galina Velikova, Andrew Bottomley; for the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium*

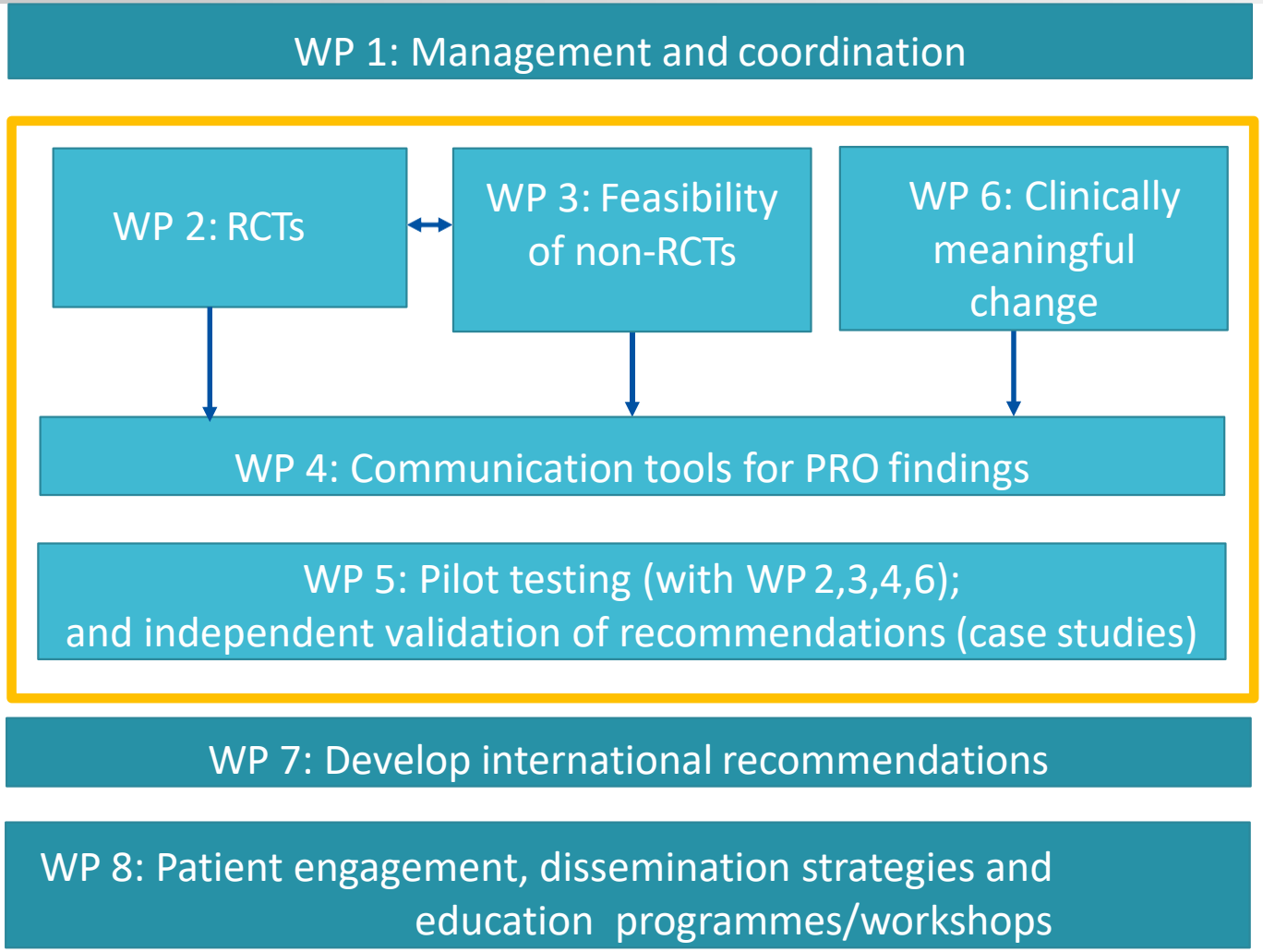
Patient-reported outcomes (PROs), such as symptoms, function, and other health-related quality-of-life aspects, are increasingly evaluated in cancer randomised controlled trials (RCTs) to provide information about treatment risks, benefits, and tolerability. However, expert opinion and critical review of the literature showed no consensus on optimal methods of PRO analysis in cancer RCTs, hindering interpretation of results. The Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium was formed to establish PRO analysis recommendations. Four issues were prioritised: developing a taxonomy of research objectives that can be matched with appropriate statistical methods, identifying appropriate statistical methods for PRO analysis, standardising statistical terminology related to missing data, and determining appropriate ways to manage missing data. This Policy Review presents recommendations for PRO analysis developed through critical literature reviews and a structured collaborative process with diverse international stakeholders, which provides a foundation for endorsement; ongoing developments of these recommendations are also discussed.

Coens et al (2020), *Lancet Oncology*, 21:e83-96.

SISAQOL-IMI: Overall project aims

- Achieve international consensus, across stakeholders, on the optimal use of PRO data in cancer clinical trials
- Improve the quality of statistical analysis of PRO data in cancer clinical trials
- Improve the standards of reporting of PRO data, improve reliable interpretation, and ultimately faster dissemination, of PRO findings
- Ultimately, better use of these data in regulatory approvals, HTA assessment, and shared decision-making

5 Scientific Work Packages, 3 Cross-Cutting Work Packages



Although the scope is focused on oncology, many statements will be applicable across disease areas

RCT = randomized controlled trial; WP = work packages

WP 6: Develop international recommendations for the terminology and definitions of clinically meaningful change in cancer clinical trials

WP 6 Objectives

1. **To harmonize** the various terminologies and definitions related to clinically meaningful change (CMC) in cancer clinical trials (clearly differentiating between **group-level and individual-level change**)

2. **To match** these terminologies and definitions **to the appropriate PRO objectives** in WP2 and in WP3

3. **To identify best practices** in the development of clinical meaningful change research objectives

WP 6 Task overview

Task 6.1: **Initial discussions** with stakeholders

Task 6.2: **Systematic literature review** of cancer studies establishing and evaluating thresholds for clinically meaningful change

Task 6.3: **Systematic literature review** of past and current practice of using clinically meaningful change and related concepts in cancer trials

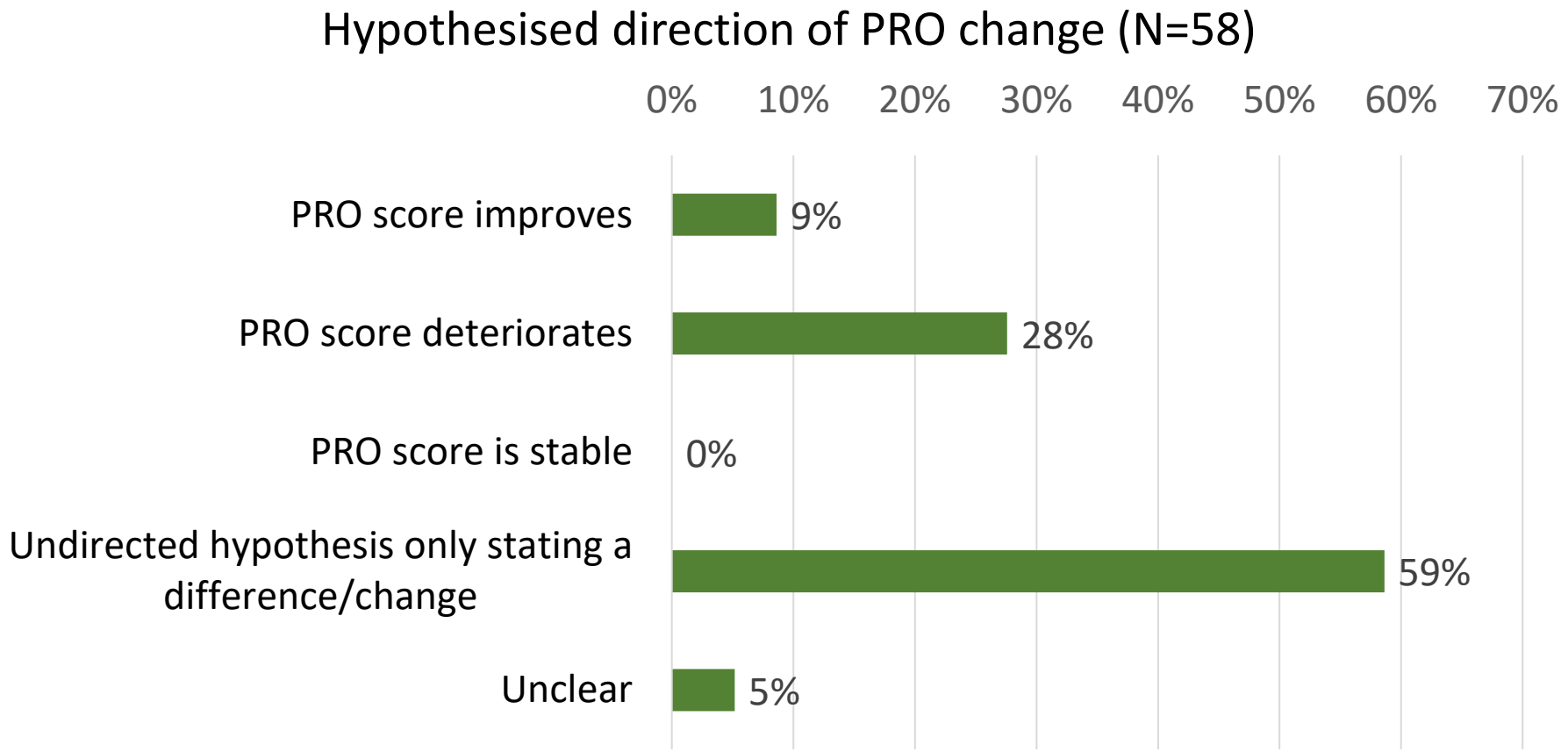
Task 6.4: **Drafting proposed statements** for clinically meaningful change recommendations based on aggregated data from the systematic reviews in task 6.2 and 6.3

