

Rater training for pediatric trials: pathology, parents, placebos, pandemics

Joohi Jimenez-Shahed, MD

Associate Professor, Neurology and Neurosurgery

Medical Director, Movement Disorders Neuromodulation & Brain Circuit Therapeutics

Rater training is a well-recognized approach to minimizing inaccuracy and variability of clinical outcomes assessments

- Outcome measures are subjective to varying degrees, and can be influenced by raters' judgments, training history and motivations
- In CNS trials, many neurology and psychiatry outcome measures rely heavily on clinician judgment
- In pediatric trials: must weigh input from multiple sources
- Major challenges:
 - Consistency (inter- and intra-rater reliability)
 - Lack of standardization
 - Rater drift
 - Accuracy (concordance with an expert rating or gold standard)

Approaches to rater training

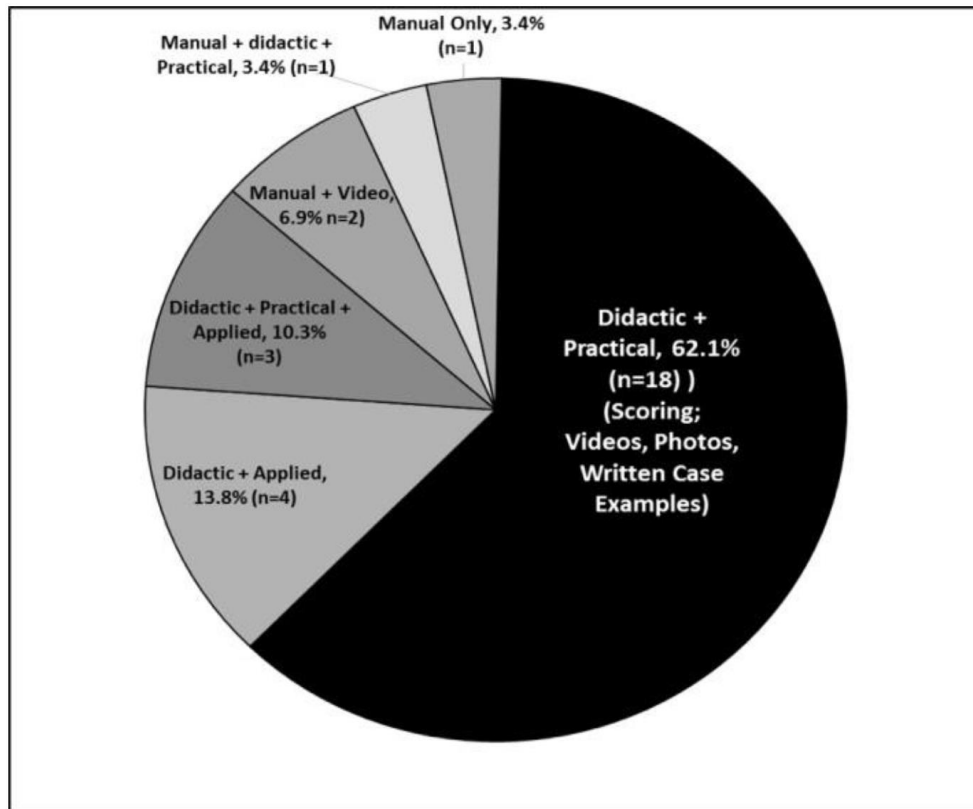


Figure 3: Percentage of studies using each training method.

- **Didactic** = lecture (live or video) on rules for test administration, scoring, w/ or w/o discussion
- **Practical** = scoring interviews/cases (video, audio, photo or written), compared to a gold standard score w/feedback
- **Applied** = conducting/scoring an interview, with live/remote observation and feedback on test administration and interviewing skills.

