The Challenge of Subject Selection in Clinical Trials: New Data

Detke M [1,2,3], Williams JBW [1,4], Kobak K [1,5], Ellis A [1], Giller E [1], Brown B [1], De Santi S [1,6], Reines S [1], Kane J [7],

MedAvante, Inc [1], Indiana University School of Medicine [2], McLean Hospital & Harvard Medical School [3], College of Physicians and Surgeons, Columbia University [4], Center for Psychological Consultation [5], New York University Langone Medical Center [6], Zucker Hillside Hospital [7]

ABSTRACT

Introduction: Clinical trials fail too frequently (up to 50% failures in trials powered at 80-90%). Signal detection might be enhanced with more reliable scales, greater rater reliability, or the use of independent assessments; here we focus on the last of these. Previous studies showed that 1/3 to 1/2 of the subjects enrolled by site raters in two MDD trials would be excluded by remote blinded clinicians' ratings of initial severity. New data on the extent and characteristics of subject ascertainment discrepancies and one method to mediate it will be presented.

Methods: Inter-rater reliability and internal consistency reliability were assessed in a MDD study. Two doses of an experimental compound were compared to placebo in a GAD study in which remote blinded clinicians and site raters assessed subjects on the HAM-A. In studies (of MDD, GAD & SZ) subjects were assessed by both site raters and by remote blinded clinicians. In two of these studies, accuracy of diagnosis was examined.

Results: In the MDD study, internal consistency reliability (Cronbach's alpha) was strong for remote blinded clinicians at baseline (.67) but much lower for site raters at baseline (.38). In studies of MDD, GAD & SZ, 39% (range: 26-56%) of subjects included by site raters would have been excluded based on remote blinded clinicians' ratings of initial severity. In one completed study of GAD, subject ascertainment by remote blinded clinicians increased the drug effect size of the site ratings from .43 to .74. In other studies, diagnostic assessments by remote blinded clinicians using the SCID-CT also revealed potential diagnostic errors in subjects previously screened for study entry by site-based raters.

Conclusion: Subject ascertainment issues are pervasive and substantial; on symptom severity alone over 1/3 of subjects enrolled in clinical trials may not meet protocol-specified inclusion/exclusion criteria. Diagnosis is an additional source of potential error. Remote blinded clinicians may be beneficial for diagnosis and symptom severity assessment across several diagnoses. Accurate subject ascertainment may substantially increase effect size.

INTRODUCTION

Literature from CNS clinical trials shows a substantial proportion of FDA-approved drugs fail to separate from placebo more frequently than their powering predicts (Khan, 2005). The failure rates for the investigational arms include:

Antidepressants: 52%
Anxiolytics: 52%
Antipsychotics: 25%

Three possible reasons for failed trials include:

- The inclusion of inappropriate subjects due to enrollment pressure, therapeutic alliance, timeline pressures, subject bias for inclusion
- Variability in administration of primary outcome measure across sites and raters resulting in poor inter-rater reliability (ICC)
- Expectation of improvement that results in scoring bias

This poster will focus on subject ascertainment as a potential reason for trial failure through reduction of drug-placebo separation and effect size. It presents recently gathered data and reviews some recent literature in mood, anxiety and psychotic disorders using blinded Remote Centralized Ratings and Continuous Quality Control (CQC).

METHODS

1. Site and Remote Centralized Ratings (Kobak, et al., 2010):

- 81 subjects with MDD participated in this study of placebo response. They were interviewed at three time points using both Site and Remote Centralized Ratings with the SIGH-D.
- At baseline site ratings were done first. Interview order was counterbalanced at other time points.
- Remote centralized ratings were conducted via videoconferencing and site ratings were conducted face to face.
- We present the distributions of the baseline HAM-D scores separately for Site and Remote Centralized Ratings for all subjects, including screen failures.
- Internal consistency reliability was assessed using Cronbach's alpha.