We are here

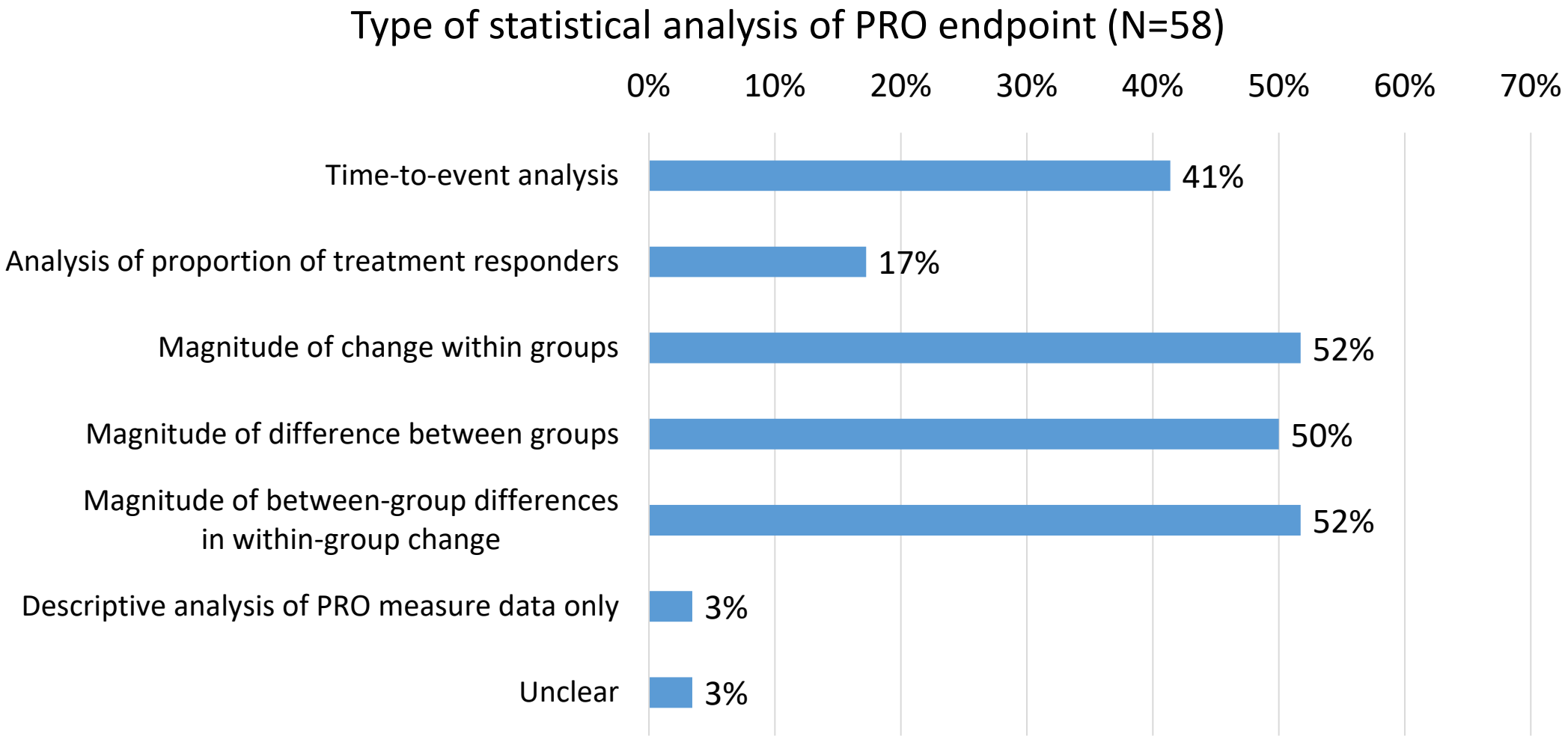
Task 6.5: **International consensus recommendations** for a harmonized terminology for clinically meaningful change and for their application in cancer clinical trials

WP 6: Selected results from the review of comparative trials

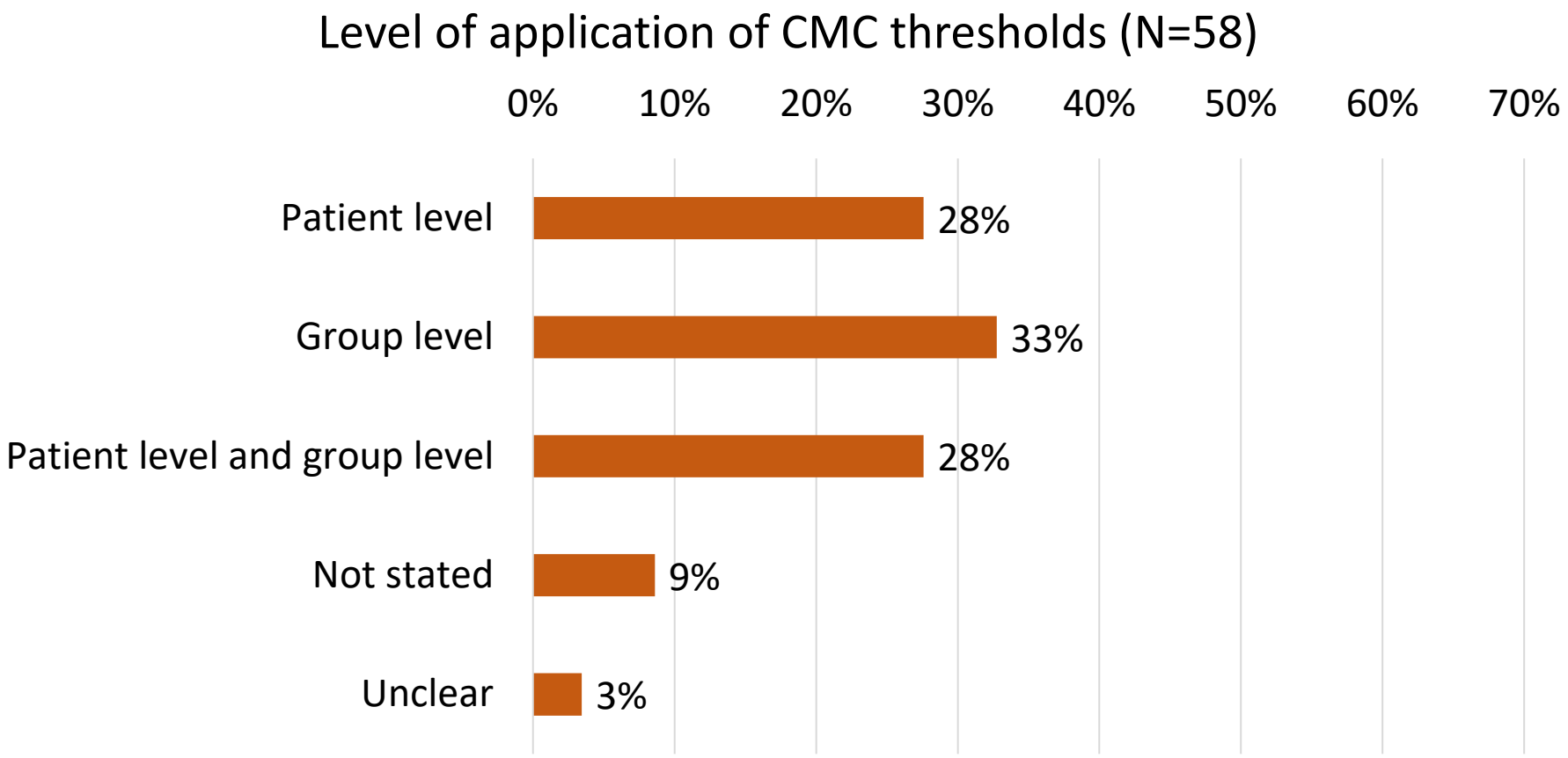
Review of comparative trial publications (N=36) and protocols (N=22) identified by Pe et al. (2018, Lancet Oncol) or provided by SISAQOL-IMI consortium members



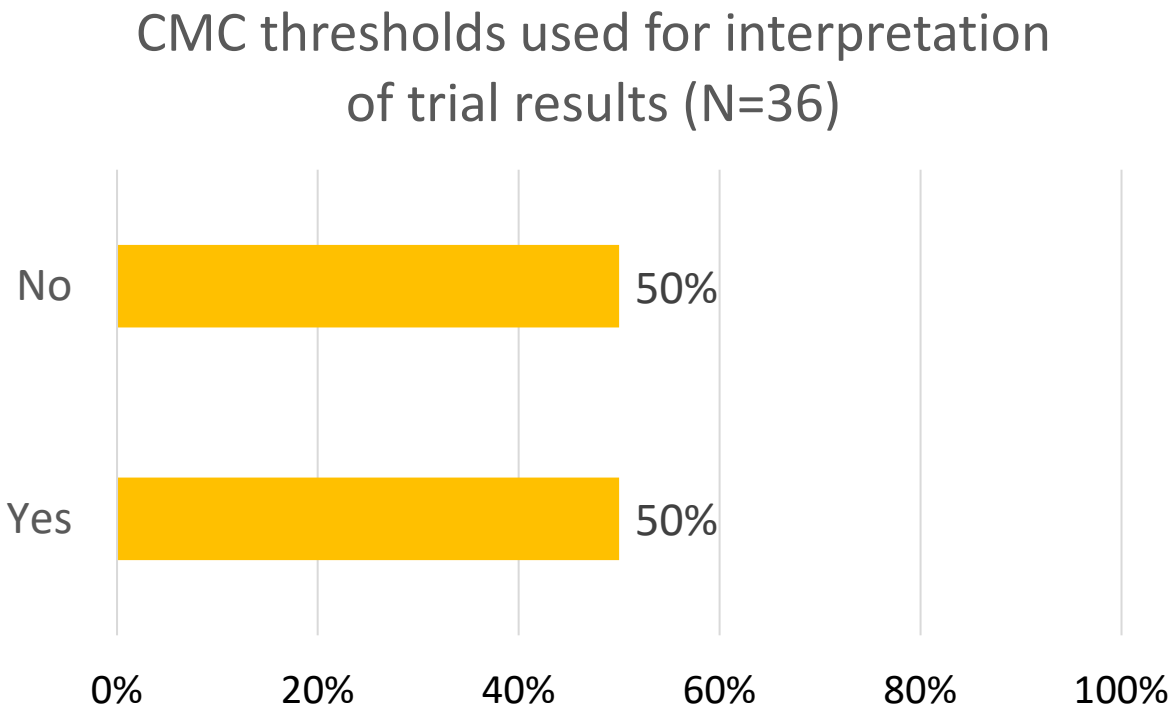
WP 6: Selected results from the review of comparative trials

