Patient-Reported Outcomes (PROs) In Clinical Trials: Challenges & Opportunities

Nancy Kline Leidy PhD
Sr. Research Leader & Sr VP Scientific Affairs
United BioSource Corporation (UBC)
Bethesda, MD

Anti-Infective Drugs Advisory Committee Meeting
December 9, 2009
Gaithersburg, Maryland
Overview

- Introductory Comments
  - Scientific Principles - Efficacy; CABP; Health Outcomes, Properties of Study Endpoints
  - Health Outcomes/Endpoints in CABP
  - Properties of Study Endpoints – PROs
  - Clinical Response in CABP

- PRO Instruments & Development/Regulatory Context
  - Development and validation of a PRO
  - Measuring symptoms of CABP – PROs

- Existing PRO Instruments for CABP
  - Pneumonia Symptom Severity Scales (Metlay et al, 1997; Marrie et al., 2004)
  - The Community-Acquired Pneumonia (CAP) questionnaire (el Moussaoui et al., 2004)
  - The Community-Acquired Pneumonia Symptom Questionnaire (Lamping et al., 2002)
  - Next Steps

- Clinical Response in CABP Trials
  - Clinician-Observered Outcomes
  - Standardization – Key Questions
Scientific Principles

- **Efficacy**
  - Trials – adequate and well-controlled
  - Design – sample; randomization; masking; non-inferiority or superiority; endpoint positioning
  - Outcome measures – well-defined and reliable
    - Reliable, valid, responsive
    - Capture the magnitude of treatment benefit

- **CABP**
  - Characterized by: selected clinical features (e.g., fever, cough, sputum production & pleuritic chest pain) supported by imaging of the lung, usually chest radiography. Physical examination is supportive (Mandell et al, IDSA/ATS Guidelines 2007)
  - Patients – hospitalized/non-hospitalized
    - Higher PORT scores = higher risk of mortality
  - Outcomes – success/clinical failure
    - Mortality (PORT sample enrichment)
    - ICU Admission
    - Duration of hospital stay
    - Clinical Response – signs and symptoms
    - Resolution – of infectious parameters
Health Outcomes

- Results or endpoints of illness – with or without treatment
- Trial endpoints – with Treatment
  » Well-defined and reliable
- Properties of study endpoints
  » Reliability – Precision
  » Validity – Measures what it purports to measure

- Is the endpoint/instrument suitable for a given purpose – e.g., clinical trial?
Scientific Principles

- Properties of Study Endpoints
  - Reliability – Precision
    » All elements of a given measure correspond/correlate with one another
    » Scores are stable over time in stable patients
    » Scores are reproducible across raters/observers
  - Validity – Measures what it purports to measure
    » Content Validity – Qualitative
      - How well the instrument measures the target concept
        - Contains the relevant & important aspects of the concept
        - “What” drives “How”
      - Evaluation – Based on the process used to develop and select items
        - Confidence in the rigor of the development methodology
    » Construct Validity – Quantitative
      - How well scores on the instrument measure (quantify) what is intended
      - Relationship to other outcome measures – similar and dissimilar
        - Known-groups; convergent, discriminant
    » Responsiveness
      - Sensitivity to change
Health Outcomes/Endpoints in CAPB

- **Mortality**
  - Sensitivity issue - Small numbers require large samples
  - Validity issue when used alone - Does not assess efficacy outcomes of survivors

- **Hospitalization**
  - ICU, duration of hospital stay, re-admission rates
  - “Noise” – health policy, hospital policy, clinician practice

- **Microbiological response**
  - Pathogen eradication
  - “Noise” – Inability to expectorate; no organism identified
  - Validity issue – correspondence to other clinical indicators of resolution

- **Chest radiograph response**
  - Sensitivity issue – timing
  - Validity issue – correspondence to other clinical indicators of resolution

- **Clinical response**
  - Time to clinical stability
    - Vital signs, O2 saturation, IV requirements, mental state
  - Resolution of signs and symptoms
    - Combination of observed and patient-reported attributes of CAPB
  - Reliability and Validity - ???
Clinical Response in CAPB Trials

- **Signs and Symptoms**
  - Sign – objectively observed
    - Detected by a clinician during a physical examination
  - Symptom – function or feeling experienced by the patient and reported to the clinician

- **Clinician Observed: Signs of pneumonia**
  - Fever, increased respiratory rate, increased pulse
  - Low oxygen saturation, cyanosis
  - Decreased breath sounds, bronchial breath sounds, crackles/rales in the upright seated position, egophony
    - Rarely: vocal fremitus, friction rub, whispered pectoriloquy
  - Dulled percussion over affected lung
  - Variable inter-observer agreement (Metlay et al., 1997; Wipf et al., 1999)

- **Patient Reported: Symptoms of pneumonia**
  - Cough, dyspnea, sputum production, pleuritic chest pain, fatigue (Metlay et al., 1997; Marrie et al., 2004)
Clinical Response in CABP Trials

- Clinical Response: Patient-reported symptoms and clinician-observed signs of CABP

