

Site vs. Centralized Raters in a Clinical Depression Trial

Kenneth A. Kobak¹, Ph.D., David DeBrotta², M.D., Nina Engelhardt¹, Ph.D., Janet B.W. Williams^{3,4}, D.S.W.

¹MedAvante, Inc., ²Eli Lilly & Co., ³Columbia University, ⁴NY State Psychiatric Institute,

BACKGROUND. The use of triple-blinding procedures, where raters are blind to study visit and design, has been suggested as a way to improve rating fidelity. Several studies have shown that using different raters at baseline and endpoint improves signal detection. The use of centralized raters, who are linked to study sites and interview patients via two-way videoconferencing, has been proposed as a means to achieve such blinding. The current study compared site and centralized ratings in a two-center depression trial.

METHOD. Patients were interviewed twice at each of three time points: screening, baseline, and endpoint, once by the site rater, and once remotely via videoconference by a centralized rater, who was blind to study visit and design. Raters were blind to each others' scores. A counter-balanced order was used at baseline and endpoint. A site HAMD score of 17 or greater was required for study entry at both screening and baseline visits.

RESULTS. Site HAMD scores were significantly higher than centralized raters scores at screening and baseline, but not at endpoint. Correlations (ICC) between site and centralized ratings at screen and baseline were .33 and .40, but improved to .75 at endpoint. Internal consistency reliabilities (coefficient alpha) for

centralized ratings were .71, .79, and .84 at screen, baseline and endpoint, and .31, .39, and .83 for site raters respectively. Forty-three percent of patients found eligible at screen by site raters were found ineligible by centralized raters, and 57% of patients found eligible at baseline were found ineligible by centralized raters.

previous findings comparing site and self-report ratings.

Table 1. Mean HAMD Score by Visit

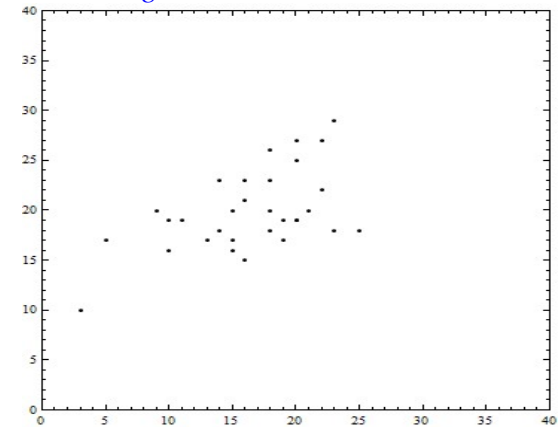
MEAN HAMD SCORE BY VISIT			
	Screen (n=32)	Baseline (n=27)	Endpoint (n=22)
Central Raters	16.28	14.26	11.68
Site Raters	19.88	18.85	11.78
DIFF	3.59	4.59	0.05
P value	.0001	.0001	.968

Table 2. Internal Consistency Reliability

Internal Consistency Reliability			
	Screen	Base-Line	End point
Cent, Raters	.71	.79	.84
Site Raters	.31	.39	.83

CONCLUSION. The use of blinded raters would result in significantly different study populations. Blinded raters generally score depression severity lower at screen and baseline. Site and centralized ratings coalesced at endpoint. Results support

Screening: Site vs Central HAMD



Endpoint: Site vs Central HAMD