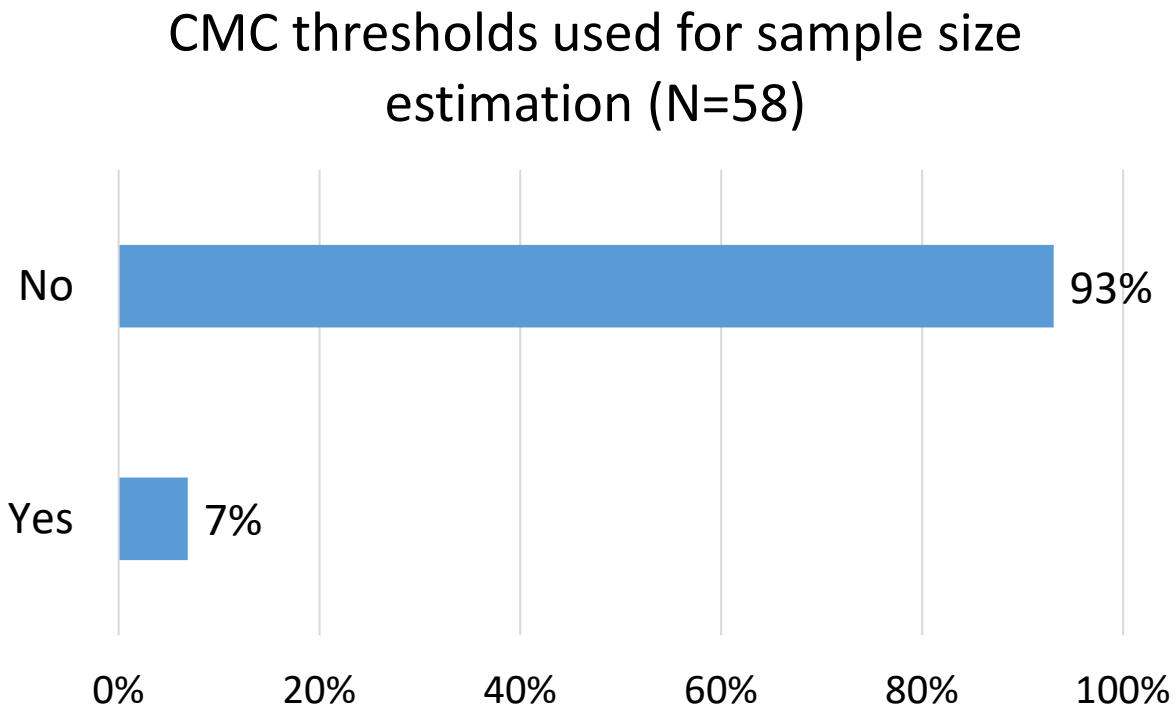


WP 6: Selected results from the review of comparative trials



WP 6: Selected results from the review of comparative trials





Findings: Frequency of CMC terms used in oncology literature

In the **methodology literature,** ***"minimum/ "minimal(ly) important difference"*** were by far the most frequently used terms

Terms related to CMC	Number of mentions
Minimal(ly)/minimum important difference	122
Clinically meaningful (change/difference)	80
Minimum clinically important difference	77
Clinically important / relevant	38
Clinically significant	19
Minimum clinically meaningful change	10
Responder	8
Meaningful (change/difference)	8
Significant	4
Other (sufficient, smallest, noticeable, subjective significance / meaningful, etc.)	19

In the **clinical trial literature**, ***"clinically meaningful"*** terms were used most frequently

Terminology: next steps

When developing terminology, we will need to distinguish different concepts:

- Patient- vs group-level CMC thresholds
- Clinically meaningful thresholds for change vs difference vs difference in change
- Thresholds for minimal CMC vs larger CMC thresholds
- CMC thresholds based on distributions vs anchors (and specify anchors)
- ….

and give a definition of CMC

Draft recommendation framework

Recommendations for the use of CMC thresholds will be developed for different types of trials and endpoints:

	Superiority endpoint			Non-inferiority endpoint	…
	Magnitude of change analysis	Responder analysis	Time-to-event analysis	…	…
Comparative trials					
Recommendation 1					
Recommendation 2					
Recommendation …					
Single-arm trials					
Recommendation….					
Descriptive					
Recommendation….					

Next steps: 2022 through 2024

- Consensus process for recommendation statements
  - Technical language
  - Plain language
- Validation and pilot testing using completed trial datasets
- Visualization and communication of PRO analyses
- Publication of final recommendation statements

Measurement properties and estimated clinically meaningful change thresholds of the *Non-Small Cell Lung Cancer Symptom Assessment Questionnaire* (*NSCLC-SAQ*)

Josephine Norquist, MS – Executive Director, Patients-Centered Endpoints & Strategy, Merck & Co., Inc.

Outline

- High level overview of *NSCLC-SAQ* development, validation, and qualification
- Overview of Phase 3 Keynote-598 and objectives related to *NSCLC-SAQ*
  - Study design and methods
  - Evaluation of
    - Psychometric properties of the *NSCLC-SAQ*
    - Assessment of clinically meaningful thresholds for within-person change in *NSCLC-SAQ* total score
- Overview of results
- Conclusions
- Lessons learned

Teams

Study and Dissemination Team	Name	Role
Merck & Co., Inc.	Thomas Burke, Josephine M. Norquist, Ayman Samkari	Study design, collection and interpretation of data; development of psychometric analysis plan; critically contributing and reviewing report and manuscript writing
IQVIA Patient-Centered Solutions	Paul Williams and Christina Daskalopoulou	*NSCLC-SAQ* psychometric analysis plan and analysis, report and manuscript writing

Dissemination Team	Name	Role
C-Path's PRO Consortium	Sonya Eremenco, Stephen Joel Coons, Rebecca M. Speck	*NSCLC-SAQ* psychometric analysis outline, interpretation of results and critically contributing to manuscript

NSCLC-SAQ timeline

Nov 2016

NSCLC-SAQ included in Keynote-598 as exploratory endpoint

Objective: "to evaluate clinically meaningful thresholds for change in symptoms" using IA1 data pooled across treatment arms

Agreement to use *NSCLC- SAQ* required Merck to send psychometric study report to PRO Consortium within 9 months after DBL

Mar 2017

NSCLC-SAQ qualification package submitted to FDA CDER

NSCLC-SAQ qualification package submitted by PRO Consortium to FDA covering the development of measure , cross-sectional measurement properties, and test-retest reliability

Apr 2018

FDA qualification of the *NSCLC-SAQ*

NSCLC-SAQ granted qualification from the FDA based on qualitative research and cross-sectional observational quantitative research

However, longitudinal data are required to evaluate *the NSCLC-SAQ* total score's ability to detect change in order to be used in drug development programs as an endpoint measure in clinical trials

Sept 2020

Keynote-598 IA1 DBL

Analyses initiated on *NSCLC-SAQ* by IQVIA using IA1 data pooled across treatment arms using pre-specified statistical analysis plan