**Key questions:**
- How is “clinical response” standardized for endpoint measurement?
- How are patient-reported symptoms and clinician-observed signs standardized and quantified to determine “clinical response” to treatment in randomized, controlled trials of CABP treatment in a regulatory context?
- How *should* clinical response be standardized for endpoint measurement in multinational trials?
Overview

- **Introductory Comments**
  - Scientific Principles - Efficacy; CABP; Health Outcomes, Properties of Study Endpoints
  - Health Outcomes/Endpoints in CABP
  - Properties of Study Endpoints – PROs
  - Clinical Response in CABP

- **PRO Instruments & Development/Regulatory Context**
  - Development and validation of a PRO
  - Measuring symptoms of CABP – PROs

- **Existing PRO Instruments for CABP**
  - Pneumonia Symptom Severity Scales (Metlay et al, 1997; Marrie et al., 2004)
  - The Community-Acquired Pneumonia (CAP) questionnaire (el Moussaoui et al., 2004)
  - The Community-Acquired Pneumonia Symptom Questionnaire (Lamping et al., 2002)
  - Next Steps

- **Clinical Response in CABP Trials**
  - Clinician-Observed Outcomes
  - Standardization – Key Questions
- Properties of Study Endpoints - **PROs**
  - Reliability – Precision
    » All elements of a given measure correspond/correlate with one another
    » Scores are stable over time in stable patients
    » Scores are reproducible across raters/observers
  - Validity – Measures what it purports to measure
    » Content Validity – Qualitative
      - How well the instrument measures the *target concept*
        - Contains the relevant & important aspects of the concept
        - “What” drives “How”
      - Evaluation - Based on the process used to develop and select items
        - Confidence in the rigor of the development methodology
    » Construct Validity – Quantitative
      - How well scores on the instrument measure (quantify) what is intended
      - Relationship to other outcome measures – similar and dissimilar
        - Known-groups; convergent, discriminant
  - Responsiveness
    - Sensitivity to change

(Patrick et al., Value in Health 2007, 10: S125-S137)
The Development & Validation Process

i. Hypothesize Conceptual Framework
   - Outline hypothesized concepts & potential claims
   - Determine the intended population
   - Determine intended application/characteristics
   - Perform literature/expert review
   - Develop hypothesized conceptual framework
   - Position in preliminary endpoint model
   - Document preliminary instrument development

ii. Adjust Conceptual Framework & Draft Instrument
   - Obtain patient input
   - Generate new items
   - Select recall period, response options & format
   - Select mode/method of administration
   - Conduct patient cognitive interviewing
   - Pilot test draft instrument
   - Document content validity

iii. Confirm Conceptual Framework & Assess Other Measurement Properties
   - Confirm conceptual framework with scoring rule
   - Assess reliability, validity, sensitivity
   - Finalize instrument content, format, scoring & training
   - Document measurement development

iv. Collect, Analyze, & Interpret Data
   - Prepare protocol & statistical analysis plan
   - Collect & analyze data
   - Evaluate treatment response
   - Document interpretation of treatment benefit in relation to claim

(L. Burke, Measuring Study Endpoints in Clinical Trials, DIA, New Orleans LA 2009)
The Development & Validation Process: Modified Wheel and Spokes (Simplified)

i. Hypothesize Conceptual Framework

ii. Adjust Conceptual Framework & Draft Instrument

iii. Confirm Conceptual Framework & Assess Other Measurement Properties

iv. Collect, Analyze, & Interpret Data

v. Modify Instrument

(L. Burke, Measuring Study Endpoints in Clinical Trials, DIA, New Orleans LA 2009)
Process and Sample Timelines

- **Development**  4 – 6 months
  - Literature review
  - Focus groups & interviews
    - Rate limiting factor – site selection, IRB, recruitment
  - Item pool development
  - Cognitive debriefing
    - Rate limiting factor – site selection, IRB, recruitment
  - Consultation with experts

- **Validation**  6 – 18 months
  - Protocol design
  - Study execution
    - Rate limiting factor – season, site selection, IRB, recruitment
  - Development of the statistical analysis plan
  - Analyses – item reduction and validation
  - Consultation with experts

- **Use in clinical trials**  Ongoing
  - Exploratory or secondary endpoint
  - With experience, use as a secondary or primary endpoint
Measuring Symptoms of CABP - PROs

- Are there existing CABP Symptom PRO Instruments?
  - Yes

- Can these CABP Symptom PRO Instruments be used in clinical trials evaluating the safety and efficacy of anti-infective agents?
  - Do they follow current standards for endpoint development and validation?
    - Are the properties consistent with PRO Guidance recommendations?
      - Content validity, reliability, construct validity, sensitivity to change?
  - Are they suitable for clinical trials in a regulatory context?