More comprehensive training programs are more likely to improve data quality

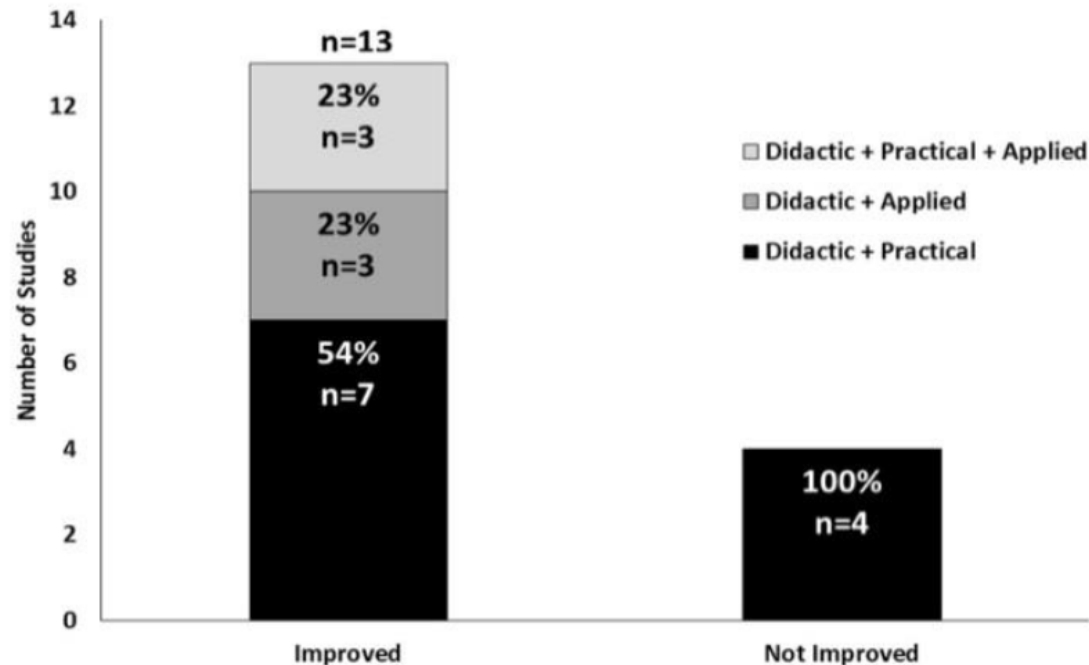


Figure 4: Effects of training methods by percentage of studies showing Improvement or No Difference (only studies in which a comparison was made either pre/post training or training/no training).

Consensus Recommendations on Rater Training and Certification (*CNS Summit Rater training and Certification Committee*)

- Rater qualification
 - Determine minimum qualifications
 - Training new raters
- Training for primary outcome measures
 - Minimum standards for training
 - Minimum standards for demonstrating competence
- Training for secondary/other outcomes
- Considerations for multinational studies
- Maintaining documentation
- Retraining and recertification

“Successful clinical trials require accurate ratings by competent raters to differentiate drug from placebo and document a beneficial drug effect.”

But....

- Many raters who participate in CNS trials have limited clinical experience and little or no training with the selected rating instruments.
- Whereas many instruments
 - require clinical experience to recognize the range of pathology
 - require knowledge about administration and scoring conventions to achieve ratings consistency
- Particular challenges in pediatric CNS trials
 - Children may not have the language/cognitive abilities to fully describe symptoms
 - Raters must weigh input from parents/caregivers and children
 - Children may be particularly susceptible to placebo response as they can be eager to show good cooperation and hence, a positive response

Busner J. CAPN 2013;18(3):p.1-4

Farchione T. CAPN 2013;18(3):p.4-7.

Targum, Journal of Clinical Psychopharmacology 2006:26(3).

Thankfully

- Rater training improves rater competency regardless of previous clinical experience
- The more you train, the better you get...