June 2021

NSCLC-SAQ psychometric analysis report from Keynote-598

Merck's report shared with PRO Consortium per the terms of use agreement

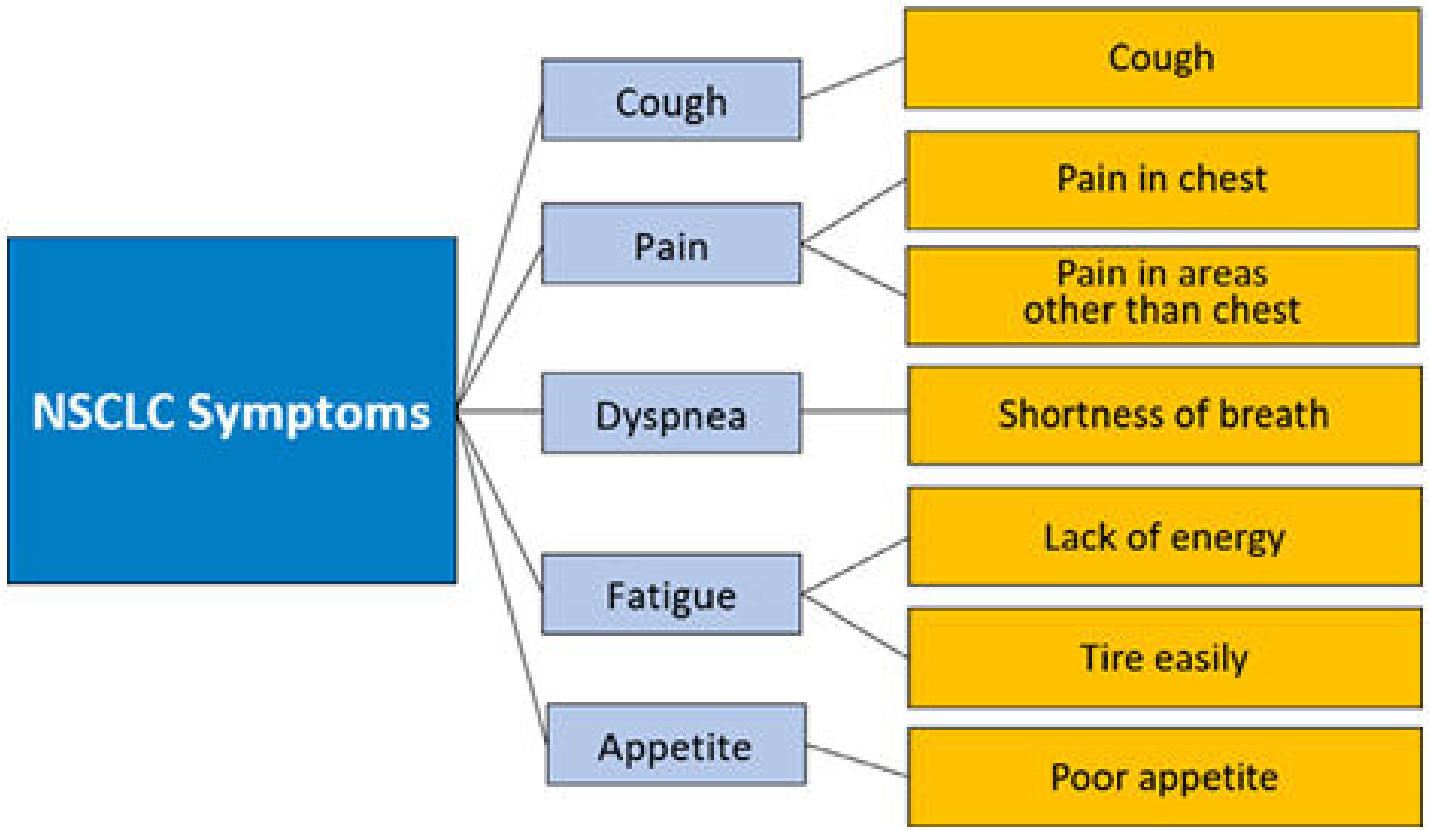
Feb/April 2022

NSCLC-SAQ psychometric manuscript publication

Pre-proof published on-line in Journal of Thoracic Oncology Clinical and Research Reports. Published 01 April 2022

DBL = Database lock; WG = Working Group; IA1 = interim analysis 1

Conceptual framework for the *NSCLC-SAQ*



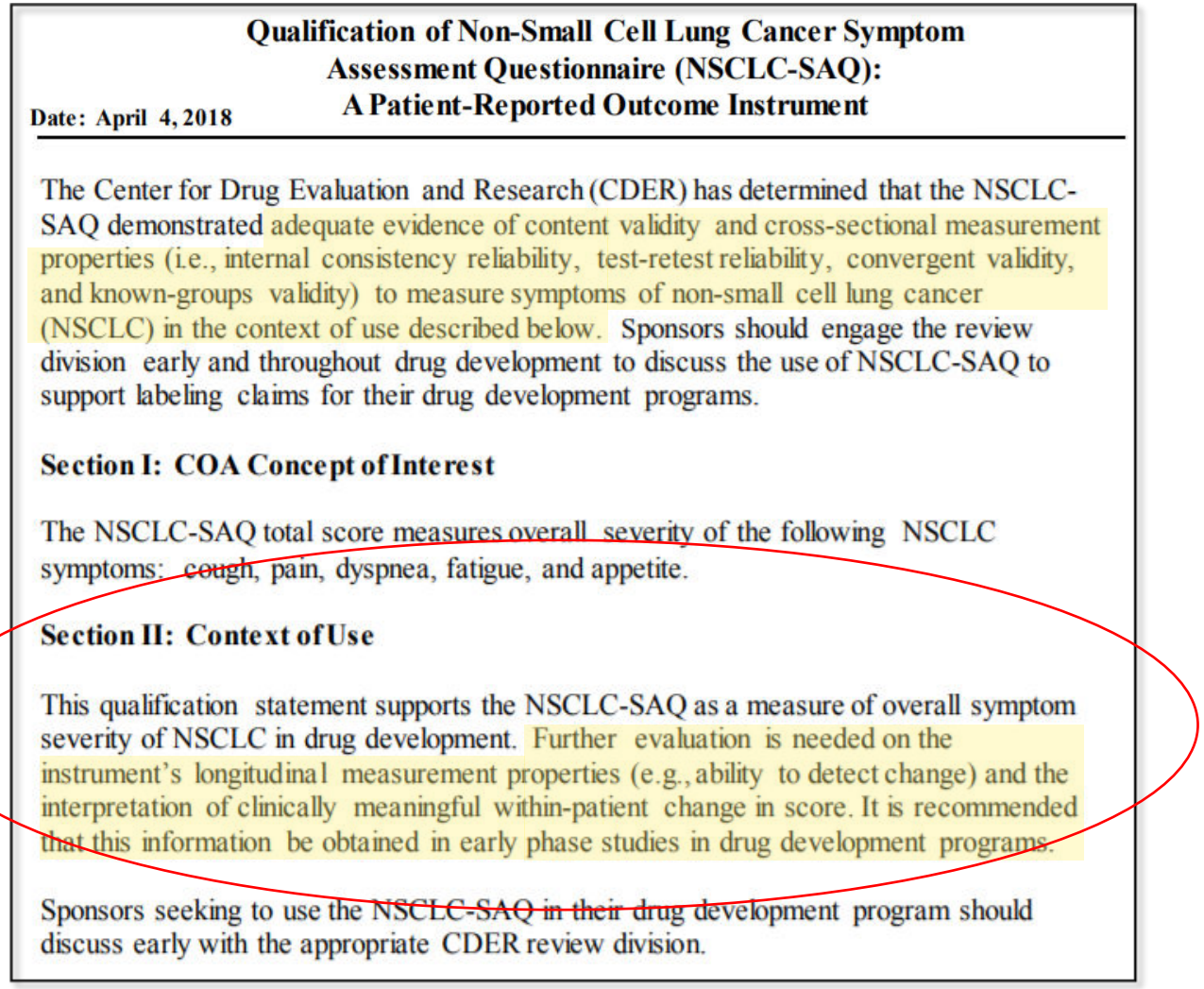
Context of Use

Adult patients (≥ 18 years) with Stage IIIB or IV NSCLC that are:

- Treatment naïve (i.e., treatment naïve to current chemotherapy and not having received chemotherapy for the past 6 months from study enrollment), or
- Treated (i.e., received chemotherapy in the last 6 months and recovered from any prior treatment related toxicities/adverse events to CTCAE v4.03 grade 1 or better)

DM. Bushnell, MA, TM. Atkinson, PhD, KP. McCarrier, PhD, AM. Liepa, PharmD, KP. DeBusk, PhD, SJ Coons, PhD, on behalf of the PRO Consortium's NSCLC Working Group. Non–Small Cell Lung Cancer Symptom Assessment Questionnaire: Psychometric Performance and Regulatory Qualification of a Novel Patient-Reported Symptom Measure. Current Therapeutic Research 95 (2021)

NSCLC-SAQ is the first PRO measure in advanced NSCLC to gain FDA qualification as measure of overall symptom severity of NSCLC



Qualification of Non-Small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ): A Patient-Reported Outcome Instrument

Date: April 4, 2018

The Center for Drug Evaluation and Research (CDER) has determined that the NSCLC-SAQ demonstrated adequate evidence of content validity and cross-sectional measurement properties (i.e., internal consistency reliability, test-retest reliability, convergent validity, and known-groups validity) to measure symptoms of non-small cell lung cancer (NSCLC) in the context of use described below. Sponsors should engage the review division early and throughout drug development to discuss the use of NSCLC-SAQ to support labeling claims for their drug development programs.

Section I: COA Concept of Interest

The NSCLC-SAQ total score measures overall severity of the following NSCLC symptoms: cough, pain, dyspnea, fatigue, and appetite.

Section II: Context of Use

This qualification statement supports the NSCLC-SAQ as a measure of overall symptom severity of NSCLC in drug development. Further evaluation is needed on the instrument's longitudinal measurement properties (e.g., ability to detect change) and the interpretation of clinically meaningful within-patient change in score. It is recommended that this information be obtained in early phase studies in drug development programs.

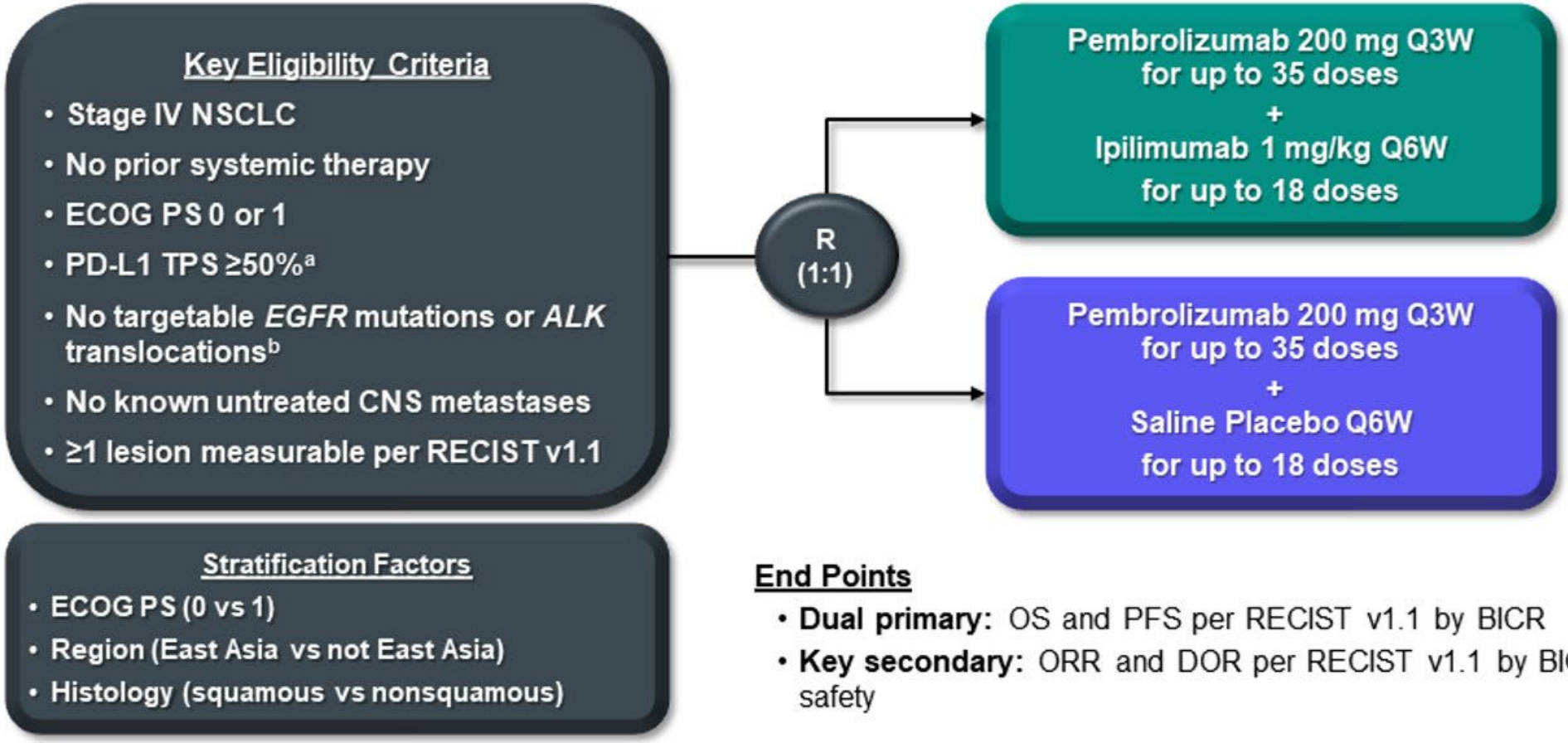
Sponsors seeking to use the NSCLC-SAQ in their drug development program should discuss early with the appropriate CDER review division.