- What are the options?
  - Examine existing instruments for consistency with standards
    - If consistent, use the instrument
  - Adapt an existing instrument
    - Make adjustments and validate the modified instrument
  - Develop a new measure
    - Using current standards and guidance documents
CAPB Symptom PRO Instruments

- **Pneumonia Symptom Severity Scales**
  - Symptom Severity Score – Metlay et al., 1997
  - PSS - Marrie et al., J of Infection, 2004

- **Community-Acquired Pneumonia (CAP) questionnaire**
  - el Moussaoui et al., Thorax, 2004; el Moussaoui et al., Chest, 2006

- **Community-Acquired Pneumonia Symptom questionnaire (CAP-Sym)**
  - Lampling et al., Chest, 2002; Torres et al., ERJ, 2003
Properties of CABP Symptom Measures

- **Reliability – Precision**
  - All elements of a given measure correspond/correlate
  - Scores are stable over time in stable patients
  - Scores are reproducible across raters/observers

- **Validity – Measures what it purports to measure**
  - **Content Validity**
    - The extent to which an instrument contains the relevant & important aspects of the concept it intends to measure.
    - The items represent a sufficient sampling of content to represent the concept
    - Evaluation – Based on the process used to develop and select items
      - Confidence in the rigor of the methodology
  - **Construct Validity**
    - How well the instrument measures what is intended
      - Scores represent the outcome
    - Relationship to other outcome measures – similar and dissimilar
      - Concurrent, Convergent, Divergent, Discriminant
  - **Responsiveness**
    - Sensitivity to change
Content Validity: Content Consensus through Qualitative Research

Concept Elicitation (Focus Groups & Interviews)

Generated Words & Phrases

Instrument Evaluation (Cognitive Interviews)

Consensus Wording
Items & Response Options
Structure
Recall, Instructions Format

Interpretation & Meaning

Developer Expertise

Value in Health – Figure 2
Symptom Assessment in CABP – Content Validity

**Literature**
- Cough
- Sputum production (color)
- Dyspnea
- Pleuritic chest pain
- Fatigue
- Tired
- Myalgia/muscle pain
- Headache
- Chills
- Shaking
- Excessive sweating
- Clammy skin
- Nausea
- Vomiting

**Pneumonia**
- **Respiratory**
  - Cough with sputum or phlegm
  - Shortness of breath
  - Pleuritic chest pain
  - Hemoptysis
- **Systemic**
  - Headaches
  - Loss of appetite
  - Mood swings
  - Nausea
  - Vomiting

# Symptom Assessment in CABP – Content Validity

**Literature**
- Cough
- Sputum production (color)
- Dyspnea
- Pleuritic chest pain
- Headache
- Chills
- Shaking
- Excessive sweating
- Clammy skin
- Myalgia/muscle pain
- Fatigue
- Tired
- Weak
- Nauea
- Vomiting

**Symptom Severity***
- **PSS***
  - Cough
  - Dyspnea
  - Sputum
  - Pleuritic chest pain
  - Fatigue

**CAP Questionnaire***
- Cough
- Sputum production
- Sputum color
- Sputum with ease
- Shortness of breath
- Severity of shortness of breath
- Feeling fit
- General health

**CAP-SYM 12***
- Coughing
- Shortness of breath
- Chest pains
- Headache
- Chills
- Sweating
- Muscle pain
- Fatigue
- Nausea
- Lack of appetite
- Trouble concentrating
- Trouble sleeping

**CAP-SYM 18***
- Coughing up phlegm
- Vomiting
- Coughing up blood
- Diarrhea
- Stomach pain
- Trouble thinking

---

* Marrie et al., 2004  
** Metlay et al., 1997  
* el Moussaoui et al., 2004  
*Lamping et al., 2002
## Symptom Assessment in CABP – Content Validity

### Literature
- Cough
- Sputum production (color)
- Dyspnea
- Pleuritic chest pain
- Headache
- Chills
- Shaking
- Excessive sweating
- Clammy skin
- Myalgia/muscle pain
- Fatigue
- Tired
- Weak
- Nausea
- Vomiting

### PSS*

<table>
<thead>
<tr>
<th>Symptom Severity**</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Cough</td>
</tr>
<tr>
<td>✓ Dyspnea</td>
</tr>
<tr>
<td>✓ Sputum</td>
</tr>
<tr>
<td>✓ Pleuritic chest pain</td>
</tr>
<tr>
<td>✓ Fatigue</td>
</tr>
</tbody>
</table>

### CAP Questionnaire*

- Cough
- Sputum production
- Sputum color
- Sputum with ease
- Shortness of breath
- Severity of shortness of breath
- Feeling fit
- General health

### CAP-SYM 12*

- Coughing
- Shortness of breath
- Chest pains
- Headache
- Chills
- Sweating
- Muscle pain
- Fatigue
- Nausea
- Lack of appetite
- Trouble concentrating
- Trouble sleeping

### CAP-SYM 18*

- Coughing up phlegm
- Vomiting
- Coughing up blood
- Diarrhea
- Stomach pain
- Trouble thinking

*Marrie et al., 2004
** Metlay et al, 1997
* el Moussaoui et al., 2004
* Lamping et al., 2002
Pneumonia Symptom Score (PSS) (2004)

- Development (content validity)
  - No development history

- Structure
  - 5 Items (fatigue, cough, dyspnea, sputum, pleuritic chest pain)
  - 6-point scale from 0 (no symptom) to 5 (very severe symptom)
  - Patient self-assessment at set intervals, e.g., Days 0 to 14, 30 and 42

