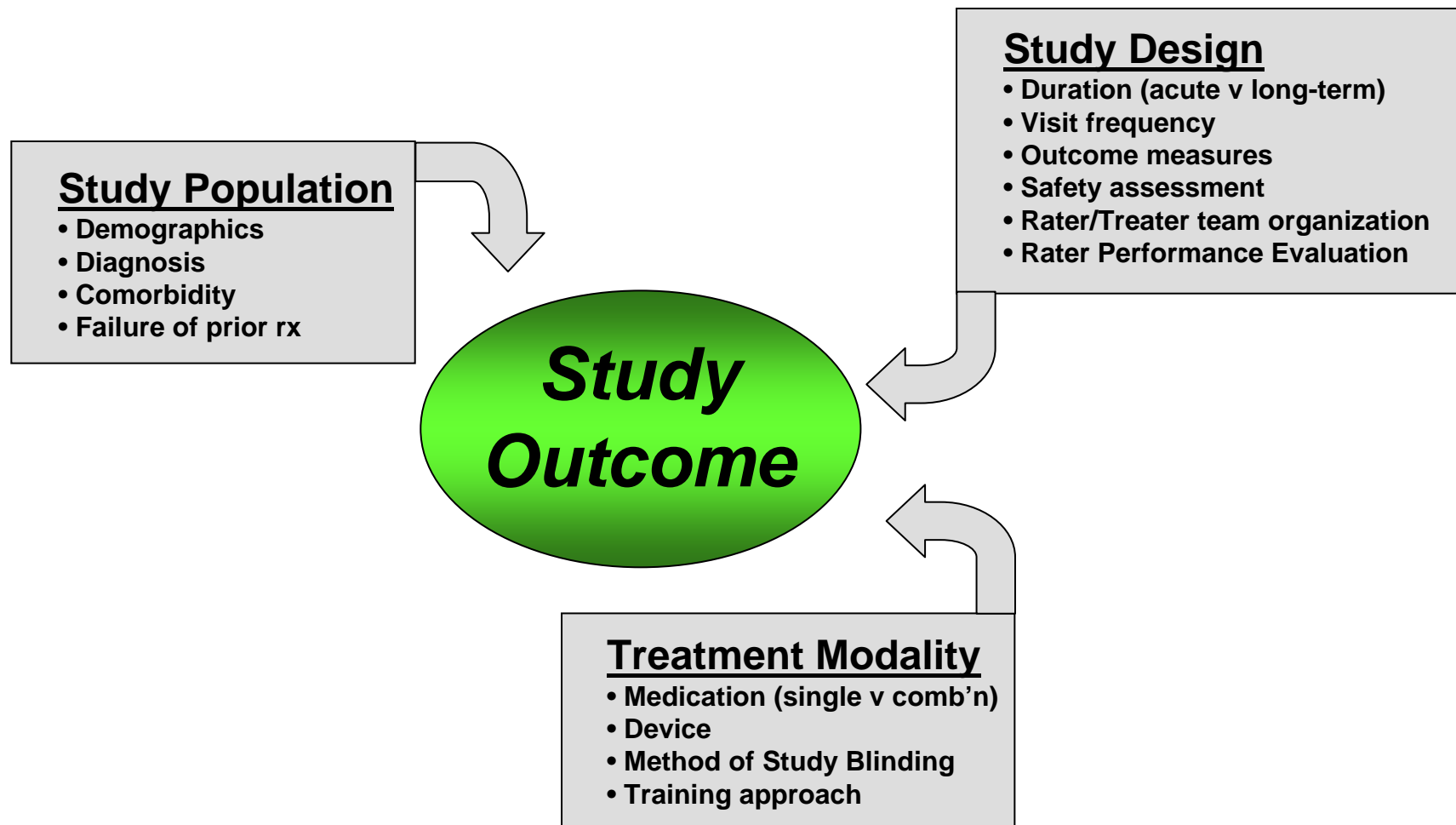


Clinical Trial Methodology and the Study of Medical Devices in Major Depression

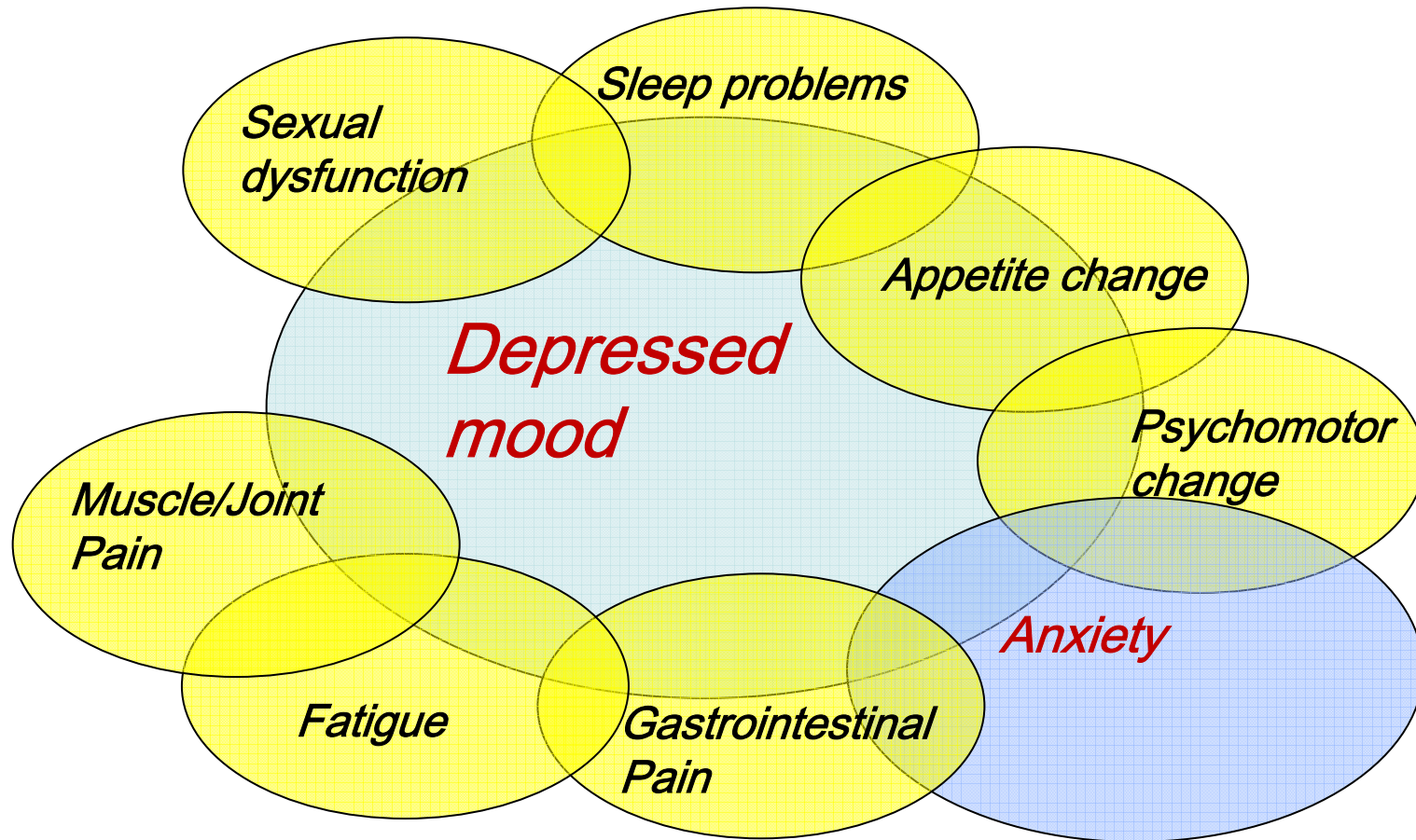
Mark A. Demitrack, MD
Vice President, Chief Medical Officer
Neuronetics, Inc.



Drugs vs Devices

Topic	Drugs	Devices
Training	<i>Rating scales</i>	<i>Rating scales</i> <i>Device performance</i>
Site Selection	<i>Minimal space needs</i> <i>'Basic' staffing configuration</i>	<i>Additional space, staff</i> <i>On-call/weekend coverage</i>
Outcome Measures	<i>'Tailored' for drugs</i>	<i>No device-specific scales</i>
Study Monitoring	<i>Document review</i>	<i>Document + Procedure review</i>
Study Design	<i>Time course well understood</i>	<i>Time course not always obvious</i>
Clinical Application	<i>Early in treatment</i>	<i>Later in treatment</i>
Regulatory environment	<i>Long history of experience, precedent</i>	<i>Most cases still first of kind</i>
Choice of comparator	<i>Many available options</i>	<i>Fewer options</i>
Study blinding	<i>Rarely examined</i>	<i>D---ed if you do...</i>