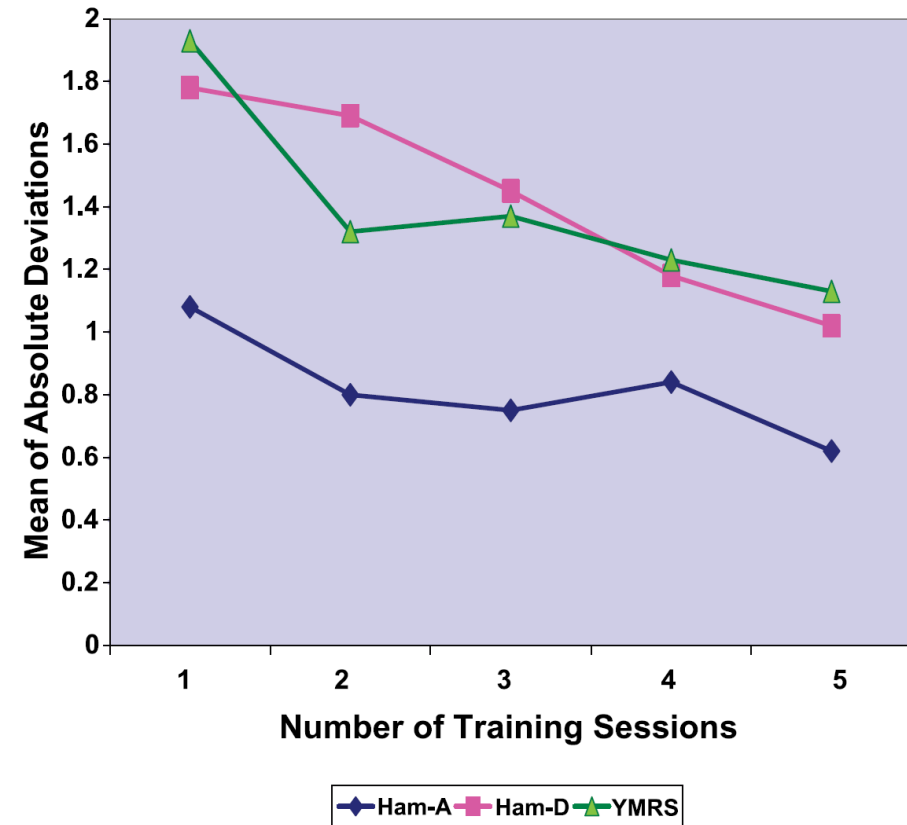
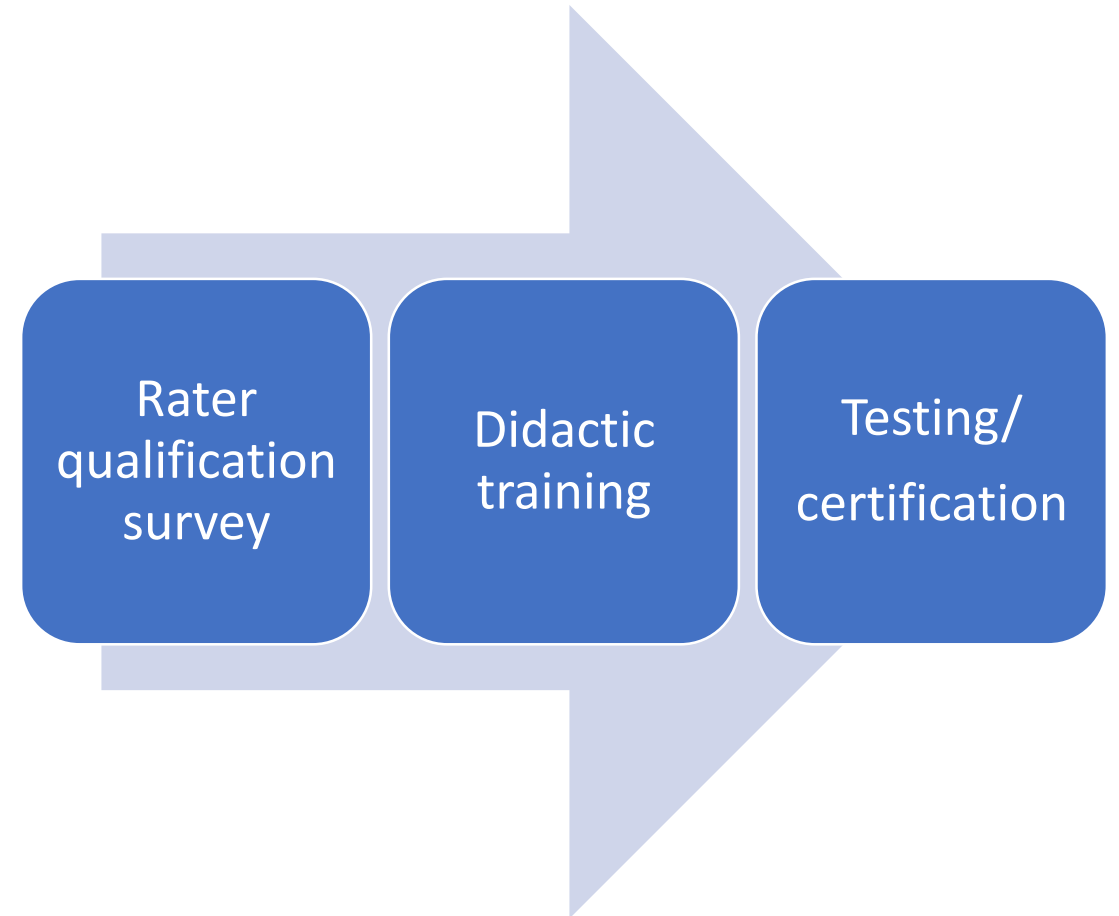


