

The Challenge of Patient Ascertainment in Clinical Trials – New Data

Michael Detke^{1,2}, Janet Williams^{1,3}, Kenneth Kobak¹, Amy Ellis¹, Earl Giller¹, Andrew Leon⁴, Scott Reines¹, John Kane⁵

¹MedAvante Research Institute, ²Indiana University School of Medicine; Harvard Medical School, ³Columbia University College of Physicians and Surgeons, ⁴Weill Cornell Medical College, ⁵Zucker Hillside Hospital & Albert Einstein College of Medicine

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 patients enrolled in two MDD trials by site raters would be excluded based on the patient's self-rating or remote blinded clinicians' ratings of initial severity. New data on the extent and characteristics of patient ascertainment discrepancies and various methods to mediate it will be presented.

Methods: Inter-rater reliability and internal consistency reliability were assessed for site and remote raters in one 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 patients on the HAMA. In ongoing studies (of MDD, GAD & SZ) patients were assessed by both site raters and by remote blinded clinicians. In two of these studies, accuracy of diagnosis was examined.

Results: Internal consistency reliability (Cronbach's alpha) was strong for remote blinded clinicians at screening and endpoint and for site raters at endpoint (approximately .6 - .8) but much lower for site raters at screening. In the completed and ongoing studies of MDD, GAD & SZ, 34% (range: 5-56%) of patients included by site raters would have been excluded based on remote blinded clinicians' ratings of initial severity. SCID-CT assessments by remote blinded clinicians also revealed potential diagnostic errors in patients previously screened for study entry by site-based raters. In one study of GAD, patient ascertainment by remote blinded clinicians increased the drug effect size from .43 to .74.

Conclusion: Patient ascertainment issues are pervasive and substantial; on symptom severity alone over 1/3 of patients enrolled in clinical trials may not meet protocol-specified inclusion/exclusion criteria. Diagnosis is an additional source of potential error. Independent assessment of symptom severity by patients appears to have potential benefit in one MDD study. Remote blinded clinicians may be beneficial for diagnosis and symptom severity assessment across several diagnoses. Accurate patient ascertainment may substantially increase effect size.

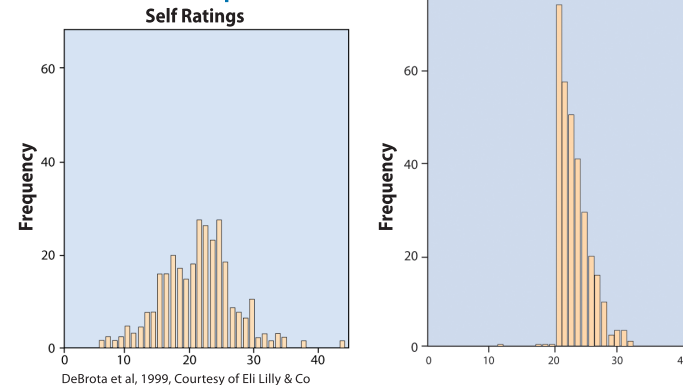
BACKGROUND

Literature Shows that CNS Clinical Trial Arms in FDA-Approved Drugs Fail to Separate More Frequently than Powering Predicts

| Results for Investigational arms | FDA Trial Database | | | |
|----------------------------------|--------------------|----------------|---------------------------|-----------|
| | Trials | Treatment Arms | Successful Treatment Arms | % Failure |
| Antidepressants | 52 | 93 | 45 | 51.6 |
| Anxiolytics | 40 | 75 | 36 | 52.0 |
| Antipsychotics | 17 | 49 | 37 | 24.5 |

Khan, et al. NCDEU, 2005.

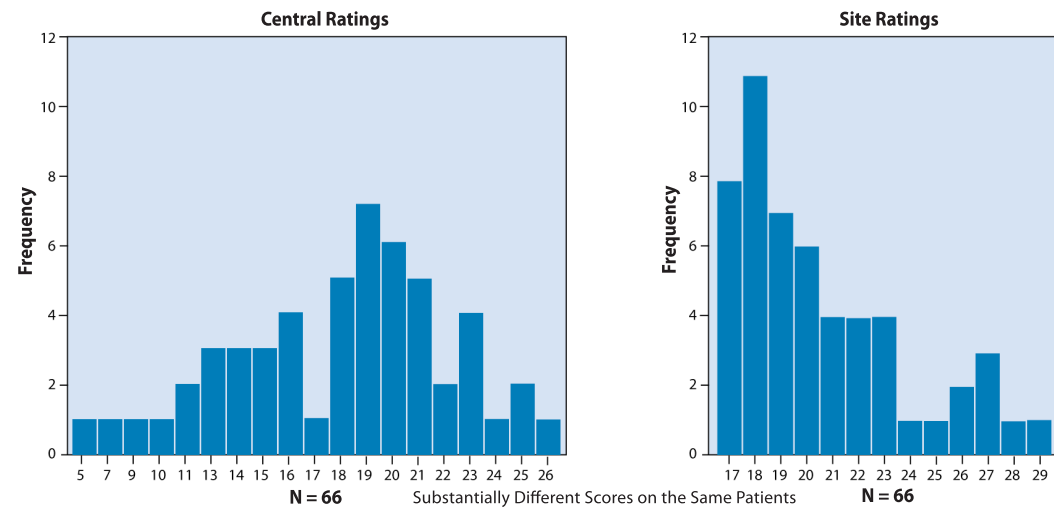
Previous Study #1 – Patient self-ratings would exclude 1/3 to 1/2 of patients in MDD