- Scoring
  - Sum; range: 0 to 25  (Transformed score 0 to 100)
  - Symptom resolution: total symptom score ≤ 20 at day 14 (untransformed score)
    - "indicates very mild individual symptoms, ≤ 1 per symptom"
  - Individual symptom resolution: ≤ 1 at day 14

- Context: clinical trial report for resolution of symptoms
  - No data on reliability or validity

(Marrie et al., Journal of Infection 2004; 49: 302-309)
N=399 CAP
- Gender: 52% male
- Mean Age: 48.6 (±15.8) years

Outpatient
- Inclusion: Signs and symptoms consistent with mild to moderate bacterial pneumonia not requiring hospitalization, radiologic evidence of new or progressive infiltrate; 2 or more of the following findings: productive cough, purulent sputum, dyspnoea or tachypnea (>20 rr), rigors or chills, pleuritic chest pain

RCT
- Efficacy and safety of 2 treatments over 10 days

(Marrie et al., Journal of Infection 2004; 49: 302-309)
PSS Change Over Time

Total Pneumonia Symptom Score Sum (PSS)

(Marrie et al., Journal of Infection 2004; 49: 302-309)
PSS Change by Resolution Status at Day 14

Total Pneumonia Symptom Score Sum (PSS)

(Marrie et al., Journal of Infection 2004; 49: 302-309)
The Development & Validation Process: Modified Wheel and Spokes (Simplified)

i. Hypothesize Conceptual Framework

ii. Adjust Conceptual Framework & Draft Instrument

iii. Confirm Conceptual Framework & Assess Other Measurement Properties

iv. Collect, Analyze, & Interpret Data

v. Modify Instrument

PSS
Symptom Severity Score (1997)

- **Development (content validity)**
  - Panel of investigators; based on prevalent symptoms
  - Response option scaling – based on Anthonisen et al, 1087

- **Structure**
  - 5 Items (fatigue, cough, dyspnea, sputum, pleuritic chest pain)
  - 2 to 5 point scales; all transformed to 6-point scales (0=none to 5=severe)
  - Mixed mode – interview, mail – at set intervals, e.g., Days 0, 7, 30 90

- **Scoring**
  - 6-point scale scores summed and transformed to a 0 to 100 summary score
  - Hypothesized meaningful change: 20 points
    - One symptom change from very severe to absent or all symptoms improving by a single severity point

(Metlay et al., J Gen Intern Med 1997; 12: 423-430)
Symptom Severity Score – Performance Properties

- **N=576 CAP**
  - Gender: 38% male
  - Age: 78% <60 years

- **Outpatient**
  - Inclusion: Acute onset of ≥1 of 18 clinical symptoms suggestive of acute illness; radiologic evidence of acute pneumonia within 24 hours of presentation

- **Multicenter prospective cohort study**
  - Pneumonia Patient Outcomes Research Team (Pneumonia PORT)

- **Mode:**
  - Mixed interviewer, in-person self, mail survey
  - Days 0 to 7, 30, 90; retrospective recall for pre-pneumonia baseline

(Metlay et al., J Gen Intern Med 1997; 12: 423-430)
Symptom Severity Score Properties (N=576)

- **Reliability**
  - Internal Consistency
  - Test-Re-test
    - Cronbach’s Alpha: 0.50 (Day 0; 0.70, Day 30 and 90)
    - Not reported

- **Validity**
  - Content Validity
    - Literature, experts – no patient input
  - Construct Validity
    - Predictive
      - Elevated scores at Day 7 or 30 predicted clinic visit
  - Responsiveness
    - Sensitive to change over time
      - Improvement consistent with health status (SF-36)
      - Stronger effect size

(Metlay et al., J Gen Intern Med 1997; 12: 423-430)
### Symptom Severity Score – Change Over Time

(Ref: Metlay et al., J Gen Intern Med 1997; 12: 423-430)

#### Table 2A. Proportion Reporting Symptoms During Resolution of Pneumonia*

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Preparation</th>
<th>Day 0</th>
<th>Day 7</th>
<th>Day 30</th>
<th>Day 90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue</td>
<td>29</td>
<td>93</td>
<td>80</td>
<td>65</td>
<td>51</td>
</tr>
<tr>
<td>Cough</td>
<td>16</td>
<td>90</td>
<td>82</td>
<td>53</td>
<td>32</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>16</td>
<td>68</td>
<td>50</td>
<td>36</td>
<td>28</td>
</tr>
<tr>
<td>Sputum</td>
<td>10</td>
<td>63</td>
<td>59</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>Pleuritic chest pain</td>
<td>3</td>
<td>47</td>
<td>22</td>
<td>12</td>
<td>8</td>
</tr>
</tbody>
</table>

#### Table 2B. Proportion Reporting Moderate to Severe Symptoms During Resolution of Pneumonia*

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Preparation</th>
<th>Day 0</th>
<th>Day 7</th>
<th>Day 30</th>
<th>Day 90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue</td>
<td>10</td>
<td>79</td>
<td>48</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>Cough</td>
<td>7</td>
<td>80</td>
<td>51</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>2</td>
<td>41</td>
<td>15</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Sputum</td>
<td>3</td>
<td>39</td>
<td>23</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Pleuritic chest pain</td>
<td>1</td>
<td>38</td>
<td>11</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

* *n = 576; patients with missing values represented <1% for each time point and were eliminated from those calculations.

