Clinical Outcome Assessments (COA) Qualification Program
DDT COA #000018: Pneumonia Patient-Reported Outcome Measure (PNEUMO-PRO)
June 6, 2018 Update

DDT # 0018
June 5, 2018

The CABP instrument has been formally renamed as PNEUMO-PRO© 2017 Foundation for the National Institutes of Health.

Along with the inclusion of the HABP-specific criteria, we have been made modifications based on the feedback we received from sites about the difficulty of enrolling patients:

- A footnote has been added to the **fever criteria**, in order to allow for "self-reported history of fever." Sites have reported that patients may have had a fever before being admitted, but the current criteria does not allow for this to be considered an eligible sign/symptom.
- **Expansion of the enrollment window from 24 hours to 48 hours after treatment initiation**: This decision has been made after multiple conversations with the clinicians and site coordinators participating in the validation studies as well as the expert opinion leaders on the FNIH Project team. The narrow window of 24 hours has made it extremely difficult to recruit patients due to the time window of their availability, especially during weekend hours (i.e. admitted on Fri/Sat). We continue to reiterate to sites that they should be enrolling patients as soon as possible; furthermore, during data analysis, we will have the ability to stratify to compare patients who were enrolled 24 hrs vs. 48 hours post treatment initiation.

The contract provides for a step-wise approach for the psychometric evaluation of the content validated instruments and our Team supports that the change to enrolment criteria will not affect the objective of psychometric validation since we are validating the measures prior to evaluating drug effectiveness. This dual approach seeks to spur standardized and harmonized use of the instruments by sponsor as the efforts move along the qualification process

CABP PRO and Protocol Comments:

1. CABP PRO Items: While we agree that it is acceptable to proceed with the full 29-item CABP PRO diary in your psychometric evaluation study at this time, note that you will need to submit a copy of your psychometric study report, along with a full data set (including longitudinal item response data from each participant, baseline information, and other measures completed by participants and clinicians) to facilitate qualification review. We remind you that we are most interested in core disease-related symptoms and impacts, as outlined in the CABP guidance. Careful attention to content validity and other measurement properties of the instrument and adherence to the agreed upon concept(s) and context of use for the instrument will be needed for successful qualification effort. Retention of items that do not closely adhere to the concept(s) could present a content validity issue.

   Response: We will submit a copy of our full psychometric evaluation study report to facilitate the qualification review.

2. Not Applicable Response Option: You have chosen to retain the "Not Applicable" response option for items 24-29 (see FDA letter of August 17, 2017) and have requested clarification regarding our recommendation.
    a. To clarify, we continue to recommend the removal of the "Not Applicable" options from items 24-29. We do not believe that "Not Applicable" is a meaningful response to these items.
        i. **Response: We have complied with your recommendation and removed the "Not Applicable" option.**
    b. The Rating Scale model and classical test theory scoring you will use in your psychometric analysis assume that items have ordinal response scales. The CABP PRO's first 23 items have a 5-point ordinal response scale (from "Not at all" to "Very Much") but items 24-29 do not have an ordinal response scale due to the inclusion of the "Not Applicable" option. Please explain how your IRT and classical test theory scoring will accommodate item responses of "Not Applicable."

       Response: The "Not Applicable" response options have been removed.

    c. Please provide us with a rationale, supported by the qualitative interview data (patient understanding of this option within the context of each question), to justify the inclusion of the "Not Applicable" option for items 24-29. We continue to have concerns that patients might have difficulty interpreting the meaning of the "Not Applicable" response option for items 24-29.

       Response: The "Not Applicable" response options have been removed.

3. **Item Interpretation for Inpatients vs. Outpatients:** We continue to have concerns regarding the interpretation of items 24 (difficulty sleeping), 25 (difficulty doing usual activities), and 27 (social activities) among inpatients and outpatients (see FDA letter of August 17, 2017). Specifically, we are concerned that inpatients' responses to these items could be limited by hospital protocol/ environment (e.g., level of independence and sleep schedules) whereas outpatients don't have these same constraints. Please explain how your scoring algorithm will account for these differences between the two subpopulations.