DeBrotta et al, 1999, Courtesy of Eli Lilly & Co

BACKGROUND continued

Previous Study #2 – Central clinician ratings would exclude 1/3 to 1/2 of patients in MDD



*Kobak, K. A., DeBrotta, D. J., Engelhardt, N., Williams, J.B.W. (2006, June). Site vs. Centralized Raters in a Clinical Depression Trial. National Institute of Mental Health, New Clinical Drug Evaluation Unit, 46th Annual Meeting, Boca Raton, FL.

METHODS

Internal consistency reliability was assessed in one MDD study. Internal consistency reliability was evaluated using Cronbach's coefficient alpha. Cronbach's alpha measures how well a set of items in a scale 'hang together', i.e., measure a single unidimensional latent construct. Cronbach's alpha is a common test of whether items are sufficiently interrelated to justify their combination in an index and will generally increase when the correlations between the items increase.

Two doses of an experimental compound were compared in a double-blind randomized placebo-controlled 30-site GAD study. Remote centralized raters (RCR) blind to study visit and entry criteria rated subjects by phone on the HAMA in addition to site raters at baseline and week 6; site raters evaluated patients in-person bi-weekly until week 8. Subjects required a minimum site rating of 20, and ≥ 2 on HAMA items 1 and 2 at screen and baseline. RCR HAMA followed site raters' evaluations at the baseline visit, and were counterbalanced at week 6.

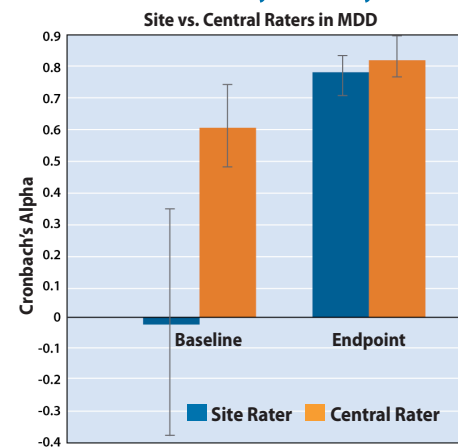
In several ongoing studies (of major depressive disorder, generalized anxiety disorder and schizophrenia) patients were assessed by both site raters and by RCR. Designs varied broadly in terms of entry criteria, inclusion criteria (determined by site raters or RCR) and other features, but are all placebo-controlled, parallel-arm acute drug treatment trials. In two of these studies, accuracy of diagnosis was also assessed by RCR.

RESULTS

Patient Ascertainment – Diagnosis

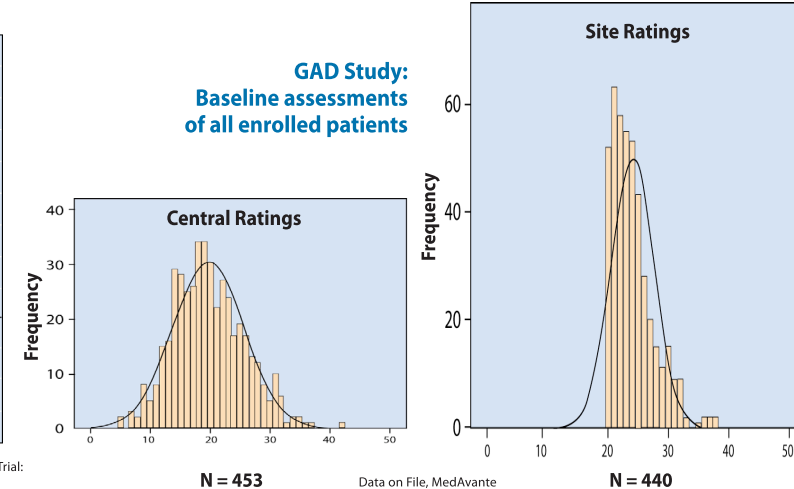
- Two large, ongoing studies in which remote central raters (RCR) are diagnosing patients using the SCID-CT.
 - Substantial proportion of patients failed to meet diagnostic criteria
- In one of the studies, 3rd Party Board-Certified Psychiatrists (compensated by sponsor) are joining by 3-way videoconference.
 - Excellent Diagnostic Agreement

Internal Consistency Reliability in MDD



Kobak, K. A., et al. Site vs. Centralized Raters in a Clinical Depression Trial: Impact on Patient Selection and Placebo Response. Under Review.