The qualification supports the *NSCLC-SAQ* as a patient-reported measure of symptoms in advanced NSCLC drug development.

Further evaluation is needed regarding the *NSCLC-SAQ*'s longitudinal measurement properties (e.g., ability to detect change) and the interpretation of clinically meaningful within-person score change.

- PRO Consortium website: https://c-path.org/programs/proc/
- Full Qualification Package: https://www.fda.gov/drugs/clinical-outcome-assessment-coa-qualification-program/ddt-coa-000009-non-small-cell-lung-cancer-symptom-assessment-questionnaire-nsclc-saq
- FDA COA Compendium: https://www.fda.gov/drugs/development-resources/clinical-outcome-assessment-compendium

Merck Keynote – 598 Study Design



End Points

- **Dual primary:** OS and PFS per RECIST v1.1 by BICR
- **Key secondary:** ORR and DOR per RECIST v1.1 by BICR and safety

[a]Assessed centrally using the PD-L1 IHC 22C3 pharmDx assay (Agilent).
[b]Patients with *ROS1* rearrangement were also excluded if *ROS1* testing and treatment were locally approved and accessible.
KEYNOTE-598 ClinicalTrials.gov identifier, NCT03302234. BICR, blinded independent central review.

Merck Keynote – 598 Study Design, cont.

- Exploratory objectives on ***NSCLC-SAQ***:
  - To evaluate **clinically meaningful thresholds** for change in lung cancer symptoms from the *NSCLC-SAQ* using the Patient Global Impression of Severity (PGIS-LC) questionnaire and the Patient Global Impression of Change (PGIC-LC) questionnaire
  - To conduct **psychometric analyses** on the *NSCLC-SAQ*
- Other PRO measures included:
  - EORTC QLQ-C30, EORTC QLQ-LC13
  - PGIS-LC and PGIC-LC
- PRO assessment schedule:
  - Baseline, at each 3-weekly treatment cycles until week 18 (cycle 7), less frequently thereafter
    - *NSCLC-SAQ*, PGIS-LC, and PGIC-LC administered up until week 18

Analyses related to the *NSCLC-SAQ* psychometric validation

- Conducted on the PRO full analysis set (FAS) population* on blinded data with pooled treatment arms from interim efficacy analysis 1 data (data cut-off date: 1-Sept-2020). N=560 patients comprised the PROFAS population
- Cross-sectional analyses were performed using cycle 1 (baseline) and week 18 (cycle 7) data
- Longitudinal analyses used change from baseline to week 18 (cycle 7)

* defined as all patients who had at least one PRO assessment available and had received at least one dose of study medication.

Estimating clinically meaningful thresholds for change in *NSCLC-SAQ* total score

PRO CONSORTIUM CRITICAL PATH INSTITUTE

Primary method

- Mean change total *NSCLC-SAQ* score at week 18 by change in the **PGIS-LC** at week 18 to establish improvement or deterioration

Supportive methods

- Empirical cumulative distribution function (eCDF) and probability density function (PDF) curves
- Effect size and standardized response mean

PGIS-LC 2.0*

How would you rate your symptoms of your lung cancer at this time?

- ☐ No symptoms
- ☐ Mild
- ☐ Moderate
- ☐ Severe
- ☐ Very severe

*PGIS-LC v 1.0 response categories = "Not severe," "Mildly severe," "Moderately severe," "Very severe," "Extremely severe"
PGIC-LC v 2.0: "Compared to your first study visit, how would you describe the symptoms of your lung cancer today?"

- 7 response options from "Much better" to "Much worse"

PRO-FAS population at baseline

Parameters/Categories	PRO FAS (N=560)	
Age (years) N Mean (SD) Median (min-max)	 560 64.0 (9.06) 65 (35-85)	
Gender, n (%) Female	 170 (30.4%)	
ECOG PS, n (%) 0 1	 203 (36.3%) 357 (63.8%)	
Histology, n (%) Squamous Non-Squamous	 157 (28.0%) 403 (72.0%)	
PGIS-LC, n (%)	**v 1.0**	**v 2.0**
	Not severe = 89 (25.9%)	No Symptoms = 38 (19.8%)
	Mildly severe = 100 (29.2%)	Mild = 66 (34.4%)
	Moderately severe = 108 (31.5%)	Moderate = 57 (29.7%)
	Very severe = 39 (11.4%)	Severe = 25 (13.0%)
	Extremely severe = 7 (2.0%)	Very Severe = 6 (3.1%)
	Missing = 217	Missing = 368

NSCLC-SAQ generally met the psychometric validation criteria in Keynote-598

Evaluation	Description	Results
Item response distributions	Measure of central tendencies (mean, median), variance (SD, normality, skewness, and kurtosis), min/max, outliers, proportion of responses/level.	*NSCLC-SAQ* item means similar at baseline Some ceiling effects indicating presence of low symptom severity No floor effects indicating high symptom severity observed at baseline
Inter-item associations	Correlation among questions within an instrument and the extent to which they belong together for scoring purposes	Inter-item Spearman correlations ranged from 0.11 (chest pain/cough) to 0.82 (tired/no energy)
Internal consistency reliability	Relationships among items and the extent to which they belong together for scoring purposes. Evaluated using Item-total and domain-total score correlation and Cronbach's alpha at baseline and week 18	Cronbach's alpha > 0.70 at both baseline (0.74) and week 18 (0.78) Item-total correlations ranged from 0.37 to 0.71 (baseline) and 0.43-0.72 (week 18)
Test-retest reliability	Stability of scores over time when no change is expected in the concept of interest. Evaluated using the ICC of *NSCLC-SAQ* scores from cycle 1 (week 0) and 2 (week 3) in participants with no change in PGIS-LC v 1.0 or v 2.0 in this period	ICC = 0.79; n=210

SD = standard deviation; ICC = intraclass correlation coefficient; ECOG PS = Eastern Cooperative Group Performance Status; ES = effect size; SRM = standardized response mean

NSCLC-SAQ generally met the psychometric validation criteria in Keynote-598, cont.

Evaluation	Description	Results
Construct Validity Convergent/ Divergent	Extent to which observed correlations among measures match hypothesized correlations in terms of sign and magnitude. Assessed with Spearman correlation coefficients between *NSCLC-SAQ* scores with items in other measures at baseline	Moderate (≥ 0.4) to high (≥ 0.6) correlations with similar domains/scores from other PRO measures; Lower correlations with less similar concepts from other PRO measures
Known-groups validity	The degree to which scores can distinguish among known groups hypothesized *a priori* to be different. Evaluated by comparing the mean *NSCLC-SAQ* score stratified by PGIS-LC responses, matched EORTC QLQ-C30/LC13 scales and ECOG PS at cycle 1	Higher (more severe) *NSCLC-SAQ* scores observed for • higher (more severe) PGIS-LC responses • lower (worse overall health) QLQ-C30 GHS/QoL scores • higher (lower level of functionality) ECOG PS
Responsiveness	Evidence measures are sensitive to detecting change. Responsiveness of the *NSCLC-SAQ* total score was assessed by identifying groups of participants who changed according to PGIS-LC/PGIC-LC collected at both baseline and week 18 for participants completing the same version of PGIS-LC/PGIC-LC at baseline and Week 18	Statistically significant mean change scores from baseline to week 18 in *NSCLC-SAQ* total score by PGIS-LC change scores groups. ES and SRM were moderate to large (>0.5) for participants improving by 1 or greater than 1 PGIS-LC category

SD = standard deviation; ICC = intraclass correlation coefficient; ECOG PS = Eastern Cooperative Group Performance Status; ES = effect size; SRM = standardized response mean

Interpretation of meaningful within-person change in score

NSCLC-SAQ Total Score Change from baseline to week 18 by PGIS-LC responses

PGIS-LC anchor response category	n*	Distribution of NSCLC-SAQ change score**	
		Mean (95% CI)	Median (Q1-Q3)
All change response categories			
<-2 (improvement)	5	-3.70 (-8.38 – 0.98)	-2.00 (-6.00 – -2.00)
-2	36	-5.90 (-7.10 – -4.71)	-5.50 (-8.50 – -3.75)
-1	68	-2.65 (-3.52 – -1.78)	-2.25 (-5.00 – 0.00)
0 (no change)	108	-1.28 (-1.99 – -0.57)	-1.00 (-3.00 – 1.00)
1	31	1.05 (-0.28 – 2.38)	1.00 (-1.50 – 3.00)
2	7	0.79 (-1.54 – 3.11)	1.00 (-1.00 – 2.00)
>2 (worsening)	0	(NC)	(NC)

