# Table of Contents

1. **PURPOSE** .............................................................................................................................................. 1
2. **TERMINOLOGY** ....................................................................................................................................... 1
3. **PROCESS RECOMMENDATIONS** ........................................................................................................... 2
   3.1. **DETERMINING MINIMUM QUALIFICATION FOR STUDY PARTICIPATION** ................................. 2
   3.2. **TRAINING OF NEW RATERS** ........................................................................................................... 2
4. **RECOMMENDATIONS FOR PRIMARY STUDY MEASURES** ................................................................. 3
   4.1. **MINIMUM STANDARDS FOR TRAINING** .......................................................................................... 3
   4.2. **MINIMUM STANDARDS FOR DEMONSTRATING COMPETENCE** .................................................. 3
5. **GUIDELINES FOR OTHER SCALES** ....................................................................................................... 4
   5.1. **MINIMUM STANDARDS FOR TRAINING** .......................................................................................... 4
   5.2. **MINIMUM STANDARDS FOR DEMONSTRATING COMPETENCE** .................................................. 4
6. **CONSIDERATIONS FOR MULTI-NATIONAL STUDIES** ......................................................................... 5
7. **GUIDELINES FOR DOCUMENTATION** ................................................................................................... 5
   7.1. **TRAINING METHODOLOGY** ............................................................................................................ 5
   7.2. **SITE TRAINING RECORDS** ............................................................................................................ 5
   7.3. **STUDY DOCUMENTATION** ............................................................................................................ 5
8. **RECOMMENDATIONS FOR RETRAINING AND RECERTIFICATION** ...................................................... 6
1. Purpose

Individuals with a wide range of skill and training routinely administer rating scales used in clinical neuroscience research studies. Further, the training and certification methodologies used in clinical trials vary in their level of rigor. There is currently no accepted standard for the clinical research industry to follow when selecting and training raters to administer rating scales. Such scales are used as primary and secondary outcome measures contributing to the registration of investigational drugs and/or for empirical studies published in peer-reviewed journal articles. Studies such as these would be better served by an industry-wide guideline.

The purpose of this document is to define terminology and to propose general process recommendations for training, qualification and certification of raters on clinician-rated scales commonly used in neuroscience clinical research. These recommendations are intended to provide a common framework for clinical trial investigators, pharmaceutical companies, contract research organizations, and other entities to follow, providing a standardized baseline approach for selecting, training and evaluating raters. The scope of the training recommended here is intended to serve a single project or protocol. Recommendations considering prior training and certification are included.

2. Terminology

- **certification**: The endorsements of an entity (e.g., business organization, sponsor, academic organization or professional group) attesting to a rater’s ability to properly administer a scale in the context of clinical research.

- **clinical neuroscience**: An area of scientific research primarily served by the disciplines of psychiatry, neurology, pain, and neuropsychology.

- **qualification**: The process of evaluating a rater against a defined set of requirements (i.e., qualification criteria).

- **qualification criteria**: The minimum educational, professional and experiential credentials considered necessary for the competent administration of a scale by a rater.

- **rater**: A person involved with clinical research who administers a scale to a research participant.

- **scale**: An instrument used to measure severity of signs and symptoms of disease or document diagnosis for purposes of clinical research; outcome measure.
testing  A process, involving a set of questions, example scenarios, or the like, used as a means of evaluating the abilities, aptitudes, skills, or performance of an individual or group; examination.

Training  The educational process of establishing accurate, precise measurement of clinical trial endpoints among raters.

3. Process Recommendations

3.1. Determining Minimum Qualification for Study Participation
Prior to study initiation, the key stakeholders should determine the minimum educational, professional and experiential credentials considered necessary for the competent administration of a scale by a rater (e.g., predefine requirements for minimum terms of relevant clinical interaction with patients who have the disorder under study and experience with administration of a rating scale of similar ilk, respectively).

The minimum qualifications for rating each scale should be clearly documented prior to the start of the study. The study sponsor and/or the CRO should consider these minimum qualifications during the site selection process as a reason for including, or excluding, a site or rater.

3.2. Training of New Raters
A research site may have raters who do not meet the minimum qualification for one or more scales. In these cases, we recommend the following:
 Establish an agreement amongst the stakeholders on the training program to be developed for the project and ensure successful completion of the training program by raters on the study before approving the rater.
 Complete in-field training:
  o Documentation of mentoring by the principal investigator and/or a designated, qualified sub-investigator who is certified to rate the scale in the study. A standard for the type of documentation to be required should be established prior to the start of the study.
  o Co-rating with the principal investigator and/or a designated sub-investigator who is certified to rate the scale in the study until such time as the investigator certifies, in writing, that the rater is competent to administer the scale on his or her own. The site should provide documentation of the co-rating process to the study sponsor to ensure proper adherence.

Note: The level of in-field training required for a study often depends on the amount of experience and training the rater has. Such training and mentoring can range widely, and study-specific guidelines should allow for this range of prior experience so as not to prematurely exclude an appropriate rater simply on the basis of a guideline or requirement that is defined too narrowly.
4. Recommendations for Primary Study Measures
This section contains our recommendations for training and testing raters prior to administration of scales that are of primary importance to the stakeholders. Importance may result from the scale being a primary outcome measure and used as the basis for regulatory submission, or from the scale being of significant interest to the study sponsor for any other reason.

4.1. Minimum Standards for Training
Training for raters can include the following for each primary study measure:
  - Didactic review of the purpose of the scale, standardized rules for administration, overview of some or many scale items, and the scoring for applicable items. A comprehensive review should be conducted for raters who are new to the scale being reviewed. This step may be waived for experienced raters who have demonstrated ability (such as through prior certification) depending on the rigor of the training program that is developed.
  - Interview skills assessment:
    - Should include a discussion of interview techniques.
    - May include demonstration of proper scale administration. This may be done in-person, using a recorded interview, or other means using either a patient\(^1\) with, or actor trained to portray, the disorder.