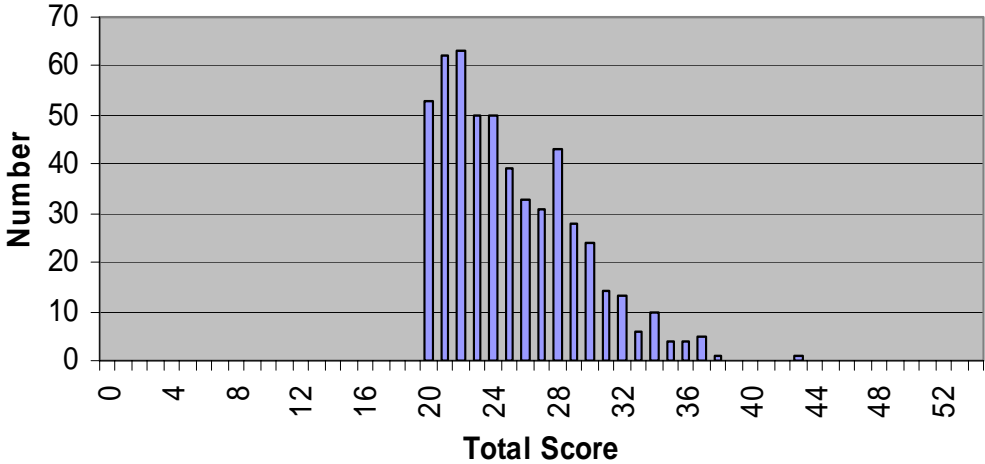
Major Depression: What's the Target?



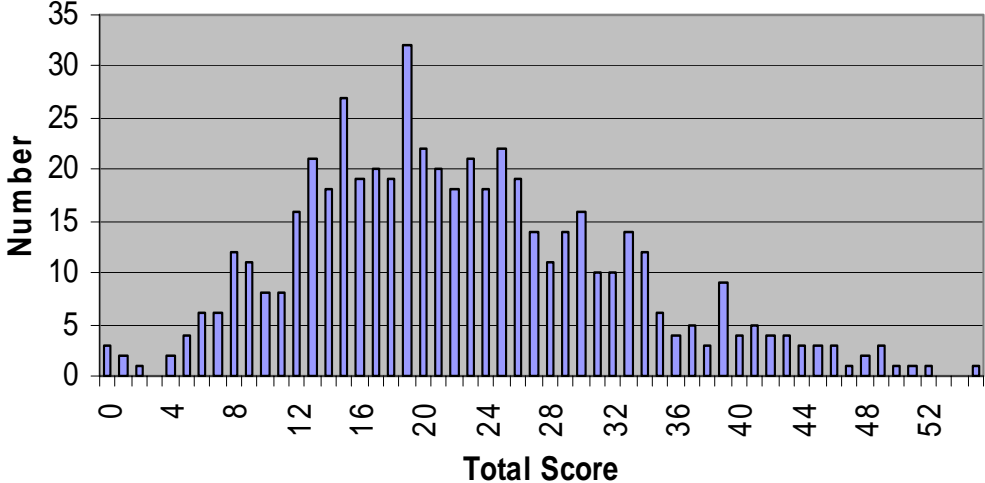
Rating Scales

- Rating scales are ‘unruly’ if not attended to...
 - Currently used scales for depression assessment were developed for specific medication profiles, not for specific devices
 - There are no examples of rating scales developed specifically for device studies
 - *Is efficacy expressed the same way for a drug and device?*
 - *Why assume that we should use the same rating scales in drug and device studies?*
-
-

Clinician HAMA visit 2



IVR-HAMA visit 2



Rating Scales – An Example...

- Large, Phase II, multisite antidepressant trial
 - 32 clinical sites
 - 3 countries
 - Each site chosen after an extensive, sponsor-driven assessment of skills, experience, and capabilities in psychiatric drug trials relevant to the needs of this study
 - Rater training performed in an extended, 6 hour session, in three groups
 - A total of 85 clinical raters were trained in this process
 - Psychopharmacology Bulletin, 1998; 34(1):19-24

Rater Training Methods

- Inter-rater reliability training sessions were conducted in a 6-hour format
 - All sessions led by the same trainer using a study-customized manual
 - Lecture on rating expectations/conventions
 - Four videotapes used: After each tape, individual items and total scores for the group were calculated and shown with discussion of areas of disagreement
 - Large group and small breakout formats employed
 - All raters judged the tapes of good visual and auditory quality to assess patient and observed rater performance
-

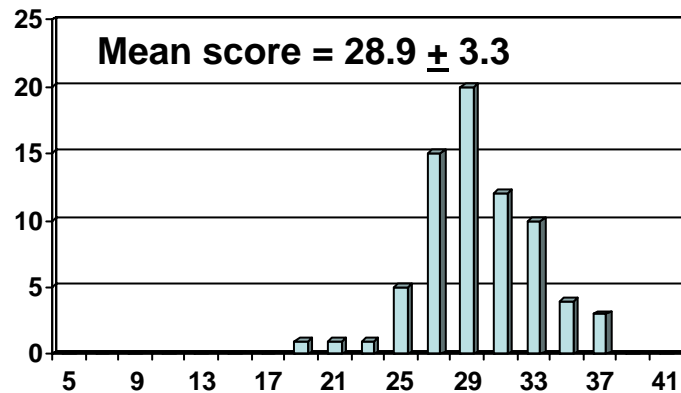
Self-Assessment of Rater Performance

How close do you predict that your HAMD total score will be to the mean total score of all raters viewing this tape at this meeting?