Day 0: 51.7 (± 20.1)  
Day 7: 31.2 (± 18.0)  
Day 30: 19.4 ± 16.9  
Day 90: 13.6 (± 16.4)
The Development & Validation Process: Modified Wheel and Spokes (Simplified)

1. Hypothesize Conceptual Framework
2. Adjust Conceptual Framework & Draft Instrument
3. Confirm Conceptual Framework & Assess Other Measurement Properties
4. Collect, Analyze, & Interpret Data
5. Modify Instrument

Symptom Severity
## Symptom Assessment in CABP – Content Validity

### Literature
- Cough
- Sputum production (color)
- Dyspnea
- Pleuritic chest pain
- Headache
- Chills
- Shaking
- Excessive sweating
- Clammy skin
- Myalgia/muscle pain
- Fatigue
- Tired
- Weak
- Nausea
- Vomiting

### Symptom Severity*  
**PSS**  
**Symptom Severity**
- Cough
- Dyspnea
- Sputum
- Pleuritic chest pain
- Fatigue

### CAP Questionnaire*
- Cough
- Sputum production
- Sputum color
- Sputum with ease
- Shortness of breath
- Severity of shortness of breath
- Feeling fit
- General health

### CAP-SYM 12*
- Coughing
- Shortness of breath
- Chest pains
- Headache
- Chills
- Sweating
- Muscle pain
- Fatigue
- Nausea
- Lack of appetite
- Trouble concentrating
- Trouble sleeping

### CAP-SYM 18*
- Coughing up phlegm
- Vomiting
- Coughing up blood
- Diarrhea
- Stomach pain
- Trouble thinking

* Marrie et al., 2004  
** Metlay et al., 1997  
* el Moussaoui et al., 2004  
* Lamping et al., 2002
CAP Questionnaire

- Development (Content Validity)
  - Textbooks, literature, experts
  - “The most specific symptoms that characterise the respiratory condition in CAP”

- Structure
  - 9 items
  - Scaling:
    » Dyspnea – yes/no
    » Fatigue and fitness – VAS
    » Others – Likert-type scale (ordinal scaling)
  - Scoring
    » Total score; respiratory score; well-being score

(el Moussaoui et al., Thorax 2004; 59:591-595)
N=67 CAP
  - Gender: 67% male
  - Mean Age: 56 (17.8) years  Range 21-96
  - PSI 56 (23.4) – Range: 20-106

4 of 8 study hospitals
  - Inclusion criteria: temp >38; clinical signs of pneumonia, new infiltrate on chest radiograph, PSI < 110

RCT
  - Comparing 2 durations of treatment of CAP
CAP Questionnaire Performance Properties (N=67)

- **Reliability**
  - Internal Consistency
  - Test-Re-test

  Cronbach’s Alpha: 0.87
  ICC: 0.83

- **Validity**
  - Content Validity
    - Literature, experts – no patient input
  - Construct Validity
    - Within Scale Analyses
    - Clinical
      - Alpha=0.87
      - Corr with physician judgment (r=0.35), temp (r=-0.43);
        respiratory rate (r=-0.34), O2 sat (r= 0.23)
      - WBC (r=-0.25); CRP (r=- 0.31); ESR (r=-0.17)

- Responsiveness
  - Change from Normal to Baseline
  - Baseline to day 10; baseline to day 28 (ES > 1)

(el Moussaoui et al., Thorax 2004; 59:591-595)
CAP Questionnaire Change Over Time

(el Moussaoui R et al., Chest, 2006; 130: 1165-1172)
The Development & Validation Process: Modified Wheel and Spokes (Simplified)

i. Hypothesize Conceptual Framework

ii. Adjust Conceptual Framework & Draft Instrument

iii. Confirm Conceptual Framework & Assess Other Measurement Properties

iv. Collect, Analyze, & Interpret Data

v. Modify Instrument

PRO
Claim

CAP-Questionnaire
## Symptom Assessment in CABP – Content Validity

### Literature
- Cough
- Sputum production (color)
- Dyspnea
- Pleuritic chest pain
- Headache
- Chills
- Shaking
- Excessive sweating
- Clammy skin
- Myalgia/muscle pain
- Fatigue
- Tired
- Weak
- Nausea
- Vomiting

### Symptom Severity*

<table>
<thead>
<tr>
<th>PSS* Symptom Severity**</th>
<th>CAP Questionnaire*</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Cough</td>
<td>✓ Cough</td>
</tr>
<tr>
<td>✓ Dyspnea</td>
<td>✓ Sputum production</td>
</tr>
<tr>
<td>✓ Sputum</td>
<td>✓ Sputum color</td>
</tr>
<tr>
<td>✓ Pleuritic chest pain</td>
<td>✓ Sputum with ease</td>
</tr>
<tr>
<td>✓ Fatigue</td>
<td>✓ Severity of shortness of breath</td>
</tr>
<tr>
<td>✓ Feeling fit</td>
<td>✓ Feeling fit</td>
</tr>
<tr>
<td>✓ General health</td>
<td>✓ General health</td>
</tr>
</tbody>
</table>