   Response: This is precisely why we initially wanted to include the N/A response options for specific items which cannot be responded to appropriately by inpatients. For our planned analyses, we should first explore if the items do indeed perform differently in inpatients vs. outpatients. We can do this by exploring, in IRT, whether the items show any differential item functioning (DIF). If DIF is found, then an appropriate scoring algorithm could be used to rescore inpatients item scores. For example, items could be rescored to reflect item scores reported by outpatients with similar level of difficulties in other items.

4. **Item Skipping:** We understand that the ePRO device used at US sites will be programmed to require participants to respond to every item.
    a. We are concerned that forcing patients to respond to each of the 29 items will add undue patient burden, increase chances for fatigue, and consequently impact the quality of response received over the 14-day period. That is, valid item responses will be discarded if a patient decides not to complete the entire PRO, and other fatigued patients will sometimes give rote responses to hasten completion. Please explain your plan for empirically evaluating the effect of response requirements on the quality, validity and reliability of the CABP PRO data. If evidence in the

literature is available to justify this approach with other PRO measures, please provide us with copies of these references for review. In addition, we recommend you consider (i) programming the ePRO devices so that the first 7 item responses measuring core disease-related symptoms are uploaded automatically once a patient completes them; this will ensure that these responses are collected for all patients even if they do not complete the entire PRO, and (ii) administering the full CABP PRO every other day, while only administering the first 7 items on the remaining, alternating days.

Response: This was also a reason for us initially wanting to keep the N/A response option. If patients are forced to respond to an item that they really cannot respond appropriately to, then this may influence both the current and subsequent item responses. ePRO devices are typically programmed not to allow item skipping, and the key strength of ePRO is the ability to prevent subjects from progressing to the next item in an instrument until they have provided a response to the current item—it offers the chance of complete PRO data at the close of the study. A well-designed instrument that addresses relevant concepts for patients and is appropriate for the context of use, is likely to result in high compliance rates ensuring that the issue of allowing skips is not one of significant concern. Because we have developed this measure according to the FDA's PRO/DDT Guidance, we believe this instrument is a well-designed one. We have also designed the system so each questionnaire (PNEUMO-PRO, EuroQol, PGI, etc.) is its own record. Once a questionnaire is completed, it is sent to the database either by action of the participant (pressing send) or during the automatic transfer at night. If the participant truly feels burdened after completing the PNEUMO-PRO, they are not forced to fill the remaining questionnaires for that day, and we will still have obtained the PNEUMO-PRO data.

5. **Language Translation and Cultural Adaptation:** Per our February 13, 2018 ICON/FNIH teleconference meeting, we acknowledge that plans for translation and cultural adaptation will be discussed later, after final item reduction is complete for the English version of the CABP PRO diary.

6. **Exclusion criteria:** We recommend that you consider exclusion of patients with post-obstructive pneumonia and known bronchial obstruction (e.g., in the setting of pulmonary malignancy) because the symptoms of the pulmonary disease process are likely to be progressive and may not respond to antibacterial drug therapy. Also, this exclusion criterion is consistent with the FDA guidance on CABP.

Response: We agree and have included this in the inclusion criteria.

SAP Comments:

7. Page 17 of your SAP briefly discusses how endpoints will be constructed from longitudinal CABP PRO data.
    a. The SAP details how the reliability and validity of the CABP PRO will be assessed. However, please be aware that the FDA will be qualifying the PRO instrument in the context of the endpoint(s) it will support. In this case, the endpoint should be in alignment with the FDA's CABP Guidance. Please propose an endpoint(s) based on the CABP PRO.

i. The FDA guidance on CABP (p. 8) notes two possible primary efficacy endpoints: (i) improvement in respiratory symptoms (corresponding to CABP PRO items 1, 2, 4, and 6), and (ii) 28-day mortality. If your intention is that the CABP PRO be used to define a primary efficacy endpoint, then the average of CABP PRO scores over 7 days (discussed in the SAP) will not correspond to the guidance recommendation for a suitable primary endpoint for a noninferiority trial. Likewise, defining a primary efficacy endpoint that includes the assessment of non-respiratory symptoms will also not be aligned with guidance recommendations. Please explain whether you aim to define a primary efficacy endpoint and discuss whether it will conform to FDA guidance recommendations.

