ISPOR Clinical Outcomes Assessment and Clinician-Reported Outcomes Emerging Good Research Practices Task Force
ISPOR COA & ClinROs Task Force Leadership Group

in alphabetical order:

Laurie B. Burke, RPh, MPH, Founder, Lora Group, LLC, Royal Oak, MD, USA and Affiliate Associate Professor, University of Maryland School of Pharmacy

Stefan Cano, PhD, CPsychol, AFBPsS, Chief Science Officer, Modus Outcomes, Stotfold, England, UK

Jeremy Hobart, PhD, FRCPJ, Professor of Clinical Neurology and Health Measurement, Peninsula College of Medicine and Dentistry, Plymouth, UK

Maria Isaac, MASC, MD, PhD, Scientific Administrator, Science Advice & Orphan Drugs Sector (SAOD), European Medicines Agency (EMA), London, UK

Patrick Marquis, MD, MBA, Chief Executive Officer, Modus Outcomes Boston, MA, USA

Elizabeth Molsen, RN, Director, Scientific & Health Policy Initiatives, ISPOR, Lawrenceville, NJ, USA
ISPOR COA & ClinRO Task Force Leadership Group (2)

Elektra Papadopoulos, MD, MPH, Acting Associate Director for Study Endpoints Review, ONDIO, CDER, FDA, Silver Spring, MD, USA

Donald Patrick, PhD, MSPH, Professor, Department of Health Services and Director, Seattle Quality of Life Group, University of Washington, Seattle, WA, USA

John H. Powers, III, MD FACP FIDSA - Chair
Associate Clinical Professor of Medicine, George Washington University School of Medicine, Washington, DC, USA

Ashley Slagle, PhD, MS, Study Endpoints Reviewer, ONDIO, CDER, FDA, Silver Spring, MD, USA

Marc K. Walton, MD, PhD, Senior Scientific Director, Janssen Research and Development, Titusville, NJ, USA
Task Force Objective

Develop an Emerging Good Research Practices Report to:

• Further improve development and use of clinical outcome assessments (COAs), including ClinROs in US and European clinical research and practice

• Improve patient outcomes and allow better assessments of benefits and harms through better measures
Why COAS?

- Rising costs require evidence of value to all stakeholders.
- Payers interested in technology assessment and evidence-based treatments.
- Consumer and patient voice rising in importance with patient-centeredness.
- Regulators like US Food and Drug Administration and European Medicines Agency require for approval.
ISPOR TASK FORCE REPORT


Marc K. Walton, MD, PhD1,*, John H. Powers III, MD, FACP, FIDSA2, Jeremy Hobart, PhD, FRCP3, Donald Patrick, PhD, MSPH4, Patrick Marquis, MD, MBA5, Spiros Vamvakas, MD6, Maria Isaac, MASc, MD, PhD6, Elizabeth Molsen, RN7, Stefan Cano, PhD, CPychol, AFBPsS8, Laurie B. Burke, RPh, MPH9,10

1Janssen Research and Development, Titusville, NJ, USA; 2Leidos Biomedical Research in support of the Division of Clinical Research, National Institutes of Health, Bethesda, MD, USA; 3Plymouth University Peninsula Schools of Medicine and Dentistry, Devon, UK; 4Seattle Quality of Life Group, Department of Health Services, University of Washington, Seattle, WA, USA; 5Modus Outcomes, Newton, MA, USA; 6European Medicines Agency, London, UK; 7International Society for Pharmacoeconomics and Outcomes Research, Lawrenceville, NJ, USA; 8Modus Outcomes, Stotfold, UK; 9LORA Group, LLC, Royal Oak, MD, USA; 10Department of Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, USA
Background

- 2009: FDA PRO Guidance for Industry
  - “A PRO instrument, like physician-based instruments, should be shown to measure the concept it is intended to measure, and the FDA will review the evidence that a particular PRO instrument measures the concept claimed.”
- 2011: Review and Qualification of Clinical Outcome Assessments; Public Workshop
- 2014: Reflection Paper on the Use of Patient Reported Outcome (PRO) Measures in Oncology Studies
Foundational Topics Covered:

- Treatment Benefit
- The Relationship Between Patient Assessments and Study Endpoints
- The Intended Treatment Benefit
- The Measured Concept of Interest
- Context of Use: Important Clinical Trial Factors
- Attributes of an OA
Definitions