FIGURE 1. Rater training improves rater competency.

Models of rater training that I have participated in

Didactic only	MoCA
Didactic + Practical	UPDRS
Didactic + Applied*	PD psychosis scale
Didactic + Practical + Applied	(none)
Repeat calibration for central reviewer/scorer services	AIMS
Central reviewer/scorer	Tremor rating for a diagnostic imaging marker

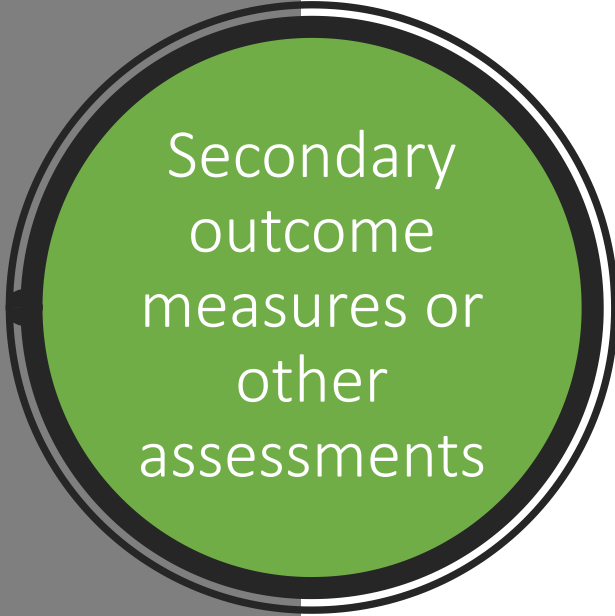
* *unusual in Neurology*



Additional note: process is rarely explained in advance

Primary Outcome Scales in recent pediatric CNS trials vs some adult measures

	CHARACTERISTICS				TRAINING POSSIBILITIES		
	Design	Scoring conventions?	Interview/ exam guide?	Routine clinical use?	Didactic	Practical	Applied
YGTSS	Semi-structured interview	No	No	Yes/No	Y	Y	Y
ICARS	Rate what you see	No	No	No	Y	Y	N
MD-CRS	Rate what you see	No	Yes	No	Y	Y	N
UWDRS	Rate what you see	No	No	No	Y	Y	N
AIMS	Rate what you see	No	Yes	Yes	Y	Y	N
UPDRS	Rate what you see	Yes	Yes	Yes	Y	Y	N



Secondary
outcome
measures or
other
assessments

Secondary outcome measures	Challenges
CGI	Raters may not have sufficient experience with the population
CSSRS	Raters may not typically assess these symptoms in routine practice
Depression	Raters may need specific training on how to administer these measures
Cognitive assessments	Rater calibration

Consideration: Psychiatric assessments in a Neurology trial
 Neurologic assessments in a Psychiatry trial

CNS diseases are complex!