-2.65, -2.25:

- Smallest estimate derived from 1-level PGIS-LC response category improvement

-5.50, -5.90:

- Smallest estimate derived from 2-level PGIS-LC response category improvement

Recommended threshold: Range of 3-5 points improvement on a 0-20 *NSCLC-SAQ* total score range

CI = Confidence interval; PGIS-LC = Patient Global Impression of Severity – Lung Cancer Symptoms; Q1 = first quartile; Q3 = third quartile; NC: not calculated

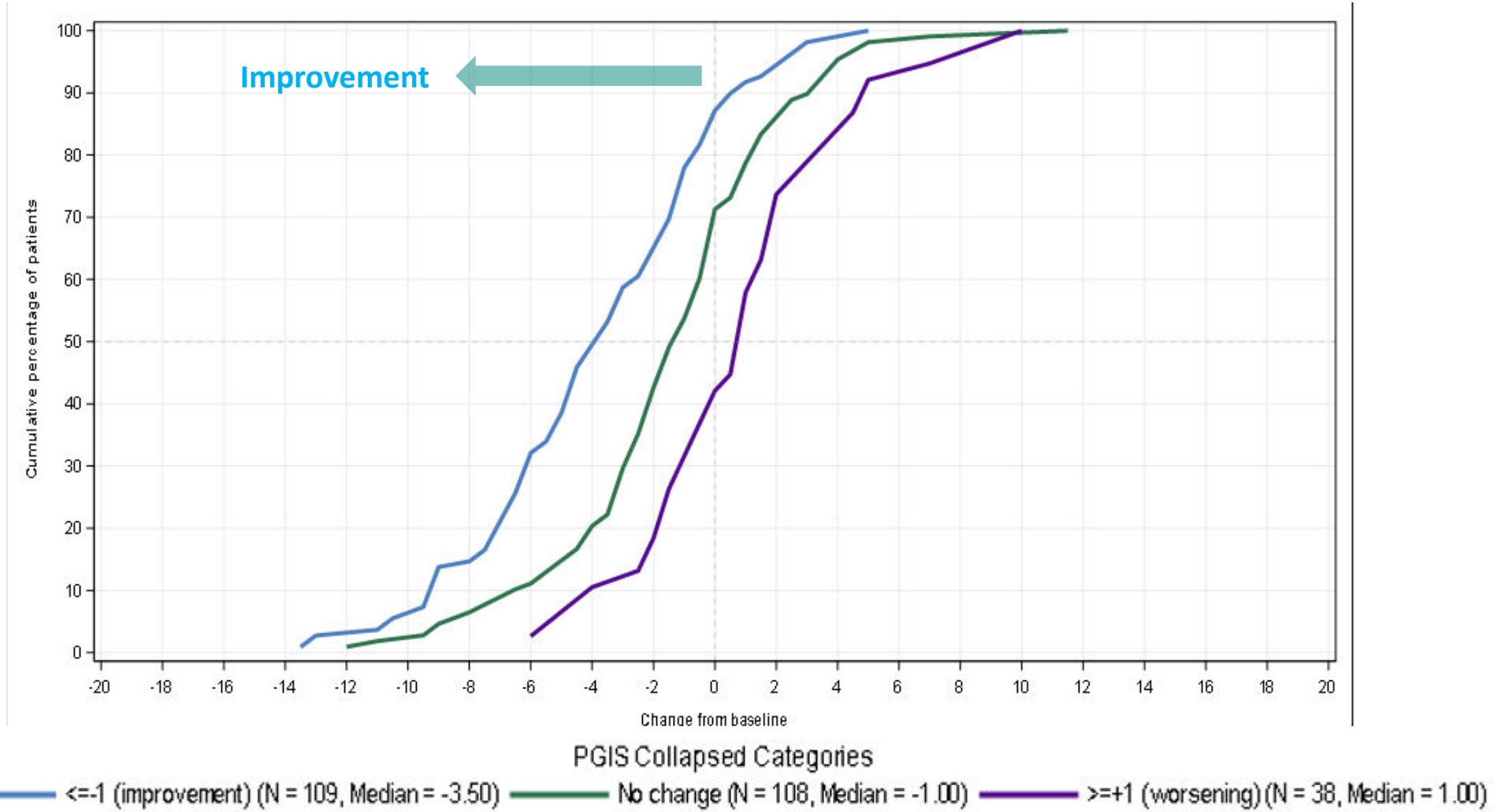
* n describes the number of participants from the PROFAS population who completed the same version of PGIS-LC at baseline and week 18.

** Change scores are calculated as week 18 score minus baseline score.

eCDF plot at week 18 showed clear separation between curves which was consistent throughout the range of change

Anchor groups based on PGIS-LC change from baseline to week 18 (in participants [n=255] who completed the same version at both timepoints)

Source: Figure 2.10.1



Clinically meaningful within-person change threshold for *NSCLC-SAQ* total score

Threshold range of 3-5 points improvement (~ 15-25% on the 0-20 scale) supported by:

- Mean and median change scores for groups of participants who had improved small (-1) / small-moderate (-2 or -1) amount on primary anchor (PGIS-LC)
- Excludes the majority of the "no change" group (25th percentile: -3; mean: -1.28, 95% CI: -1.99 to -0.57)
- Understanding thresholds around symptom worsening was limited
  - Small sample of participants worsened and large variance around the *NSCLC-SAQ* total change score for this group

Sponsors should engage the review division early and throughout drug development to discuss the use of *NSCLC-SAQ* to support labeling claims for their drug development programs

Conclusions

- **Novel evidence from this study**
- Psychometric results showed **consistency with previous findings** of adequate cross-sectional measurement properties
- Data contribute to addressing a gap in **defining meaningful within-person change** in *NSCLC-SAQ* total score in the target patient population
- Main limitations of this study were
  - symptom worsening, as assessed by all PRO measures, not generally observed
  - few participants reported high severity in symptoms at baseline
  - PGIS-LC and PGIC-LC version changed during the study

Manuscript published on 1 April, 2022, with data to support the measurement properties of the *NSCLC-SAQ* including defining meaningful within-person change

Lessons learned

Prior to Study Start

- Obtain internal cross-functional alignment on including a new PRO measure in clinical trial
- Ensure a license agreement is in place with the PRO Consortium to include the *NSCLC-SAQ*
  - Coordinate with PRO Consortium for translations using their translation vendor
- Ensure wording of the anchor measures (PGIS/PGIC) are clearly defined before the start of the study
  - In Keynote-598 the PGIS-LC response categories needed to be changed mid-study to align with FDA feedback

During Protocol Development

- Pre-specify timepoint of interest
  - In Keynote-598, week 18 was pre-specified within the protocol; most *NSCLC-SAQ* score changes occurred within first 6 weeks
  - Administration of the *NSCLC-SAQ* and PGIS/PGIC only up to week 18 (compromise on assessment times) while all others PRO measures continued to be administered
- As *NSCLC-SAQ* was never used at Merck, new data specification for inclusion in a RCT was required
- Internal alignment was necessary for the *NSCLC-SAQ* analyses
  - In Keynote-598 these psychometric analyses were pre-specified in a separate analysis plan and conducted by IQVIA
  - Datasets were not transferred to IQVIA but kept on Merck's server for IQVIA to access

Before/After IA1/DBL

- Per the terms of use agreement, Merck provided a psychometric analysis outline to PRO Consortium prior to IA1/DBL and then a summary report within 9 month from IA1/DBL

Meaningful within-person change thresholds of the Worst Itching Intensity Numerical Rating Scale (WI-NRS) for assessing itch in patients with chronic kidney disease-associated pruritus

Margaret Vernon, PhD – Senior Vice President, General Manager, Evidera, Inc.

Outline

- Background on chronic kidney disease-associated pruritis (CKD-aP)
- Introduction of Worst Itching Intensity Numerical Rating Scale (WI-NRS) to assess clinical benefit
- Methods used to ascertain meaningful change threshold in CKD-aP
  - Phase 2 psychometric analysis
  - Phase 3 psychometric analysis
  - Phase 3 exit interview study
    - 2 interview approaches
- Results and triangulation of evidence
- Key take-aways

Background

- CKD-aP is a common and distressing symptom in people with CKD receiving hemodialysis
  - Affects more than 60% receiving hemodialysis
  - 20-40% have moderate to severe pruritis
- No approved treatment (historically), and not adequately controlled by topical or oral antihistamines or steroids, or off-label treatments
- Cara Therapeutics, an early commercial-stage biopharmaceutical company developed KORSUVA™ (difelikefalin) injection for treatment of pruritis

Assessing clinical benefit

- Pruritis is a symptom that only patients themselves can report on
  - A patient-reported outcome (PRO) measure is required to evaluate efficacy of any new investigational treatment

Worst Itching Over the Past 24 Hours

Please indicate the intensity of the **WORST ITCHING** you experienced over the past 24 hours.