### CAP Questionnaire*
- ✓ Cough
- ✓ Sputum production
- ✓ Sputum color
- ✓ Sputum with ease
- ✓ Shortness of breath
- ✓ Severity of shortness of breath
- ✓ Feeling fit
- ✓ General health

### CAP-SYM 12*
- ✓ Coughing
- ✓ Shortness of breath
- ✓ Chest pains
- ✓ Headache
- ✓ Chills
- ✓ Sweating
- ✓ Muscle pain
- ✓ Fatigue
- ✓ Nausea
- ✓ Lack of appetite
- ✓ Trouble concentrating
- ✓ Trouble sleeping

### CAP-SYM 18*
- ✓ Coughing up phlegm
- ✓ Vomiting
- ✓ Coughing up blood
- ✓ Diarrhea
- ✓ Stomach pain
- ✓ Trouble thinking

* Marrie et al., 2004  
** Metlay et al., 1997  
* el Moussaoui et al., 2004  
* Lamping et al., 2002
**CAP-Sym Development – Content Validity**

- **Qualitative Interviews**
  - Telephone or face-to-face
  - Daily life with CAP
  - Symptoms
  - Circumstances most bothered/limited due to CAP (pre-defined format)

- **N=33 with CAP**
  - US & France
  - Different stages of CAP (0 – 7 days; 8-21 days; > 28 days and end of oral treatment
  - Mean age: 52 Years
  - Gender: 58% men
  - Treatment: Oral antibiotics; n=8 additional IV treatment

- **Translations**
  - 12 languages using forward/backward methodology

*(Lamping et al., Chest 2002; 122: 920-929)*
## CAP-Sym Structure

### Item Content
In the past 24 hours, how much have you been bothered by….

- Coughing
- Shortness of breath
- Chest pains
- Headache
- Chills
- Sweating
- Muscle pain
- Fatigue
- Lack of appetite
- Nausea
- Trouble concentrating
- Trouble sleeping
- Coughing up phlegm
- Coughing up blood
- Vomiting
- Diarrhea
- Stomach pain
- Trouble thinking

### Response Options

| 0. | Did not have |
| 1. | Not at all  |
| 2. | A little    |
| 3. | Moderately  |
| 4. | Quite a Bit |
| 5. | Extremely   |

### Scoring

- **Summation**
- **0 to 90**
- **Higher Scores = Poorer Outcome**

---

(Lamping et al., Chest 2002; 122: 920-929)
N=556 CAP
- Gender: 58% male
- Mean Age: 50.41 (18.65) years  Range: 17-97

Outpatient clinics, general practice, hospital centers
- Inclusion criteria: Fever, elevated WBC, signor or symptoms of pneumonia, and a new or progressive infiltrate on chest radiograph

64 Centers; 13 Countries

RCT
- Moxifloxacin (400 mg QD) vs Standard Treatment
- Standard Treatment:
  » Amoxicillin, 1g tid, and/or
  » Clarithromycin, 500 mg bid
- Treatment up to 14 days

(Lamping et al., Chest 2002; 122: 920-929)
CAP-Sym Performance Properties (N=556)

- **Reliability**
  - Internal Consistency
  - Test-Re-test
  
  - Cronbach’s Alpha: 0.82
  - ICC: 0.96

- **Validity**
  - Content Validity
  - Qualitative Research

  - Construct Validity
    - Within Scale Analyses: Alpha=0.82; Inter-item Correlations; EFA
    - Known-Group Differences: Scores Cure > Failure (n=7) (p=0.034)
    - Convergent Validity: Corr with SF-36 Vitality = 0.33; PCS -0.35; MSC-0.25
    - Discriminant Validity: No corr with Age, Gender

- **Responsiveness**
  - Change from Baseline to days 3-5; 7-10; 28-35 ES> 1.0

(Lamping et al., Chest 2002; 122: 920-929)
CAP-Sym Responsiveness in the RCT

(Torres et al., Eur Respir J 2003; 21: 135-143)
CAP-Sym Change Over Time (RCT)

Clinical Cure: 12

Clinical Failure: 21

(Torres et al., Eur Respir J 2003; 21: 135-143)
The Development & Validation Process: Modified Wheel and Spokes (Simplified)

i. Hypothesize Conceptual Framework

ii. Adjust Conceptual Framework & Draft Instrument

iii. Confirm Conceptual Framework & Assess Other Measurement Properties

iv. Collect, Analyze, & Interpret Data

v. Modify Instrument

PRO \[\downarrow\] Claim

CAP-Sym
The Development & Validation Process (CAP-Sym)

i. Hypothesize Conceptual Framework
   - Outline hypothesized concepts & potential claims
   - Determine the intended population
   - Determine the intended application/characteristics (type of scores, mode, frequency of administration)
   - Perform literature/expert review
   - Develop hypothesized conceptual framework
   - Position in preliminary endpoint model
   - Document preliminary instrument development

iv. Collect, Analyze, & Interpret Data
   - Document interpretation of treatment benefit in relation to claim

v. Modify Instrument
   - Change wording of items, populations, response options, recall period, or mode/method of administration/data collection
   - Translate & culturally adapt to other languages
   - Evaluate modifications as appropriate
   - Document all changes
Measuring Symptoms of CABP