GAD Study: Baseline assessments of all enrolled patients



Site vs. Remote Central Raters Across Disorders

Patients Included by Site Raters that would have been Excluded by Remote Independent Blinded Clinicians – one study per line

| Indication | % Patients that would be excluded | Indication | % Patients that would be excluded |
|---------------------------------|-----------------------------------|------------|-----------------------------------|
| GAD | 43% | Psychosis | 27% |
| | 52% | | 32% |
| MDD | 18% | | 5% |
| | 26% | | 56% |
| 51% | | | |
| Range = 5% -- 56% | | | |
| Weighted Mean = 34% (1013/2995) | | | |

6/10/09, Data on File, MedAvante; All Ongoing or Completed Studies in these Diagnoses with >50 patients. Depending on protocol, Site or Central determined actual enrollment.

Effect of Patient Ascertainment by Remote Central Raters: Increased Effect Size

Remote, Independent, Blinded Clinicians: Significantly increased effect size. Same drug change; Lower placebo change

| HAM-A Baseline to Endpoint Δ | Site as Gatekeeper, Site Ratings (N=145) | Central Rater as Gatekeeper, Site Ratings (N=62) |
|-------------------------------------|------------------------------------------|--------------------------------------------------|
| Mean Placebo Δ | -9.6 ± 7.6 | -7.9 ± 7.8 |
| Mean Escitalopram Δ | -12.1 ± 8.0 | -12.8 ± 7.4 |
| Mean Esc - Placebo Δ | -3.1 ± 7.2 | -5.3 ± 7.2 |
| Effect Size | 0.43 | 0.74 |

Coric, et al. (2008) ACNP Annual Meeting.

Increasing Effect Size Allows for Reduction in N (and/or Increased Power)

Powerful benefit for either increasing power (certainty) or reducing sample (cost/time)

| N per group = 16 / (ES) ² | |
|--------------------------------------|-------------|
| Effect Size | N per group |
| .20 | 400 |
| .30 | 178 |
| .43 | 87 |
| .50 | 64 |
| .60 | 44 |
| .74 | 29 |

Assuming: 80% power; Two-tailed alpha = .05; t-test; equal N/group; Lehr R. (1992). 16 s-squared over d-squared: a relation for crude sample size estimates. Statistics in Medicine, 11, 1099-1102

CONCLUSIONS

- Patient ascertainment issues are pervasive and substantial
 - On symptom severity alone, over 1/3 of patients enrolled in clinical trials may not meet protocol-specified inclusion/exclusion criteria.
 - Diagnosis is an additional source of potential error in patient ascertainment.
- Independent assessment of symptom severity by patients appears to have potential benefit in one MDD study
- Remote raters appear to be beneficial for diagnosis and symptom severity assessment across several diagnoses.
- Accurate patient ascertainment may substantially increase effect size, which can improve power and/or reduce N required

ACKNOWLEDGEMENTS

Madeline Alexander, David DeBrotta, Nina Engelhardt, Arif Khan, Josh Lipsitz, Craig Mallinckrodt, Karen Meares

Author Disclosure Information: M. Detke, MedAvante, Eli Lilly, Part 1; MedAvante, Eli Lilly, Part 2; MedAvante, Eli Lilly, Part 3; MedAvante, Part 5; J. Williams, MedAvante, Part 1; MedAvante, Part 2; MedAvante, Part 3; MedAvante, Part 5; K. Kobak, MedAvante, Part 1; MedAvante, NIMH, Part 2; MedAvante, Part 5; A. Ellis, MedAvante, Part 1; MedAvante, Part 2; MedAvante, Part 3; MedAvante, Part 5; E. Giller, Pfizer, Wyeth, Memory Pharmaceuticals, Marinus Pharmaceuticals, MedAvante, Part 1; MedAvante, Part 2; MedAvante, Part 5; A. Leon, Astra-Zeneca (DSMB), Cyberonics, Dainippon Sumitomo (DSMB), MedAvante, Pfizer (DSMB), Schering-Plough (DSMB), Takeda, Part 1; NIMH, MedAvante, Pfizer, Part 2; J. Kane, Astra-Zeneca, BMS, Lilly, Wyeth, Otsuka, Janssen, Lundbeck, Pfizer, Cephalon, Proteus, Dainippon Sumitomo, Targacept, PGX Health, GSK, Vanda, MedAvante (equity interest), Part 1; Astra-Zeneca, Lilly, Janssen, Otsuka, BMS, Part 2.



MEDAVANTE RESEARCH INSTITUTE
Calibrated for Certainty