2. Remote Centralized Ratings in Schizophrenia:

- Baseline scores by centralized raters of 800 subjects with Schizophrenia from three different clinical trials are presented, including screen failures.
- Primary outcome measure: Positive and Negative Syndrome Scale (PANSS).
- The baseline distributions of primary outcome measures by Remote Centralized Raters were analyzed for normality to evaluate the possibility of bias in enrollment decisions.

3. Continuous Quality Control in MDD of Site-Administered Baseline Interviews (Brown, et al., 2010):

- Continuous Quality Control (CQC), is a new approach to monitoring and remediating the administration and scoring of clinical outcome measures. Data from ongoing clinical trials are presented and evaluated for baseline distribution normality, as well as for potential score inflation.
- 17 calibrated quality reviewers were rigorously trained and continuously calibrated on scale scoring and interview quality.
- Site raters audio recorded all scale administrations and uploaded the recordings to a central server.
- Blinded reviewers scored the tapes for interview quality and scoring accuracy. Feedback was provided to the site raters.

4. Site and Remote Centralized Ratings in GAD (Coric, et al., 2008):

- 290 subjects with GAD were interviewed by both Site and Remote Centralized Raters with the HAM-A at 4 visits: Screening, Baseline, Weeks 4 & 8 [or early termination]; site raters went first and determined enrollment.
- Remote centralized ratings were conducted via videoconferencing and site ratings were conducted face to face.
- We present the HAM-A change scores from baseline to endpoint for placebo and active comparator arms, segmented by cohort of subjects enrolled by the sites and cohort that central raters agreed met entry criteria.

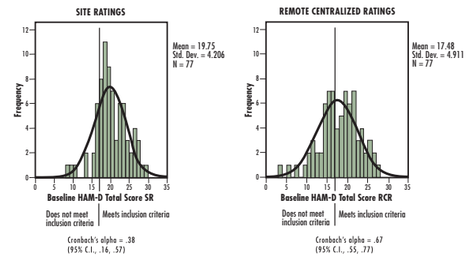
5. Subject Ascertainment – Diagnosis:

- In two large ongoing studies, subjects referred by site personnel were diagnosed by Remote Centralized Raters using the SCID.
- In one of the studies, 3rd Party Board-Certified Psychiatrists (compensated by the sponsor) observed each interview by 3-way videoconference.

RESULTS

1. Distribution of Baseline Scores by Site and Remote Centralized Raters

- In an MDD study, 35% (23/66) of the subjects included based on site ratings would have been excluded based on blinded Remote Centralized Ratings.
- Site ratings had lower Cronbach's alpha (internal consistency reliability) than Remote Centralized Ratings. This is consistent with potential score inflation of site ratings.



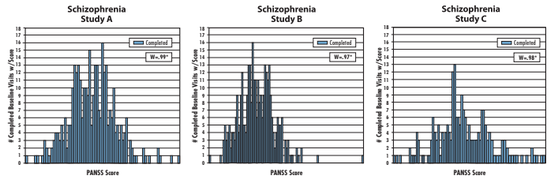
Percentage of Study Subjects That Would Have Been Excluded by Remote Blinded Centralized Raters Based on Symptom Severity Scores
(Rating only those subjects included by Site raters)

- Across several studies, an average of 39% of subjects who would have been included in studies based on site ratings of initial severity would have been excluded by blinded Remote Centralized Ratings of initial severity.

Indication	% Subjects who would be Excluded
GAD	**Trial 1: 43%** **Trial 2: 52%**
MDD	**Trial 1: 26%** **Trial 2: 35%**
Psychosis	**Trial 1: 32%** **Trial 2: 56%**
	Range 26% - 56% **Weighted mean = 39%** **(811/2054)**

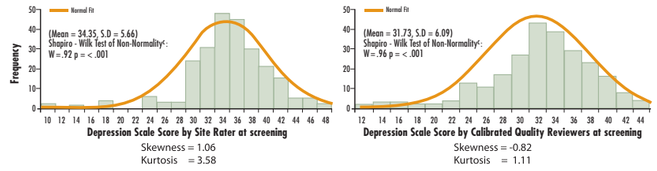
2. Remote Centralized Ratings' Baseline Score Distributions

- Uniformly produce normally distributed baseline scores (as assessed by Shapiro-Wilk statistic - W).
- Distributions are not skewed toward any inclusion criterion.