- Are there existing CABP Symptom PRO Instruments?
  - Yes – Examples: PSS, CAP Questionnaire, CAP-Sym
  - All responsive to change

- Can these CABP Symptom PRO Instruments be used in clinical trials evaluating the safety and efficacy of anti-infective agents?
  - Do they follow current standards for endpoint development and validation?
  - Are they suitable for clinical trials in a regulatory context?
  - CAP-Sym – Closest to FDA Draft Guidance for PRO measures
    » Qualitative empirical foundation; quantitative evaluation
    » Key issues – Content validity relative to target claim (“bothersome” vs “severity” rating); documentation (evaluation limited to the publication); limited information on interpretation

- What are the options?
  - Further examination of the CAP-Sym for consistency with standards
    » Full evidence dossier for detailed assessment and regulatory review
    » If consistent, move forward with the measure
  - Consider adapting the instrument
    » Make adjustments and validate the modified instrument; documentation relative to guidance
  - Develop a new measure
    » Using current standards and guidance documents
Considerations: Population, Claims, Positioning

- **Population – CABP**
  - Hospitalized vs Outpatient
  - Presenting vs enriched (PORT)
  - Diagnostic criteria – Signs & Symptoms + chest radiograph?
    » Symptoms – standardized (CABP Symptom PRO Instrument)

- **Claims**
  - Clinical response – Recovery
    » Time to clinical response
    » Clinical response (success/failure) at Day X
  - Measurement of “Recovery”
    » Symptom resolution
    » Sign and symptom resolution – composite symptoms + sign (e.g., afebrile)

- **Positioning**
  - Primary
  - Secondary
Symptom Outcomes: Next Steps

- **Options**
  - Further examination of the CAP-Sym for consistency with standards
  - Consider adapting the instrument
  - Develop a new measure

- **Next Steps**
  - Determine the population & claim
  - Select from the options
  - Transition to the outcome
    » Exploratory ↔ Secondary ↔ Primary
  - Refine or replicate during the transition

- **Possible Path**
  - Collaboration
  - Collaboration with the FDA through the C-Path Institute PRO Consortium
    » [http://www.c-path.org/PRO.cfm](http://www.c-path.org/PRO.cfm)
Clinical Response in CABP Trials

Key questions:
- How is “clinical response” standardized for endpoint measurement?
- How are patient-reported symptoms and clinician-observed signs standardized and quantified to determine “clinical response” to treatment in randomized, controlled trials of CABP treatment in a regulatory context?
- How should clinical response be standardized for endpoint measurement in multinational trials?

Addressed:
- Patient Reported: Symptoms of pneumonia

To be considered:
- Clinician Observed: Signs of pneumonia
Clinician-Observed: Signs of Pneumonia

- Clinician-reported outcome (ClinRO) – A standardized rating of directly observed aspects of a patient’s health status that require clinical assessment and judgment.
  - Behaviors, signs, or observable symptoms
  - Definition: Measuring Study Endpoints in Clinical Trials, DIA, New Orleans LA 2009
  - “To be meaningful, however, there should be evidence that the [PRO] instrument effectively measures the particular concept that is studied.” (US FDA PRO Draft Guidance, 2006)
Properties of Study Endpoints

- Reliability – Precision
  » All elements of a given measure correspond/correlate with one another
  » Scores are stable over time in stable patients
  » Scores are reproducible across raters/observers

- Validity – Measures what it purports to measure
  » Content Validity – Qualitative
    - How well the instrument measures the target concept
      - Contains the relevant & important aspects of the concept
      - “What” drives “How”
    - Evaluation – Based on the process used to develop and select items
      - Confidence in the rigor of the development methodology
  » Construct Validity – Quantitative
    - How well scores on the instrument measure (quantify) what is intended
    - Relationship to other outcome measures – similar and dissimilar
      - Known-groups; convergent, discriminant

» Responsiveness
  - Sensitivity to change
Reproducibility of Chest Findings

Table 1.—Precision of Physical Examination Findings in Examination of the Chest*

<table>
<thead>
<tr>
<th>Physical Examination Finding</th>
<th>Agreement, %†</th>
<th>K Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tachypnea</td>
<td>63</td>
<td>0.25</td>
</tr>
<tr>
<td>Reduced chest movement</td>
<td>70</td>
<td>0.38</td>
</tr>
<tr>
<td>Increased tactile fremitus</td>
<td>85</td>
<td>0.01</td>
</tr>
<tr>
<td>Dullness to percussion</td>
<td>77</td>
<td>0.52</td>
</tr>
<tr>
<td>Decreased breath sounds</td>
<td>. . ‡</td>
<td>0.43</td>
</tr>
<tr>
<td>Wheezes</td>
<td>79</td>
<td>0.51</td>
</tr>
<tr>
<td>Crackles</td>
<td>72</td>
<td>0.41</td>
</tr>
<tr>
<td>Bronchial breath sounds</td>
<td>. . ‡</td>
<td>0.32</td>
</tr>
<tr>
<td>Whispered pectoriloquy</td>
<td>. . ‡</td>
<td>0.11</td>
</tr>
</tbody>
</table>