- Empathy
- Hope
- Optimism
- Quality of life
- Belief
- Encouragement

- Emphasize how to minimize the influence of these biases on ratings
 - Need for blinded raters
 - Standardized interview & examination techniques
 - Impact of encouragement on motor performance
- Include discussions about minimizing the placebo response
- Consider how patient-related factors may influence rater's ability to conduct or rate the examination

Site Rater (trainee): additional perspectives

IMs/virtual training – time away from family, work

- Training should be engaging, efficient, and beneficial
- E.g., ARS, videos/multimedia

Learn from an expert who uses the scale

- Multiple scales = multiple trainers – e.g., psychiatry, neurology
- Opportunity for discussion/discourse is often welcome (e.g., “pearls”)

Consider different levels of experience in the audience

- ?groups/tracks

Applied training considerations

- Readily accessible/easy to schedule
- Explain need for feedback procedure and ensure it is constructive

Central Reviewer/Scorer: additional perspectives

- MD-CRS
- AIMS
- UDys-RS

- Importance of scoring conventions and rater calibration, re-calibration
- Appropriate training of other stakeholders
 - E.g., videographers, research assistants
 - Training on equipment
- Video protocol
 - Early feedback on quality, technique, duration

Trainer perspective: Challenges to YGTSS rater training

1. Scale is not typically used in routine clinical practice
 - Semi-structured interview
 - Can take 40-60 minutes to complete the entire scale
 - Most prior training programs: scale didactic + 1hr demo video + “quiz” recording the answers identified in the video
 - Training fatigue

Trainer perspective:

Challenges to YGTSS rater training

2. No “manual” with standard definitions for terminology used in the scale or scoring conventions

- Site raters may have been instructed differently in other trials

YGTSS Dimensions: 4. Complexity

COMPLEXITY

	Motor	Phonic	
NONE If present, all tics are clearly "simple" (sudden, brief, purposeless) in character.	<input type="checkbox"/>	<input type="checkbox"/>	0
BORDERLINE Some tics are not clearly "simple" in character.	<input type="checkbox"/>	<input type="checkbox"/>	1
MILD Some tics are clearly "complex" (purposive in appearance) and mimic brief "automatic" behaviors, such as grooming, syllables, or brief meaningful utterances such as "ah huh," "hi" that could be readily camouflaged.	<input type="checkbox"/>	<input type="checkbox"/>	2
MODERATE Some tics are more "complex" (more purposive and sustained in appearance) and may occur in orchestrated bouts that would be difficult to camouflage but could be rationalized or "explained" as normal behavior or speech (picking, tapping, saying "you bet" or "honey", brief echolalia).	<input type="checkbox"/>	<input type="checkbox"/>	3
MARKED Some tics are very "complex" in character and tend to occur in sustained orchestrated bouts that would be difficult to camouflage and could not be easily rationalized as normal behavior or speech because of their duration and/or their unusual, inappropriate, bizarre or obscene character (a lengthy facial contortion, touching genitals, echolalia, speech atypicalities, longer bouts of saying "what do you mean" repeatedly, or saying "fu" or "sh").	<input type="checkbox"/>	<input type="checkbox"/>	4
SEVERE Some tics involve lengthy bouts of orchestrated behavior or speech that would be impossible to camouflage or successfully rationalize as normal because of their duration and/or extremely unusual, inappropriate, bizarre or obscene character (lengthy displays or utterances often involving copropraxia, self-abusive behavior, or coprolalia).	<input type="checkbox"/>	<input type="checkbox"/>	5

“Bout” = a period of intense tic activity

- Succession of multiple tics when one movement or noise may morph into a subsequent one

“Camouflage” = attempt to perform the tic in a more socially acceptable way, or mask the tic

Factors to consider

- | | |
|--|---|
| <ul style="list-style-type: none">• Simple vs complex• Ability to camouflage• Length of tic displays | <ul style="list-style-type: none">• Inappropriate/bizarre/obscene• Coprolalia/copropraxia/SIBs |
|--|---|

Trainer perspective: Challenges to YGTSS rater training

3. Patients/parents/caregivers may be new to diagnosis and symptomatology
 - Complex neuropsychiatric disorder
 - Need to balance patient's report with parent/caregiver's observations
 - Symptom rating focuses on prior week only
 - Consider tics alone (exclude influence from features related to co-morbidities)



