Performance During "Applied" Training on the MADRS/SIGMA as a Predictor of Change in a Global Depression Trial

Rothman B[1], Lord-Bessen J[1], Sanchez R[2], Peters-Strickland T[2], Opler M[1,3]

[1]ProPhase LLC; [2]Otsuka Pharmaceutical Commercialization and Development; [3]New York University

ISCTM 13th Annual Scientific Meeting, 21 – 23 February, 2017, Washington, D.C.

The Question

Which aspects of applied training performance were most associated with change in MADRS?

Introduction

Many rater training programs rely on passive scoring tasks, such as video interview scoring, to assess inter-rater reliability. It is widely accepted, however, that monitoring and review of a rater's "applied" performance (e.g., via a simulated interview) is a more accurate and ecologically valid way to judge rater quality and thus his or her ability to detect change over time. When accompanied by standardized tools such as the Rater Applied Performance Scale (RAPS), the evaluator can assess a rater's capacity to administer and score an assessment in a way that better approximates in-study behavior. Using applied performance, multiple domains of performance are evaluated in addition to scoring – including adherence to an interview guide (if used), follow-up and clarification technique, research rapport and neutrality in questioning (avoidance of leading questions). However, it is unclear as to which aspects of rater performance during applied training are most associated with change in the assessment under consideration. Using data from a multi-site, global depression trial, we sought to identify which RAPS domains were most correlated to change in the MADRS, the study's primary outcome.

Methods

Eighty qualified raters participating in a global depression trial completed a 1-1 "applied" training session via telephone as part of pre-study certification on the MADRS. Each interview used the SIGMA and was scored according to six RAPS domains (Lipsitz, et al., 2004) – Adherence, Follow-up, Clarification, Neutrality, Rapport, and Accuracy of Scoring. Prior to applied training sessions, applied trainers participated in a calibration process consisting of: a) an inter-rater reliability scoring exercise on the MADRS; b) a practice Applied Training session with a standardized actor; c) a practice feedback session whereby the trainer's capacity to provide accurate and constructive feedback was evaluated by a master trainer. Results from applied training were compared to in-study data from 195 subjects, assigned to one of two active comparator arms (escitalopram + placebo or aripiprazole + placebo) or to a combination therapy arm (aripiprazole + escitalopram). Note that there were no significant differences found between study arms in this trial.

For this analysis, we looked at subjects who had the same raters between Weeks 8 and 14 (double-blind randomization phase, where subjects were assigned to one of three treatment arms) and determined the amount of absolute change in the MADRS for that time period; this resulted in a subset of 195 subjects and 78 raters. Then using the Pearson correlation we looked at the relationship between MADRS change and each RAPS domain and the total score of the RAPS.

Results

Table 1 shows the correlation between the change in MADRS between weeks 8 and 14 and each of the RAPS domain as well as the total RAPS score. Of the six RAPS domains, both Follow-Up (r=-0.17, p=0.031) and Accuracy (r=-0.29, p<0.001) were negatively correlated with absolute value of change in MADRS. For all domains but one, the higher the score on the domain, the less change in MADRS from Week 8 to 14. The only domain with a positive relationship with change was Rapport, but it was not significant (r=0.06, p=.489).

Table 2 breaks down the findings in Follow-Up and Scoring domains by treatment arm. The combination therapy (aripiprazole + escitalopram) showed the largest relationship between the RAPS Scoring domain and MADRS change registered (r=-0.44, p=0.001). Both the active comparator arms showed larger correlations between MADRS Change and RAPS Scoring than the combination therapy arm.

Table 1

		Correlation with Change in MADRS	
		r	*p*
RAPS Domains	Adherence	-0.14	0.086
	Follow-Up	-0.17	0.031
	Clarification	-0.11	0.158
	Neutrality	-0.08	0.340
	Rapport	0.06	0.489
	Scoring	-0.29	<0.001
	RAPS Total Score	-0.15	0.056

Table 2

	Correlation with Change in MADRS		
	r	*p*	*n*
Escitalopram + Placebo			67
Follow-Up	-0.26	0.056	
Scoring	-0.34	0.016	
Aripiprazole + Placebo			64
Follow-Up	-0.21	0.147	
Scoring	-0.03	0.838	
Aripiprazole + Escitalopram			64
Follow-Up	-0.08	0.554	
Scoring	-0.44	0.001	

Conclusions

In this sample, Follow-Up and Accuracy were the only RAPS domains significantly correlated with change in MADRS scores**.** This correlation was negative – in other words, higher scores on these domains were associated with less change in the total MADRS from Week 8 to Week 14. This suggests that raters who perform well in these domains during applied training may be less likely to contribute unwanted variability in study outcomes. Ultimately this provides some support for the idea that high performance during training is related to less noise in-study, which will impact a rater's ability to detect signal. Training and in-study surveillance programs involving the MADRS should focus on these domains when evaluating rater quality. Further research should also attempt to corroborate these findings with RAPS validation studies as well as with a larger sample size, to determine whether they can be replicated when separation from drug and placebo is significant.

References

Lipsitz, J, Kobak K, Feiger A, Sikich D, Moroz G, & Engelhardt N. (2004). The Rater Applied Performance Scale: development and reliability. *Psychiatry Research, 127*, 147-155.

Disclosure: This project was funded by Otsuka Pharmaceutical Commercialization and Development and by ProPhase, LLC. The authors report no conflicts of interest for this work.

ProPhase
Precision. Innovation. Impact.