Minimum training is dependent on early agreements about the training program with the relevant stakeholders.

4.2. Minimum Standards for Demonstrating Competence
We recommend that a rater should demonstrate competency to properly administer and score a scale through multiple knowledge and skill demonstrations:
  - Meets or exceeds the minimum qualifications needed for the scale as defined by the study sponsor.
  - Scores one or more sample video interviews in a manner that demonstrates ability to accurately apply the scale rules with a high degree of agreement with colleagues and/or expert consensus\(^2\). Adjustments made for video quality, linguistic and cultural factors or in response to outlier analysis may be appropriate.
  - Established competence with scale administration, based on previous performance, may waive all other training requirements if agreeable to the stakeholders involved. Such competence should be established based on an established standard of testing and training that takes into account appropriate skills:
    - Agreement with consensus or expert panel

\(^1\) If a patient is used, the person being interviewed should provide written consent to the use of their recorded interview for training purposes.
\(^2\) The degree of agreement, and the basis for consensus (e.g., group or expert panel) should be determined \textit{post hoc} and used as a basis for evaluating each rater against the testing objective.
5. Guidelines for Other Scales

This section contains our recommended minimum standards for training and testing raters prior to administration of secondary outcomes, safety and other scales as defined in the protocol for a clinical research study.

5.1. Minimum Standards for Training

Training for raters should include the following for each scale:

- Didactic review of the purpose of the scale, standardized rules for administration, overview of the scale items, and the scoring for each item. As recommended for scales of primary importance, a comprehensive review should be conducted for raters who are new to the scale being reviewed. This step can be waived for experienced raters who have demonstrated ability (such as through prior certification).
- A legally licensed copy of the scale, matching the version identified in the protocol.
- Citation of a relevant, up-to-date publication, as well as the original publication, and a copy of such if considered useful.

5.2. Minimum Standards for Demonstrating Competence

We recommend that a rater should meet or exceed the minimum qualifications needed for the scale as defined by the study sponsor.

- If a scale is measuring an outcome of particular interest to the study sponsor, the sponsor may choose to have the potential raters score one or more sample video interviews in a manner that demonstrates ability to accurately apply the scale rules with a high degree of agreement with colleagues and/or expert consensus. Adjustments made for video quality, linguistic and cultural factors or in response to outlier analysis may be appropriate.
- At the option of the key stakeholders, demonstration of proper scale administration may be required. This may be done in-person or using a recorded interview, using either a patient with, or actor trained to portray, the disorder.

If testing is performed, the methodology and results should be documented in the study report.
6. Considerations for Multi-National Studies

Multi-national studies introduce a number of variables that may affect study outcomes, and should be addressed as part of the training effort. These variables affect scale validity in the culture and include:

- Linguistic differences in the scale version used in the study.
- Cultural and behavioral norms applied by clinicians.
- Clinical training and experience of raters with research trials.

Study sponsors should consider the benefit of using a single test case across all languages versus using multiple, culture/language-specific examples for each culture/language group. A single video example provides a common basis for evaluating test data and necessitates some adjustment for each cultural/language group due to interpretational differences (e.g., translation and subtitling, culturally-adjusted acceptable scores). Conversely, multiple culture/language-specific examples allow for establishment of a higher degree of agreement within the culture/language group; however, they likely preclude any cross-cultural analysis. Study sponsors are encouraged to meet with their study statistician and make their rater evaluation process consistent with the statistical analysis plan of the study.

7. Guidelines for Documentation

Training and certification should be documented for each site. In addition, a comprehensive training report should be prepared at the end of the study and maintained as part of study documentation.

7.1. Training Methodology

The study sponsor or delegate should document the training methodology used for the study prior to the beginning of the study. For each scale, the document should specify the qualifications required, the contents of the training provided, and the testing methods used to determine certification (if required by the sponsor).

7.2. Site Training Records

Each site should receive a training record document that contains:

- Name(s) of rater(s) trained and/or certified
- For each rater, the scale(s) trained and/or certified and date of training or certification
- Sponsor name and protocol number
- Name of trainer or training entity

Site training records should be maintained at the site as part of the regulatory binder. This document should be reviewed by the study monitor on a periodic basis to ensure that ratings are being conducted by training and/or certified raters (as specified by the sponsor, by scale).

The study sponsor should also maintain a copy of each site’s training record as part of their study documentation.
qualification requirements for each scale used in the study, the training provided, and the certification results for all raters who participated in the study. The report should also document inter-rater reliability for scales for which testing data was collected, through statistical analysis such as Cohen’s kappa coefficient, intra-class and/or inter-class correlation coefficients (ICC), or Pearson’s $r$.

8. Recommendations for Retraining and Recertification

We recommend that periodic retraining and/or recertification may be relevant to raters participating in longer-term studies.

Study sponsors should evaluate the burden placed on site personnel by retraining and/or recertification with the need to maintain and document ongoing adherence to scale rules and inter-rater reliability during the study. It is desirable that any requirement for retraining and/or recertification be clearly stated to the sites prior to the beginning of the study.

Retraining, to include scoring conventions and guidelines, may be particularly desirable when a study is of longer duration and the frequency of scale administrations is low.

Study sponsors might wish to consider the implications of recertifying raters during a study. For example, if a rater no longer meets certification criteria, what does this mean for the study data previously collected by the rater? Is the rater allowed to continue in the study?

All retraining and results of recertification activities should be documented by the study sponsor and maintained as part of the study documentation.