Response: We do aim to define a primary efficacy endpoint. This endpoint will conform to FDA guidance recommendations, such as the degree of improvement (e.g., at day 3 to day 5 from baseline) in respiratory symptoms (CABP PRO items 1,2,4,6). In terms of the definition of this endpoint, we will refer to the FDA guidance document where improvement is defined as at least a one-point improvement from baseline on each item.

b. Any PRO assessing symptoms of serious diseases will sometimes have missing data due to patient incapacity or death. For example, CABP inpatients will be unable to complete the CABP PRO on any days they are ventilated. Endpoints should be defined so that they still have meaningful values in the face of PRO data that are missing due to patient incapacity or death. Please explain how you will do this.

Response: We will need to explore in the analysis how to define a primary endpoint in the face of missing PRO data by examining the data obtained. It is possible, for example, that PRO data that are missing due to patient incapacity or death will be assigned the worst possible value

c. If you define an efficacy endpoint in terms of averaging some number of daily CABP PRO scores, then participant status over the days that are averaged should be stable. Otherwise, for some participants, the endpoint could represent the average of poor and good status at different time points, which would be hard to interpret.

Response: While we are indeed proposing to average item scores over a day, it is not necessarily the case that we would average scores over a number of days. As stated here, this may obscure both high and low scores on any one day. Please see our response to a. above as to our approaches to define an efficacy endpoint.

d. As you plan to recruit CABP and HABP patients together to develop a single instrument that can be used by CABP and HABP patients, you will need to submit a letter explaining your new plan to develop your instrument in both populations and a revised SAP reflecting these changes and how the two populations will be analyzed.

Response: We will be submitting the updated protocol and revised SAP.

Statistical Analysis Plan
Psychometric Evaluation of a new PNEUMO-PRO© Instrument

1. Introduction

The Foundation for the National Institutes of Health Biomarkers Consortium (FNIH BC) is interested in developing reliable, well-defined and clinically relevant endpoints that measure tangible benefits for patients in clinical trials of antibacterial drugs. FNIH BC identified Community-Acquired Bacterial Pneumonia (CABP) and Hospital-Acquired Bacterial Pneumonia (HABP) as priority indications, and subsequently developed a candidate list of endpoints for use in clinical trials. As part of this effort, the FNIH BC seeks to develop a patient reported outcome (PRO) symptom instrument in accordance with the Food and Drug Administration (FDA) guidance for PRO measures used to support labeling claims (FDA PRO guidance, 2009) for use in clinical trials of antibacterial interventions. The intention is that the PRO instrument will be used to identify and assess symptoms related to clinically relevant endpoints for CABP and HABP.

In collaboration with the FNIH BC, ICON Patient Centered Outcomes (ICON PCO) has created a new PRO instrument to assess both CABP and HABP symptoms and impacts in a clinical trial setting. Having evaluated the content validity of the instrument in qualitative interviews, the next step is to validate the instrument psychometrically in line with FDA PRO guidance (FDA, 2009). The purpose of this statistical analysis plan (SAP) is to provide full details of the statistical analyses that have been outlined in the study protocol. The scope of this plan includes all the proposed analyses to be executed by ICON PCO. The SAP outlines the rationale for the statistical tests that will be performed and the criteria that will be used to interpret the results. A list of tables summarizing the analyses is also provided. Associated statistical programming code for undertaking these analyses will be prepared by ICON PCO based on the SAP.

2. Objectives

The objective of this study is to evaluate the psychometric properties of the new PNEUMO-PRO©1 instrument, which will be administered as a daily diary. The psychometric properties of the PNEUMO-PRO© will be measured in CABP and HABP patient populations. This is part of a broader effort between ICON PCO and FNIH BC to support an FDA label claim submission used in clinical trials for antibacterial interventions and other studies as appropriate. The psychometric properties the study will assess include:

- Item level properties
- Domain Structure
- Reliability
- Construct validity
- Ability to detect change
- Responder definition

Document updated on 03/18/2019.