- **Treatment benefit**
  - Favorable effect on meaningful aspect of how patients *feel* or *function* in their lives, or how long patients *survive*
    - Improved feel, function, survive
    - Not worsened feel, function, survive related to disease
    - Decreased harm compared to an alternative treatment
  - **Meaningful aspect** – Meaningful to the person
  - **In their life** – In the patient’s usual (typical) life
Types of Outcome Assessments

- Survival

- PRO – Patient-Reported Outcome
- ClinRo – Clinician-Reported Outcome
- ObsRO – Observer-Reported Outcome
- PerfO - Performance Outcome

COA: Assessment that depends on Rater judgment or Patient active participation to perform measured action

- Biomarkers
  - Survival and biomarkers not in scope of this TF
“Endpoint” - a particular method for assessment of the patients at one or more times during the study and analyzed according to a specified statistical method to provide a comparison between groups NOT the assessment itself

“Outcome assessment” - the measuring instrument that provides a rating or a score that is intended to represent some aspect of the patient’s medical status
Context of Use

• The specifics of study design and other study features
  – Disease or condition
  – Patient subpopulation
  – Culture, language, geography
  – Standard clinical practice
  – Endpoint positioning
  – Manner of use within the endpoint
  – Measurement setting
  – Method of administration

• The context of use may alter the OA performance characteristics and interpretation
Attributes of the Outcome Assessment

Influence the measurement properties and how the measurement properties are evaluated

- Whether the OA is dependent on patient active involvement or rater participation
- Whether the rater applies judgment to form the measurement
- Relationship between the OA and treatment benefit (feels, function, survives)
Goals:

– Define ClinRO Outcome assessments and types of ClinROs

– Outline emerging good measurement practices to consider when developing or deciding to use ClinRO Outcome Assessments as (part of) endpoint in clinical trial

Little clinical research found specifically related to development, evaluation of ClinROs though commonly used in clinical trials.
**Definitions**

- **Clinician Reported Outcomes**
  - Report that comes from a trained health-care professional after observation of patient’s health condition
  - Involve a *judgment or interpretation* of the observable signs, behaviors, or other physical manifestations thought to be related to a disease or condition
    - Judgment relies on professional training or experience
  - ClinRO measures cannot *directly* assess symptoms that are known only to the patient (e.g., pain intensity)
When is a ClinRO Appropriate?

- When clinicians can form accurate assessments (that reflect feels, functions, survival) that becomes the measurement on which a study endpoint is based

- For example,
  - Where patients are unable to self-report on their own health status due to mental or physical impairment or age (e.g., neonates)
  - Or where patients cannot comment on a specific sign (e.g., Parkinson’s disease)
  - When composite of disease signs used as outcome
Definitions

- **Types of ClinROs:**
  - *Readings* – dichotomous report: fracture on a radiography, absence of presence skin lesions
  - *Ratings* – categorical or continuous scoring report: spleen size, change in skin lesions, severity of schizophrenia (PANSS); improvement in vein appearance
    - In endpoint rating may be dichotomized to “responder/non-responder” but measurement itself is still a rating
  - *Globals* – Clinician Global Impressions (CGI): overall judgment on patient status; what is measured and how assessments made not clearly defined
Definitions

- “Clinicians” can be various types of trained members of an investigative team (anyone who has a role in the study, e.g. blinded raters)
  - Physicians
  - Nurses, nurse practitioners, physicians assistants
  - Clinical research associates
  - Various types of therapists
  - Etc.
Overall Outline: Good Measurement Practices

GMP 1: Define the context of use of the measure
GMP 2: Define concept of interest
GMP 3: Explain relationship of assessment to direct treatment benefit
GMP 4: Documentation of content validity
GMP 5: Documentation of other measurement properties
GMP 6: Interpretation and analysis clinical trial results
GMP 7: Implementation in clinical trials
ClinROs to Construct Endpoints

- Little empirical research on development, evaluation, measurement properties of ClinROs – recommendations based on currently available information
- ClinROs commonly used to construct endpoints or part of endpoint in clinical trials
- ClinRO often part of composite endpoints (e.g. ACR 20, 50)
  - Composite only as good as its weakest link so ClinRO development important
Areas of Emphasis with ClinROs

- Most are indirect measures of patient benefit –
  - Requires defining direct benefit that ClinRO is supposed to reflect (GMP-3)
  - Needs evaluation of relationship between direct benefit (feels, function, future survival) and concept measured (GMP-3)
- Content validity – clinician (and patients) GMP-4
- Measurement properties – inter- and intra-rater reliability (GMP-5)
- Implementation – need for training of persons making assessments (GMP-7)
Examples of ClinRO
Usage and Evaluation
Overall Approach

