

THE IMPACT OF SYSTEMATIC RATER TRAINING ON INTER-RATER RELIABILITY

Rachel Delahay, Shun Yan Toto To, Rajaei K Sharma, Caroline Zangani, Katharine Smith
and with thanks to the NIHR-funded PETRUSHKA study team



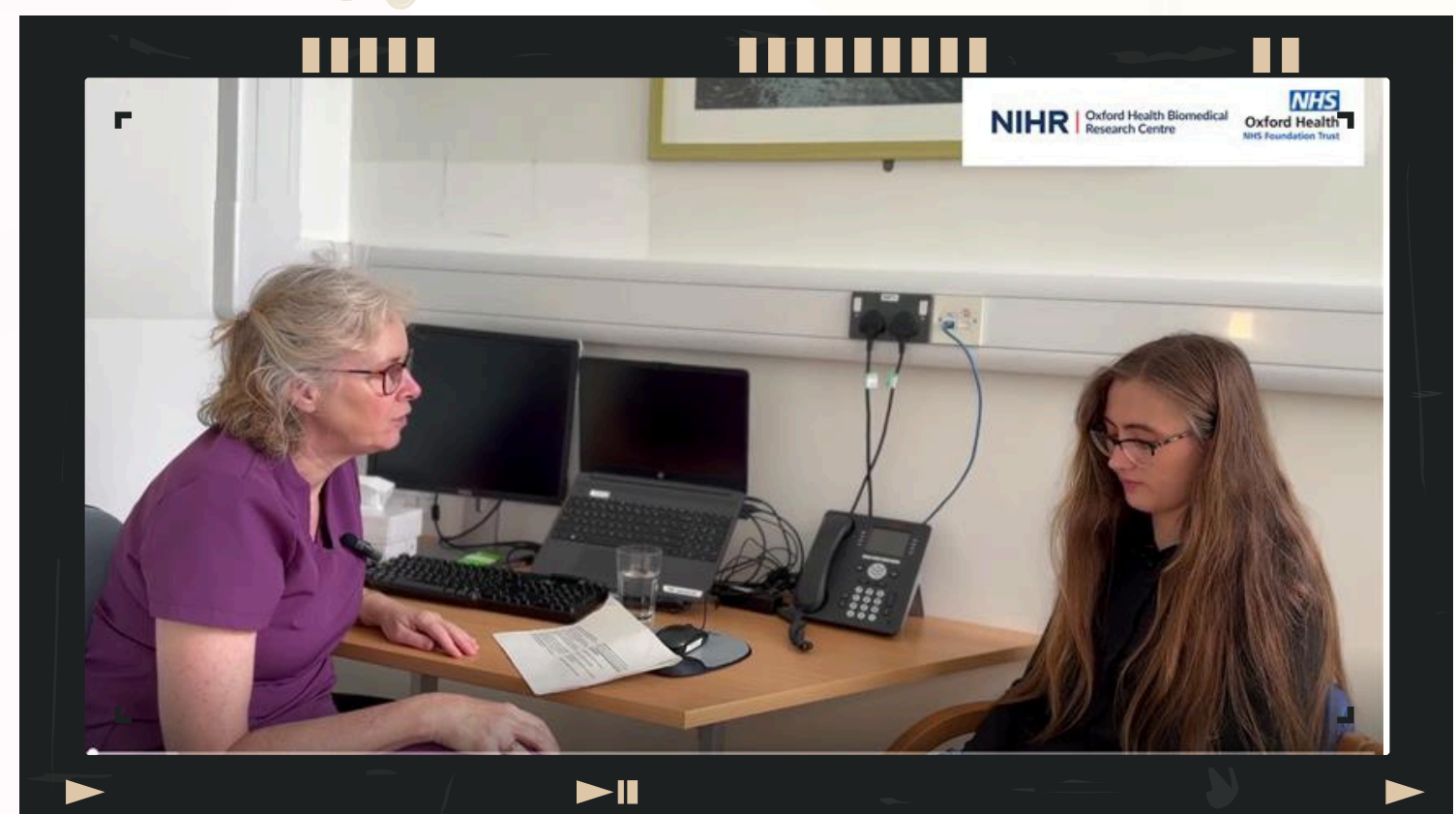
BACKGROUND

- The NIHR Oxford Health Clinical Research Facility (OHC RF) is the only NIHR-funded mental health (MH) CRF in the UK
- Specialized nature of MH rating scales administered → Training is often limited to raters with previous experience
- Good inter-rater reliability (IRR) is essential in eligibility criteria and trial outcomes
- The NIHR-funded PETRUSHKA study (Personalising antidepressant Treatment for Unipolar depression combining individual choices, risks, and big data) (Tomlinson et al., 2020; Ostinelli et al., 2025) → Involved raters from a range of backgrounds with limited experience of MH scales
- OHC RF therefore developed and piloted a systematic rater training programme

METHODS

- Hamilton-Depression Rating Scale (HAM-D) and Hamilton-Anxiety rating scale (HAM-A)
- Simulated interview videos - created with an expert rater
- Consultant psychiatrist-led training + bespoke video
- Trainees submitted independent scores of the simulated interview
- IRR was calculated with intraclass correlation coefficient (ICC) using Microsoft Excel
- 2 independent consultant psychiatrists provided 'gold-standard' ratings for comparison

Training Video



Feedback from attendees:

"I felt comfortable to participate and ask questions."

"The trainer was very knowledgeable and the training was thorough."

"Training provided us with example questions to ask participants for all questions."

RESULTS

- 8 raters submitted the scores of both scales
- 1 rater was excluded - omitted items
- ICC: **0.75**; IRR: **0.74**
- More variability were shown in some sub-scores compared to gold-standard scores
- ✓ Top-up training

CONCLUSION

- The IRR obtained across both scales demonstrated how the combination of one training session and bespoke videos, the focused top-up training with expert input facilitated efficient rater training for different backgrounds (Cicchetti, 1994)
- OHC RF now has expanded the rater training by offering a series of training activities for the sites across Mental Health Mission network

References:

Cicchetti DV (1994). "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology". *Psychological Assessment*. 6 (4): 284–290. doi:10.1037/1040-3590.6.4.284

Ostinelli, E. G., Jaquierey, M., Liu, Q., Elgarf, R., Haque, N., Potts, J., ... & PETRUSHKA Team. (2025). Personalising Antidepressant Treatment for Unipolar Depression Combining Individual Choices, Risks and big Data: The PETRUSHKA Tool: Personnalisation du traitement antidépresseur de la dépression unipolaire associant choix individuels, risques et mégadonnées: l'outil PETRUSHKA. *The Canadian Journal of Psychiatry*, 07067437251322399.

Tomlinson, A., Furukawa, T. A., Efthimiou, O., Salanti, G., De Crescenzo, F., Singh, I., & Cipriani, A. (2020). Personalise antidepressant treatment for unipolar depression combining individual choices, risks and big data (PETRUSHKA): rationale and protocol. *BMJ Ment Health*, 23(2), 52-56.

NIHR | Oxford Health Clinical Research Facility

NHS
Oxford Health
NHS Foundation Trust