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
NO ITCHING										WORST ITCHING IMAGINABLE

- While psychometric properties and meaningful within-person change thresholds have been evaluated in psoriasis and atopic dermatitis, these have not been evaluated in CKD-aP

Evolution of endpoint definition and Agency interactions

Clinical Trial	Primary endpoint
Phase 2 CR845-CLIN2005	Visual analog scale (100-mm) **absolute change from baseline** to the average at Week 2, completed at morning and night
Phase 2/3 CR845-CLIN2101	**Change from baseline** in the weekly mean of the daily 24-hour WI-NRS to the last week of the treatment period (Week 8 for Part A, Week 12 for Part B)
Phase 3 CR845-CLIN3102	Reduction of itch intensity as assessed by the **proportion of patients achieving an improvement from baseline ≥3 points** with respect to the weekly mean of the daily 24-hour WI-NRS at Week 12
Phase 3 CR845-CLIN3103	

- Type C guidance meeting 14-Dec-2015

the summary results of the completed Phase 2 clincial trial (CR845-CLIN2005). The Agency provided several recommendations on PRO endpoints and the proposed development program. In particular, the Agency recommended the sponsor to use a responder analysis instead of mean change from baseline, as interpretation of mean change is problematic and would unlikely be acceptable. Post the EOP2 meeting, the sponsor submitted a protocol for

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2021/214916Orig1s000AdminCorres.pdf

Evolution of endpoint definition and Agency interactions

Clinical Trial	Primary endpoint
Phase 2 CR845-CLIN2005	Visual analog scale (100-mm) **absolute change from baseline** to the average at Week 2, completed at morning and night
Phase 2/3 CR845-CLIN2101	**Change from baseline** in the weekly mean of the daily 24-hour WI-NRS to the last week of the treatment period (Week 8 for Part A, Week 12 for Part B)
Phase 3 CR845-CLIN3102	Reduction of itch intensity as assessed by the **proportion of patients achieving an improvement from baseline ≥3 points** with respect to the weekly mean of the daily 24-hour WI-NRS at Week 12
Phase 3 CR845-CLIN3103	

- End of phase 2 meeting 6-Sep-2017

Meeting Discussion:

There was general discussion about defining the primary endpoint and the success criterion. The Agency clarified that the primary endpoint should be based on a responder definition, which is based in change from baseline to Week 12 with a threshold level of 4 point change. The sponsor advocated using a 2 point change. The Agency responded that the mean change for subjects in the placebo arm from your Phase 2 trial was 1.9. The Agency proposed that the sponsor may submit data from their clinical trials to support their argument of using a lower threshold level than the Agency recommended 4 point change. The sponsor should also provide a rationale for anchoring the worst itch intensity NRS to the PGIC "minimally improved" anchor response category.

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2021/214916Orig1s000AdminCorres.pdf

Methods overview

Phase 2 analysis

- Patient Global Impression Severity (PGIS) and Patient Global Impression Change (PGIC) used as anchors to evaluate meaningful within-person change of WI-NRS

Phase 3 analysis

- Confirmatory anchor-based analysis using PGIC

Phase 3 exit interviews

- Modified-PGIC approach (Koochaki et al. 2018)
- Review of WI-NRS change (McCarrier et al. 2019)

Phase 2 secondary analysis

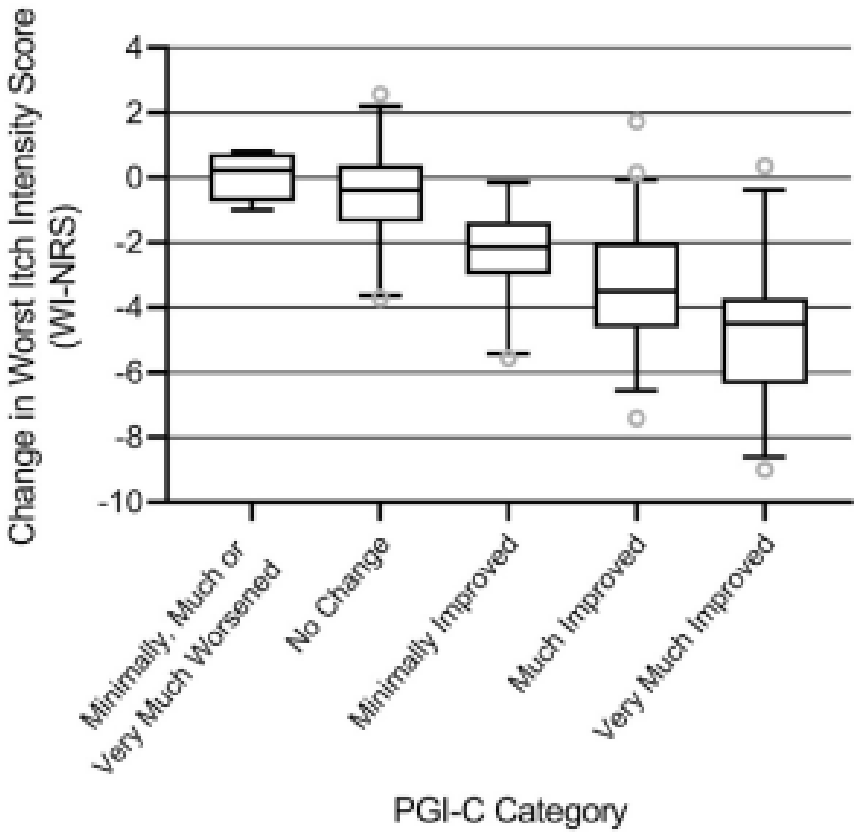
RESEARCH LETTER | VOLUME 84, ISSUE 4, P1132-1134, APRIL 01, 2021

Clinically meaningful change in itch intensity scores: An evaluation in patients with chronic kidney disease–associated pruritus

Margaret Vernon, PhD • Sonja Ständer, MD • Catherine Munera, PhD • Robert H. Spencer, PhD • Frédérique Menzaghi, PhD

J Am Acad Dermatol. 2021 Apr;84(4):1132-1134.

Criteria	WI-NRS change score (week 8 – baseline), mean	Change from baseline, mean, %	Effect size (Cohen *d*)
Primary anchor			
PGI-C minimally improved	−2.26	−33.56	1.29
PGI-C minimally and much improved	−3.02	−42.99	1.65
PGI-C much improved	−3.41	−47.81	1.83
Secondary anchors			
PGI-S improved 1 point	−2.49	−37.10	1.40
PGI-S improved at least 1 point	−3.45	−49.61	1.75



PGIC: Since the start of the study, my itch is?

PGIS: Please choose from the following options the one that best describes your worst itch in the last 24 hours. response options: Very severe, severe, moderate, mild, none

Phase 3 analysis

Psychometric validation and meaningful change thresholds of the Worst Itching Intensity Numerical Rating Scale for assessing itch in patients with chronic kidney disease-associated pruritus

Margaret K. Vernon[1], Laura L. Swett[1], Rebecca M. Speck[1], Catherine Munera[2], Robert H. Spencer[2], Warren Wen[2] and Frédérique Menzaghi[2*]

Journal of Patient-Reported Outcomes (2021) 5:134

Criteria	N	Mean WI-NRS change score[a] (SD)	Mean % change from baseline	Effect size (Cohen's d)
Primary anchor-based approach				
PGI-C minimally improved	198	−1.85 (1.73)	−25.73	1.09
Secondary anchor-based approach				
PGI-C much improved	209	−3.54 (2.08)	−51.02	2.04
PGI-C minimally or much improved	407	−2.72 (2.09)	−38.72	1.48

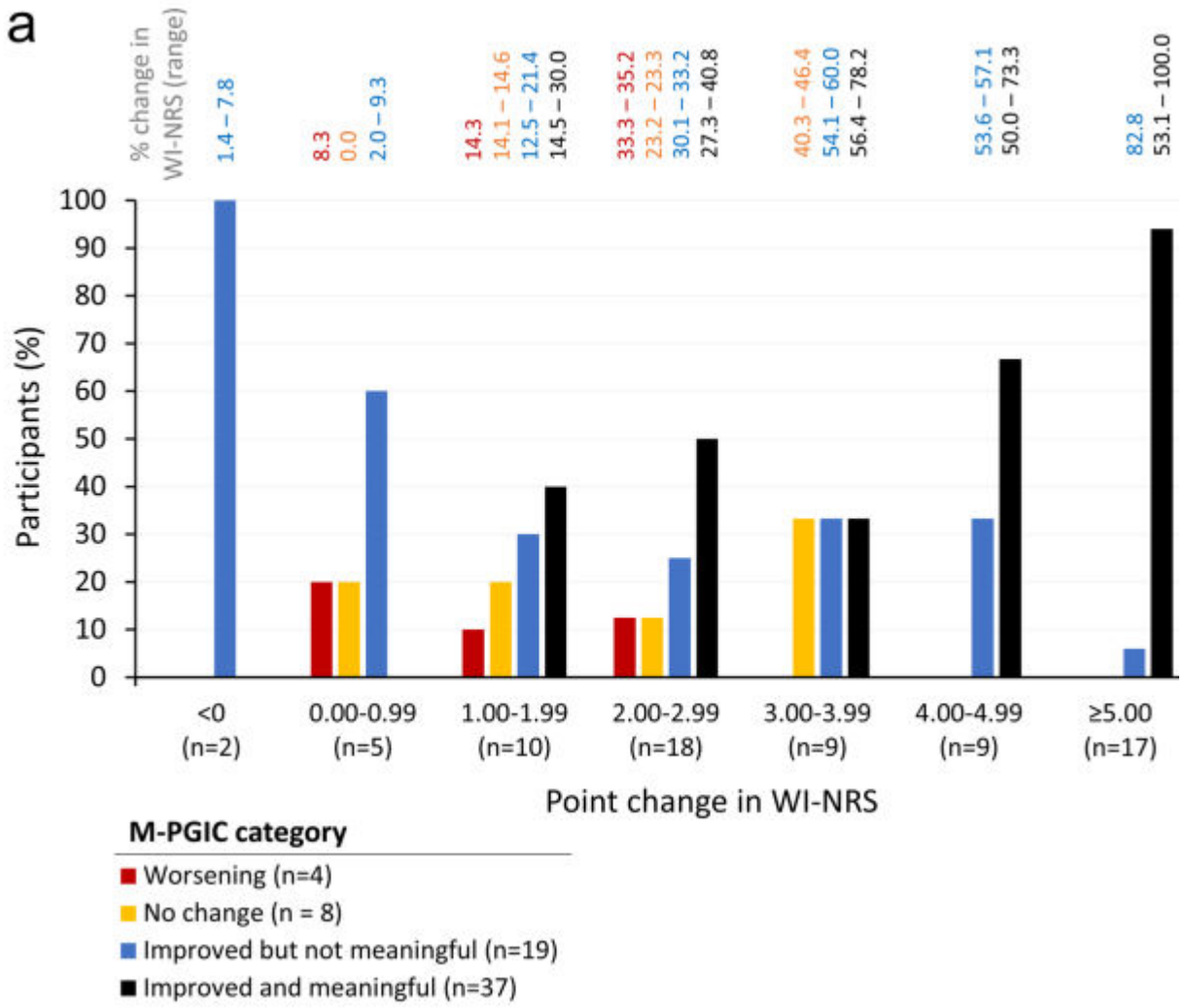
PGI-C, Patient global impression of change; SD, standard deviation; WI-NRS, Worst Itching Intensity Numerical Rating Scale