Context of Use

Treatment benefit (feels, function, survival)

concept(s) of interest

outcome assessment(s)

endpoint(s)
ClinRO Reading

- COU – patients with cystic fibrosis
- Treatment benefit – patient functional disability and symptoms
- COI – pulmonary exacerbation (occurrence or not)
- COA – Clinician judged symptoms based on Fuchs’ criteria (check list of 12 sino-pulmonary signs and symptoms)

Endpoint – comparison of proportions of patients in test and control groups on time to first exacerbation of CF (presence of at least 4 signs/symptoms resulting in administration of medication)
ClinRO Rating

- COU: Acute Bacterial Skin and Skin Structure Infections in adults who can self-report
- Treatment benefit: patient pain and decreased function in daily life
- COI: skin lesion size
- COA: scalar measurement (rating) of skin lesion size with paper ruler at baseline and during study by clinicians (judgment required to perform measurement)

Endpoint: Proportion of patients in test and controls groups with >20% decrease in lesion size by 48-72 hr post-randomization (PRO in development)
ClinRO Clinician Global

- COU – Hospital-acquired pneumonia in adult patients who can self-report
- Treatment benefit – undefined; serious and life threatening disease so all-cause mortality key; symptoms and/or function in survivors not defined as part of ClinRO CG
- COI – undefined
- COA – “clinical response” variables undefined: clinician judgment of improvement in signs, symptoms, biomarkers (body temperature, WBC, CXR) such that no additional antimicrobial therapy is prescribed

Endpoint – difference of proportions between test and control group at various time points called “clinical response”
Evaluation of ClinRO Definition

- Acute bacterial skin and skin structure infections (ABSSSI)
- ClinRO of “lesion area” but how is “lesion defined? How is “area” measured?
- Two different definitions and two different ways of measurement:
  - Erythema only vs. erythema plus induration plus swelling
    - Patient qualitative interviews show all three are important signs to patients.
  - Head to toe measurement vs. longest axis measurement
Example of Application of GMPs for CLinROs

- **GMP 1: Context of Use:**
  - Acute bacterial skin and skin structure infections in adults who can self-report

- **GMP 2: Concept of Interest:**
  - Lesion Area
  - Based on demonstration of treatment effects in antibiotic studies from 1930s on decreasing lesion area to justify use of noninferiority hypotheses
  - Should COI be based on demonstration of treatment effects or defined based on patient benefit first?
Example of Application of GMPs for ClinROs

GMP 3 – Relationship between Concept of Interest and Direct Measures of Patient Benefit
- How does Lesion Area relate to patient-relevant concepts like pain?
- Comparison of pain measured with Visual Analog Scale (VAS) and Faces Rating Scale (FRS) to lesion area at baseline and over time

Results:
- Poor relationship between lesion area and pain at baseline (large lesions can have little pain and vice versa)
- Good relationship between decrease in pain and decrease in lesion area over time regardless of initial lesion area at baseline

Figure 2. Baseline visual analog scale pain scores compared with baseline lesion area, illustrating lack of association between these two variables at baseline.

VAS, visual analog scale.
Example of Application of GMPs for ClinROs

A  Worst pain

<table>
<thead>
<tr>
<th>Improvement</th>
<th>Median Visual Analog Scale (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>n = 1295</td>
</tr>
<tr>
<td>Day 2</td>
<td>n = 1293</td>
</tr>
<tr>
<td>48-72 h</td>
<td>n = 455</td>
</tr>
<tr>
<td>Day 4-6</td>
<td>n = 823</td>
</tr>
<tr>
<td>Day 7-9</td>
<td>n = 1203</td>
</tr>
<tr>
<td>Day &gt;14</td>
<td>n = 39</td>
</tr>
<tr>
<td>Day 10-13</td>
<td>n = 1169</td>
</tr>
</tbody>
</table>

Decrease in Median Lesion Area (%)
Example of Application of GMPs for ClinROs

- GMP 4: Document Content Validity
  - Qualitative interviews with clinicians and patients regarding which signs (and symptoms) of disease important
    - Patients can observe these signs so interviews with patients as well as clinicians
  - Confirmed that signs of disease of erythema (redness), induration (hardness of skin), and edema (swelling) all important