*Adapted from Spiteri et al. 23
†Calculated based on data provided in table 1 of Spiteri et al. 23
‡ mean pair agreement rates were not calculated for the signs for which 2 or more physicians in a group failed to report the presence or absence of the sign.

(Spiteri et al., Lancet 1988, in Metlay et al, JAMA 1997; 1440-1445)
Reproducibility of Chest Findings - Pneumonia

K: 0=chance; 1=perfect agreement

Table 3. Physician Agreement on Findings as Reflected by κ Values

<table>
<thead>
<tr>
<th>Finding</th>
<th>Physician A vs Physician B</th>
<th>Physician A vs Physician C</th>
<th>Physician B vs Physician C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bronchi breath sounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0.14</td>
<td>-0.14</td>
<td>-0.14</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.03</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>Bronchophony</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>-0.12</td>
<td>-0.06</td>
<td>-0.14</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.16</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>Egophony</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>-0.08</td>
<td>0</td>
<td>0.18</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.03</td>
<td>-0.10</td>
<td>0.03</td>
</tr>
<tr>
<td>Rales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0.35</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.24</td>
<td>0.49</td>
<td>0.65</td>
</tr>
<tr>
<td>Lateral decubitus rales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0.23</td>
<td>0.47</td>
<td>0.23</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.32</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>Wheezes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0.49</td>
<td>1.0</td>
<td>0.65</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.17</td>
<td>0.65</td>
<td>-0.05</td>
</tr>
<tr>
<td>Rhonchi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0.13</td>
<td>0</td>
<td>-0.06</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.18</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>Percussion (fingertip)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0</td>
<td>=</td>
<td>0</td>
</tr>
<tr>
<td>Right lung</td>
<td>0.40</td>
<td>1.0</td>
<td>0.28</td>
</tr>
<tr>
<td>Percussion (auscultatory)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>0</td>
<td>0</td>
<td>0.45</td>
</tr>
<tr>
<td>Right lung</td>
<td>-0.04</td>
<td>0.28</td>
<td>0.10</td>
</tr>
<tr>
<td>Pneumonia diagnosis (%)</td>
<td>0.18 (60)</td>
<td>0.31 (68)</td>
<td>0.43 (72)</td>
</tr>
</tbody>
</table>

*κ is not defined due to expected agreement of 1.0.

(Wipf et al, Arch Intern Med 1999, 1082-1087)
Clinician-Observed: Signs of Pneumonia
A Road Map for Standardization?

i. Hypothesize Conceptual Framework

ii. Adjust Conceptual Framework & Draft Instrument

iii. Confirm Conceptual Framework & Assess Other Measurement Properties

iv. Collect, Analyze, & Interpret Data

v. Modify Instrument
Standardizing Clinical Response in CABP: Key Questions

- **What is the Population?**
  - Hospitalized vs Outpatient
  - Presenting vs enriched (PORT)
  - Diagnostic criteria – Signs & Symptoms + chest radiograph?
    » Symptoms – standardized (CABP Symptom PRO Instrument)

- **What are the Claims?**
  - Clinical response – Recovery
    » Time to clinical response
    » Clinical response (success/failure) at Day X
  - Measurement of “Recovery”
    » Symptom resolution
    » Sign and symptom resolution – composite symptoms + sign (e.g., afebrile)

- **How are the Outcomes Positioned?**
  - Primary/secondary? In the short and long term?
Overview/Summary

- **Introductory Comments**
  - Scientific Principles - Efficacy; CABP; Health Outcomes, Properties of Study Endpoints
  - Health Outcomes/Endpoints in CABP
  - Properties of Study Endpoints – PROs
  - Clinical Response in CABP

- **PRO Instruments & Development/Regulatory Context**
  - Development and validation of a PRO
  - Measuring symptoms of CABP – PROs

- **Existing PRO Instruments for CABP**
  - Pneumonia Symptom Severity Scales (Metlay et al, 1997; Marrie et al., 2004)
  - The Community-Acquired Pneumonia (CAP) questionnaire (el Moussaoui et al., 2004)
  - The Community-Acquired Pneumonia Symptom Questionnaire (Lamping et al., 2002)
  - Next Steps

- **Clinical Response in CABP Trials**
  - Clinician-Observed Outcomes
  - Standardization – Key Questions
Conclusions

- John Glenn, Friendship 7
- February 20, 1962, Cape Canaveral
- First American to Orbit Earth

- Space Shuttle Endeavour
- June 15, 2002