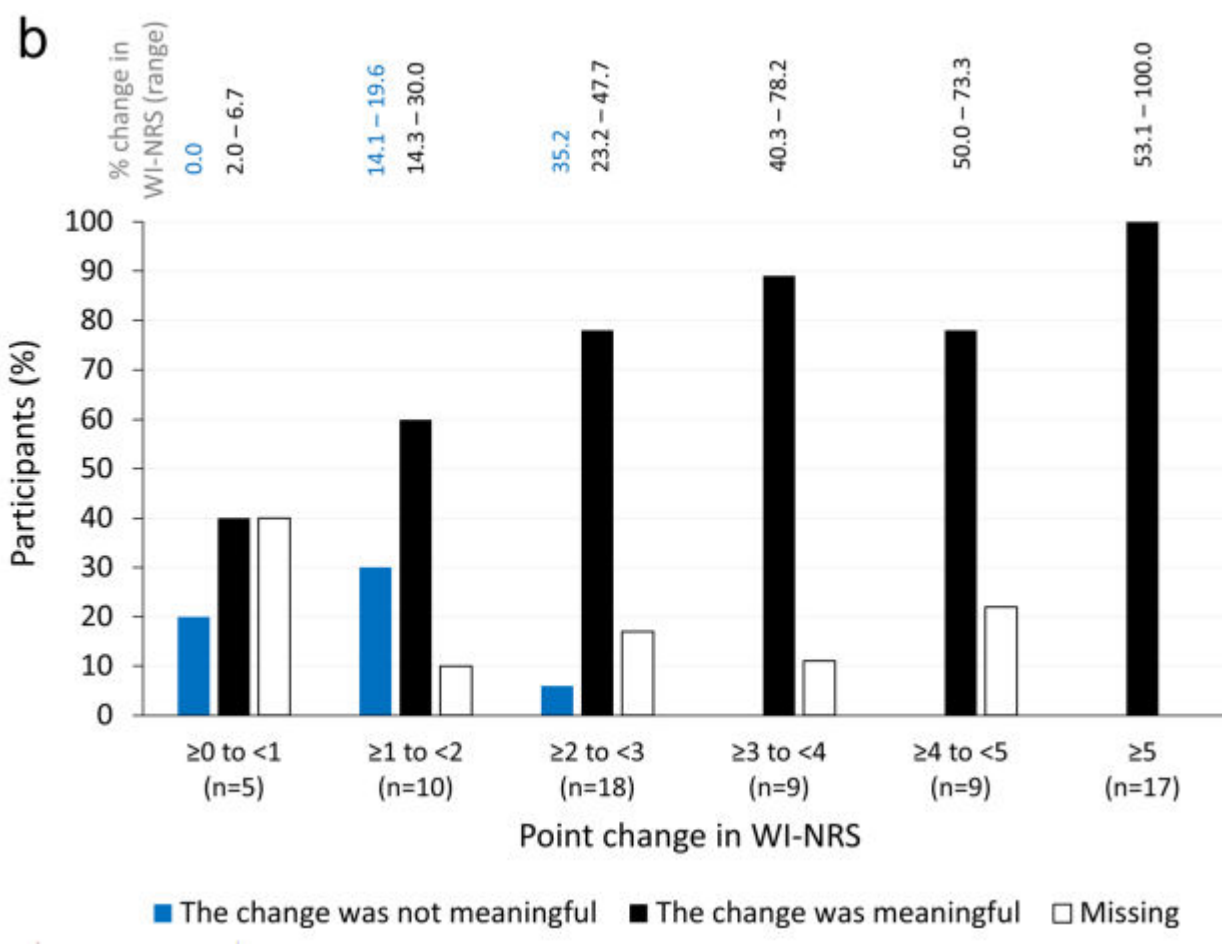
[a] Change from baseline to end of treatment

Phase 3 exit interviews

- Mixed-method telephone-based exit interviews (N=70), enrollment stratified to represent different levels of WI-NRS change
  - 10–12 patients reporting a 1-point improvement and 15–20 reporting a 2-, 3-, and 4-point improvement on the WI-NRS from baseline to Week 8–10
- Complete Modified Patient Global Impression of Change (M-PGIC)
  - “My itch got worse”
  - “No change”
  - “My itch got better but the amount of improvement was not meaningful to me”
  - “My itch got better and the amount of improvement was meaningful to me”
- Review and discuss their actual WI-NRS change from clinical trial

Phase 3 exit interviews





Triangulation of results

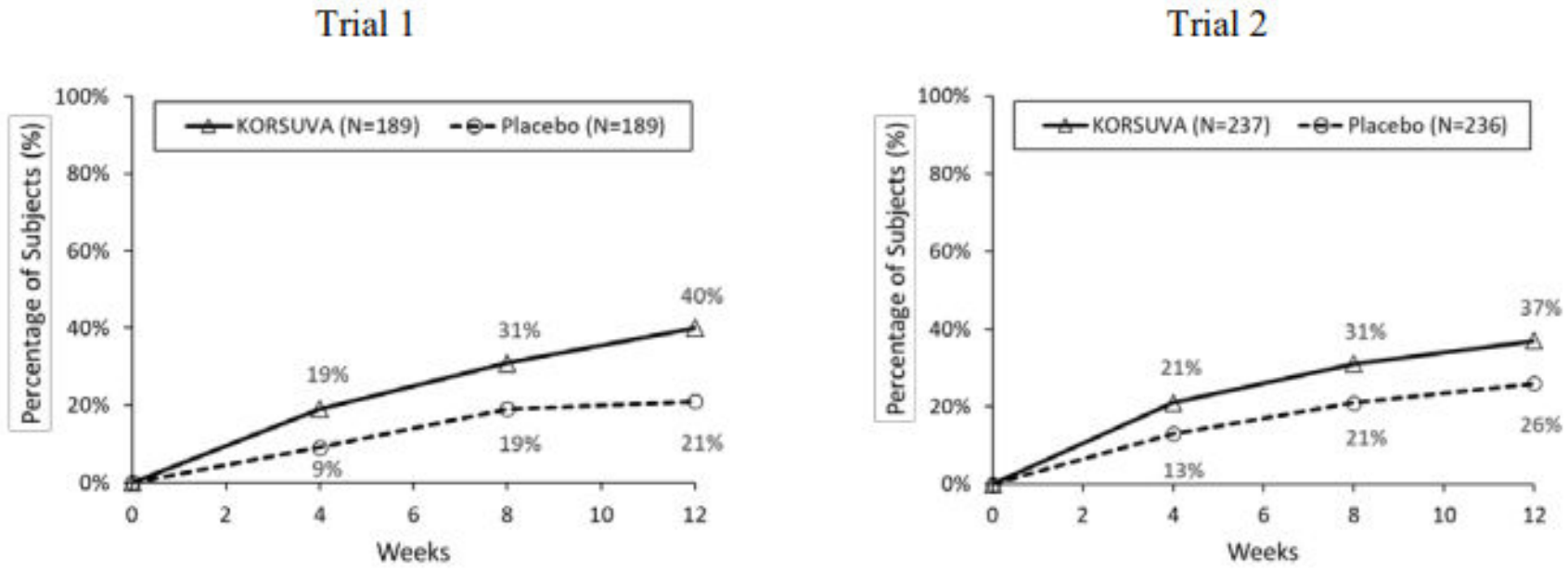
Method	Results
Phase 2 analyses	Meaningful change estimates ranged from 2.26 (minimally improved) to 3.45 (much improved) based on PGIC; 2.49 based on a 1-point improvement on a 5-point PGIS
Phase 3 analyses	Meaningful change estimates ranged from 1.85 (minimally improved) to 3.54 (much improved) based on PGIC
Phase 3 exit interviews	When reviewing their actual change in WI-NRS, the majority of patients with a 2-point improvement on the WI-NRS found the change to be meaningful; all patients with a 3-point or greater improvement found the change to be meaningful

KORSUVA label claim

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use KORSUVA safely and effectively. See full prescribing information for KORSUVA.

KORSUVA™ (difelikefalin) injection, for intravenous use
Initial U.S. Approval: 2021

Figure 1: Percentage of Subjects with Moderate-to-Severe CKD-aP Undergoing HD with a ≥4-point Improvement from Baseline on the WI-NRS in Trial 1 and Trial 2



Itch reduction was seen by Week 4 and sustained through Week 12.

https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/214916s000lbl.pdf

Take-aways

- Come early and come often!
  - Endpoint definition evolved, Agency advice was critical
- Include anchors (PGIS, PGIC) in Phase 2 and leverage that data
- Look at all levels of change on your anchors
- Plan for the possibility of exit interviews
  - Opportunity to leverage real data and real patients and avoid the hypothetical
- Benefits and drawbacks of different methods; different methods yield somewhat different results

Panel Discussion

Moderator

- *Rebecca M. Speck, PhD, MPH* – Clinical Outcome Assessment Scientist, Clinical Outcome Assessment Program, C-Path

Presenters

- *Carla Mamolo, PhD* – Director, Health Economics & Outcomes Research, Pfizer, Inc.
- *Johannes Giesinger, PhD* – Assistant Professor, Medical University of Innsbruck
- *Josephine Norquist, MS* – Executive Director, Patient-Centered Endpoints & Strategy Lead, Merck & Co., Inc.
- *Margaret Vernon, PhD* – Senior Vice President, General Manager, Evidera, Inc.

Additional Panelists

- *Selena Daniels, PharmD, PhD* – Clinical Outcome Assessment Team Leader, Division of Clinical Outcome Assessment, U.S. Food and Drug Administration
- *Lili Garrard*, *PhD* – Lead Mathematical Statistician, Division of Biometrics III, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration

Thank you!
Day 1 Wrap Up