Example of Application of GMPs for ClinROs

- GMP-5: Document other measurement properties
- Goals of study:
  - Evaluate intra-cluster correlation coefficients of inter-rater reliability
    - using various definitions of “lesion area” based on head to toe compared to longest length and width measurements
    - Using various definitions of lesion area based on signs of disease (erythema only vs. erythema, induration, edema [EIE])
The light gray area represents the center of the lesion. The dark gray area represents the erythema. The dotted line represents the area of induration or edema.
Example of Application of GMPs for ClinROs

- ICC for inter-rater reliability of lesion area 0.982 to 0.992 at baseline regardless of definition

- ICC for inter-rater reliability of lesion area 0.960 to 0.989 at primary endpoint timing of 48-72 hours post randomization regardless of lesion area definition
Challenges with Clinician Globals

- Clinician Global ClinROs do not define the variables measured or the criteria for “success” on those variables
- Challenging to know what is measured, whether it relates to direct patient benefit, and how to interpret
- May have poor measurement properties given greater potential for variability
Comparison of ClinRO Rating and ClinRO Global

- Noninferiority RCT in Acute Exacerbations of COPD comparing antibiotic Drug A and Drug B
- Primary endpoint: “Clinical response” defined as clinician judgments that all signs and symptoms cured or improved such that further antibiotics not needed
  - ClinRO global since signs, symptoms and criteria for further therapy not defined
- Case Report Form (CRF) included ClinRO rating of specific variables:
  - Cough, Dyspnea, Sputum Production, Sputum Purulence
Comparison of ClinRO Rating and ClinRO Global

- Allowed for direct comparison of ClinRO Rating in CRF with overall ClinRO global
- CRF included clinician rating of each variable as cured/improved
- Compared with overall global assessment on CRF of “cured/improved”
- Evaluation of misclassification using ClinRO rating as reference standard (defined variables and criteria)
Comparison of ClinRO Rating and ClinRO Global

- Kappa correlation between ClinRO Rating and ClinRO Global = 0.63
- Appears “good” but what is effect on study results
- ~50% of “failures” on ClinRO Rating misclassified as “success” with ClinRO Global
Effect of Misclassification on Trial Results (point estimates & 95% CI)

Lower bound of 95% CI changes from -3.8 to -8.3

Lower bound of 95% CI changes from -3.8 to -8.3

ClinRO Global

ClinRO Global

ClinRO Rating

ClinRO Rating

-3.8%

-3.8%

-2.2%

-2.2%

+1.0%

+1.0%

+5.7%

+5.7%

+3.4% ClinRO Rating

+3.4% ClinRO Rating

-8.3%

-8.3%

-10% -5% 0 +5% +10%

Treatment difference

← Favors Drug B

Favors Drug A →
ClinRO Globals

- Is ClinRO Global more “comprehensive” than what is being measured?
  - But would expect more failures, not less with ClinRO Global if more “comprehensive”

- Is some important aspect of disease measured by ClinRO Global? If so, what? Or does it ignore some aspect of ClinRO rating?

- Or does ClinRO Global rely on biomarkers or other surrogate outcomes that do not relate to direct patient benefit?

- Or are differences between ClinRO Rating and Global due to variability of what is measured from one clinician to another or within clinicians?
Study Results

- Even a small amount of misclassification changes points estimates and confidence intervals substantially.
- Differential misclassification (different between study arms) is possible.
- Not addressed by randomization since misclassification not random and assessments performed after randomization.
- Preference to replace ClinRO globals with clearly defined ClinRO readings or ratings (or PROs) so what is measured, how it is measured, and measurement properties are clear.
Conclusions

- Good measurement practices are important to ClinROs as they are for all COAs
- Development of ClinROs based on GMPs can help ensure that treatment effects are interpretable (direct patient benefits)
- Can increased efficiency of trials if decreased variability and better measurement properties can decreased sample size
- Some points of increased emphasis or differences with ClinROs compared to PROs
  - Relation to direct patient benefits
  - Measurement properties – inter- and intra-rater variability
Please join the review group!

1. Go to the ISPOR homepage:  www.ispor.org
2. Click on the GREEN TASK FORCE menu at the TOP of the homepage
3. Select JOIN on the pull-down menu.
Forum Slides

FORUM SLIDES are AVAILABLE!

http://www.ispor.org/Event/ReleasedPresentations/2016Washington
Questions