Trainer perspective: Challenges to YGTSS rater training

4. Experienced sites with an appropriate population may have less experienced raters (e.g., sub-Investigators or back-up raters)

5. International clinical trials program
 - YGTSS is not translated to all languages from involved countries
 - ?Cultural influences on or local conventions about language surrounding tic symptoms or reporting
 - Unable to conduct training in native language of each involved country



Pandemic considerations



- CNS disorders lend themselves to remote assessment / telehealth visits for obtaining primary outcome and some safety measures
 - Exception: labs, imaging, EKGs
- Conducting study visits remotely
 - Videography/telemedicine – video quality?
 - Patients/caregivers may need to be trained on equipment as well as scales
- Use of site study-specific tablets or other direct EDC (e.g., web portal) to document ratings
 - PROs + Clinician-measured assessments
 - Associated training
- Virtual rater training or a virtual IM for a “virtual” or “remote” study
 - Video-based, Interactive, Short sessions
 - Periodic refresher training
 - Investigate rater consistency across different/changing clinical scenarios

RECOMMENDATIONS AND FUTURE DIRECTIONS: A MODEL FOR RATER TRAINING AND MONITORING IN [pediatric] CLINICAL TRIALS

A standard set of scoring conventions

- Competing scoring conventions – raters having to use different guidelines in different simultaneous studies

Use of a structured or semistructured interview guide

- Standardized initial probes + improvised follow-up questions as necessary

Didactic training

- Testing didactic knowledge: assess the trainee's conceptual understanding of scoring conventions and general conventions for assessing patients

Applied training

- Must include the trainee administering the rating scale with observation and feedback
- Testing of applied skills with actual or standardized patients

RECOMMENDATIONS AND FUTURE DIRECTIONS: A MODEL FOR RATER TRAINING AND MONITORING IN [pediatric] CLINICAL TRIALS

Testing the efficacy of the training intervention

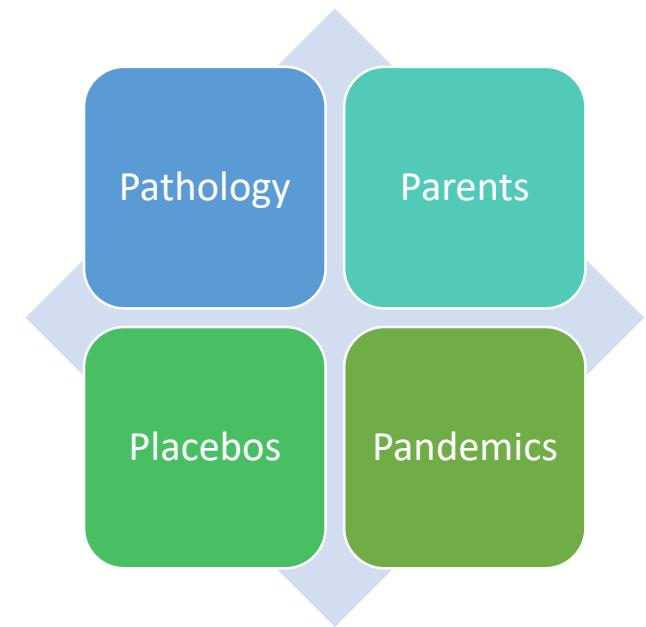
- E.g., pre- and post-testing

Post-training monitoring for interview quality and rater drift

- Monitoring and calibration
- “Refresher” training

Final considerations

- **Know the training audience**
 - Multi-disciplinary and expert training team
 - Avoid the “long single training video”
 - Enriched training for new raters
 - Discuss measures to reduce bias and placebo effect
- **Education/training of additional stakeholders**
 - Patients and parents/caregivers
 - Research assistants, videographers
 - Include disease state, placebo effect, equipment and/or video training
- **Be prescriptive, proactive, and transparent about the training program**
 - Develop a scale-specific training program (didactic, practical, applied)
 - Include plan for refresher and recalibration sessions
 - Consider frequency and nature of refresher training



Investing in rater training:

“While there would be increased costs associated with proper rater training, the costs of failed trials are significantly higher”