3. Distribution of Screening Scores for Total Depression Scale Score (N=245) for Site Raters compared to Central Raters

- Uniformly produce normally distributed baseline scores (as assessed by Shapiro-Wilk statistic - W).
- Distributions are not skewed toward any inclusion criterion.



4. Effect Size is Increased with Remote Centralized Ratings

- The effect size for subjects qualified by the blinded Remote Centralized Raters increased substantially over the effect size of subjects qualified by Site Raters in one study of GAD.

Generalized Anxiety Study
SITE RATINGS

Site Scores used to Enroll Students	All Subjects deemed eligible by sites ǂ			Subset Subjects deemed eligible by Central Raters		
		SD	n =		SD	n =
Change from Baseline						
Placebo	**9.6**	(±7.6)	98	**7.9**	(±7.8)	45
Lexapro	**12.1**	(±8.0)	47	**12.8**	(±7.4)	17
Pbo-Lexapro	**2.5**	(±7.7*)		**4.9**	(±7.7*)	
Drug Effect Size Least Square Means & root MSE	**0.32** **0.43**			**0.64** **0.74**		

ǂ Requested permission from BMS to publish the Central Ratings of "All Subjects" in addition to Site Ratings. Reply pending as of press.
* Pooled Standard Deviation

5. Subject Ascertainment Requires Careful Diagnosis: Data from two large studies

- Across these studies, Remote Centralized Raters found that 25-53% of site-referred study candidates did not meet diagnostic criteria.
- In the one study in which a third party observed all Centralized Raters diagnostic interviews, there was excellent diagnostic agreement between the Board Certified Psychiatrists and the Remote Centralized Raters.

CONCLUSIONS

- Subject selection issues are substantial – both in initial severity and diagnostic realms.
  - Over 1/3 of subjects may not meet protocol-specified inclusion/exclusion criteria for Initial severity.
  - 1/4 to 1/2 of subjects may not meet protocol-specified diagnostic criteria.
- An index of normality and internal consistency reliability of initial severity scores suggest that Remote Centralized Ratings (RCR) and continuous quality control (CQC) may minimize bias and improve subject ascertainment.
- Diagnosis by Remote Centralized Raters may improve appropriate subject ascertainment as determined by Independent third party verification.
- Subject ascertainment with Remote Centralized Raters roughly doubled the effect size between active comparator and placebo in one study of GAD.

References

Khan A, Kolts RL, Rapaport MH, Krishnan KR, Brodhead AE, Browns WA. Magnitude of placebo response and drug-placebo differences across psychiatric disorders. Psychol Med. 2005;35(5):743-9

Kobak K, Leuchter MD, DeBrota, D, Engelhardt E, Williams JBW, Cook I, Leon AC, Alpert J. Site vs. centralized raters in a clinical depression trial: Impact on patient selection and placebo response. Journal of Clinical Psychopharmacology, 2010; 30(2), 193-197.

Brown B, De Santi S, Detke M, Williams JBW. Assessing Interview Quality and Scoring Accuracy in Clinical Trials with Continuous Quality Control (CQC). (2010) The International Society for CNS Clinical Trials Methodology, Washington, DC.

Coric V, Stock E, Shekhar A, Pultz J, Dockens R. A randomized double-blind placebo-controlled and active comparator trial of pexacerfont, a novel corticotropin releasing factor receptor-1 antagonist, in the treatment of Generalized Anxiety Disorder. (2008) American College of Neuropsychopharmacology, Scottsdale, AZ.

This study was supported by MedAvante, Inc.

MEDAVANTE® RESEARCH INSTITUTE
Calibrated for Certainty