<u>Response</u>	<u>Tape 1</u>	<u>Tape 2</u>	<u>Tape 3</u>	<u>Tape 4</u>
• Within 1 point	0	0	5	1.3
• Within 2 points	28.2	21.8	30	15
• Within 3 points	52.6	51.3	52.5	50
• Within 5 points	19.2	25.6	12.5	28.8
• Not within 5 points	0	1.3	0	5

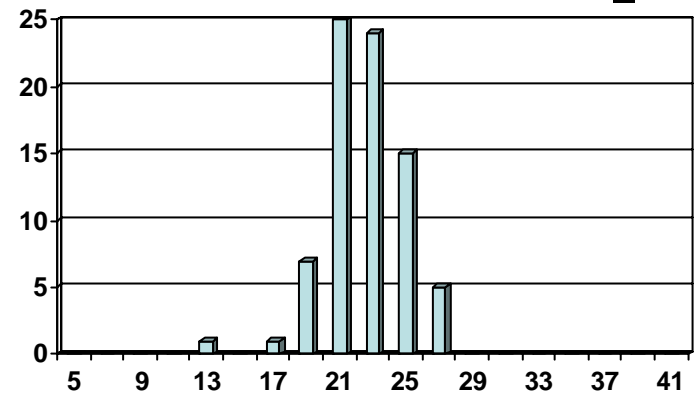
Distribution of Total HAMD Scores

TAPE 1

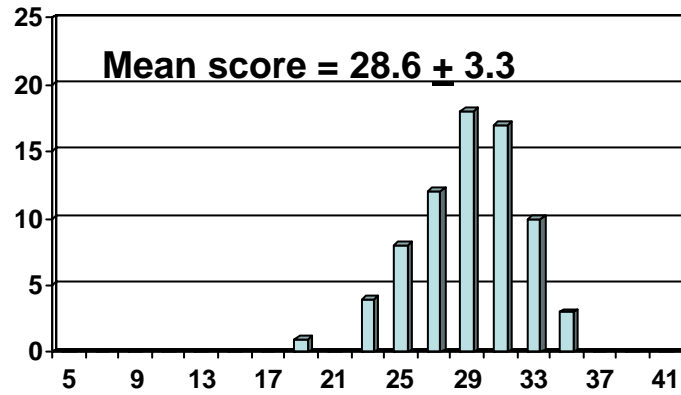


TAPE 3

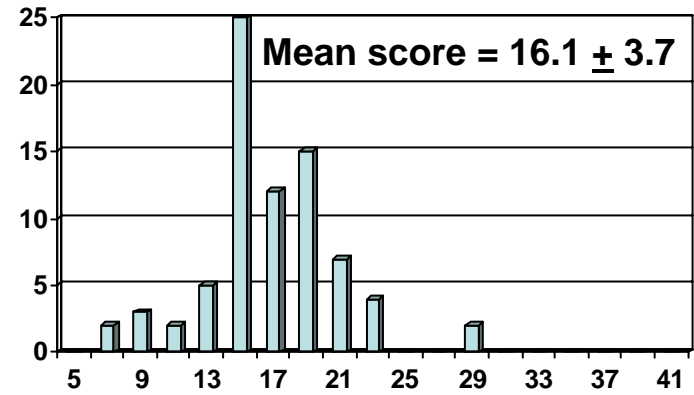
Mean score = 21.9 ± 2.3



TAPE 2



TAPE 4

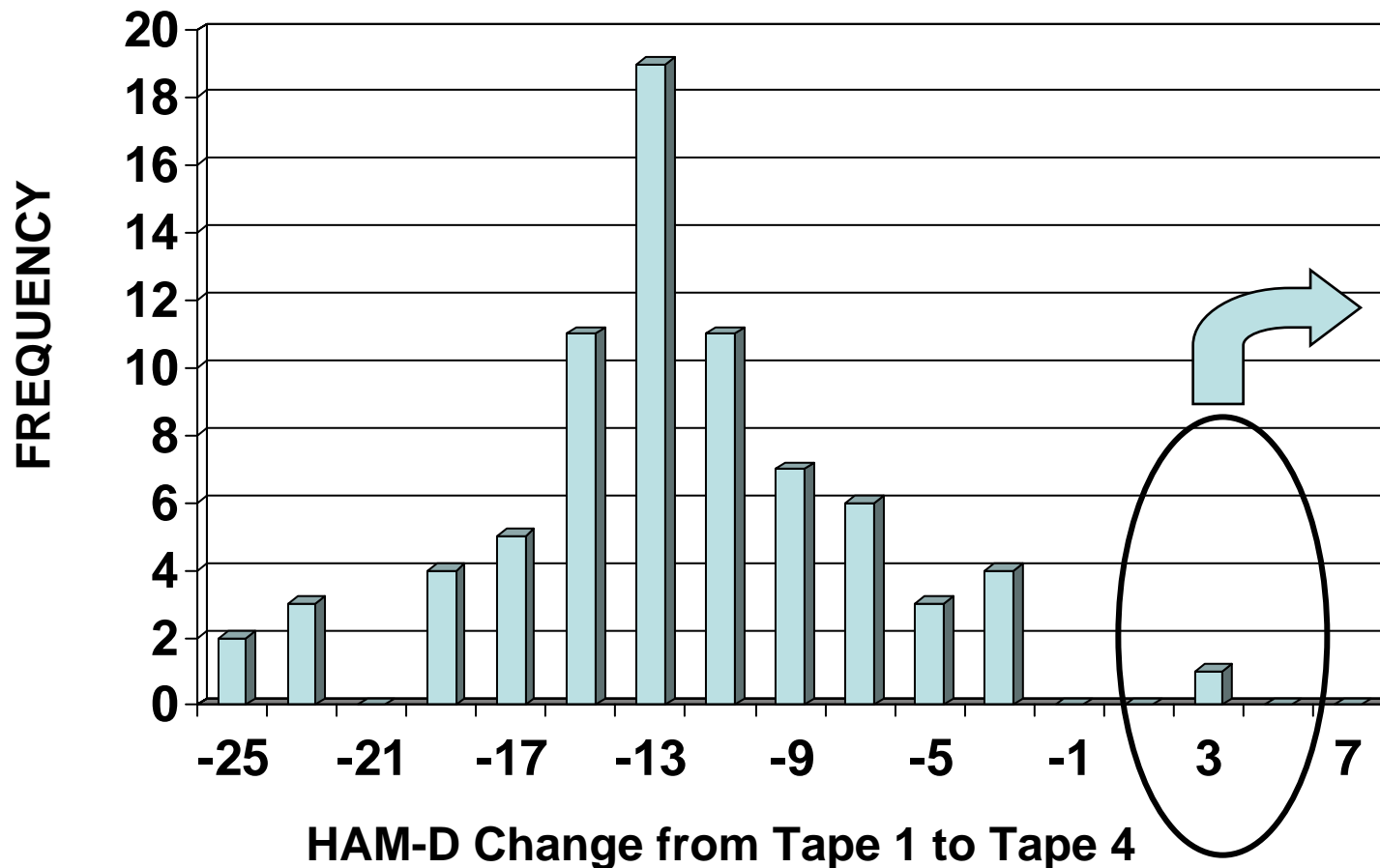


% Rater Agreement

Item	Tape 1	Tape 1	Tape 1	Tape 1	Average
Depressed Mood	.44	.54	.45	.42	.48
Feelings of Guilt	.46	.56	.74	.44	.56
Suicide	.50	.47	.69	.39	.52
Insomnia, early	.86	.76	.75	.71	.77
Insomnia, middle	.53	.49	.81	.36	.55
Insomnia, late	.45	.57	.80	.58	.61
Work and Activities	.79	.71	.82	.39	.68
Retardation	.41	.44	.49	.46	.45
Agitation	.31	.46	.43	.61	.46
Anxiety Psychic	.40	.44	.84	.60	.58
Anxiety Somatic	.45	.60	.64	.88	.65
Somatic symptoms, GI	.61	.74	.76	.88	.76
Somatic symptoms, general	.57	.54	.52	.47	.53
Genital symptoms	.51	.95	.93	.88	.82
Hypochondriasis	.40	.44	.42	.77	.51
Loss of weight	.90	.75	1.00	.98	.91
Insight	.70	.88	.76	.88	.81

*(% agreement refers to the overall proportion of time that any pair of raters
Had concordant scores for the item in question)*

HAMD Change Score Detected from Tape 1 to 4



This rater saw the patient as more ill in the tape where she was actually recovered!

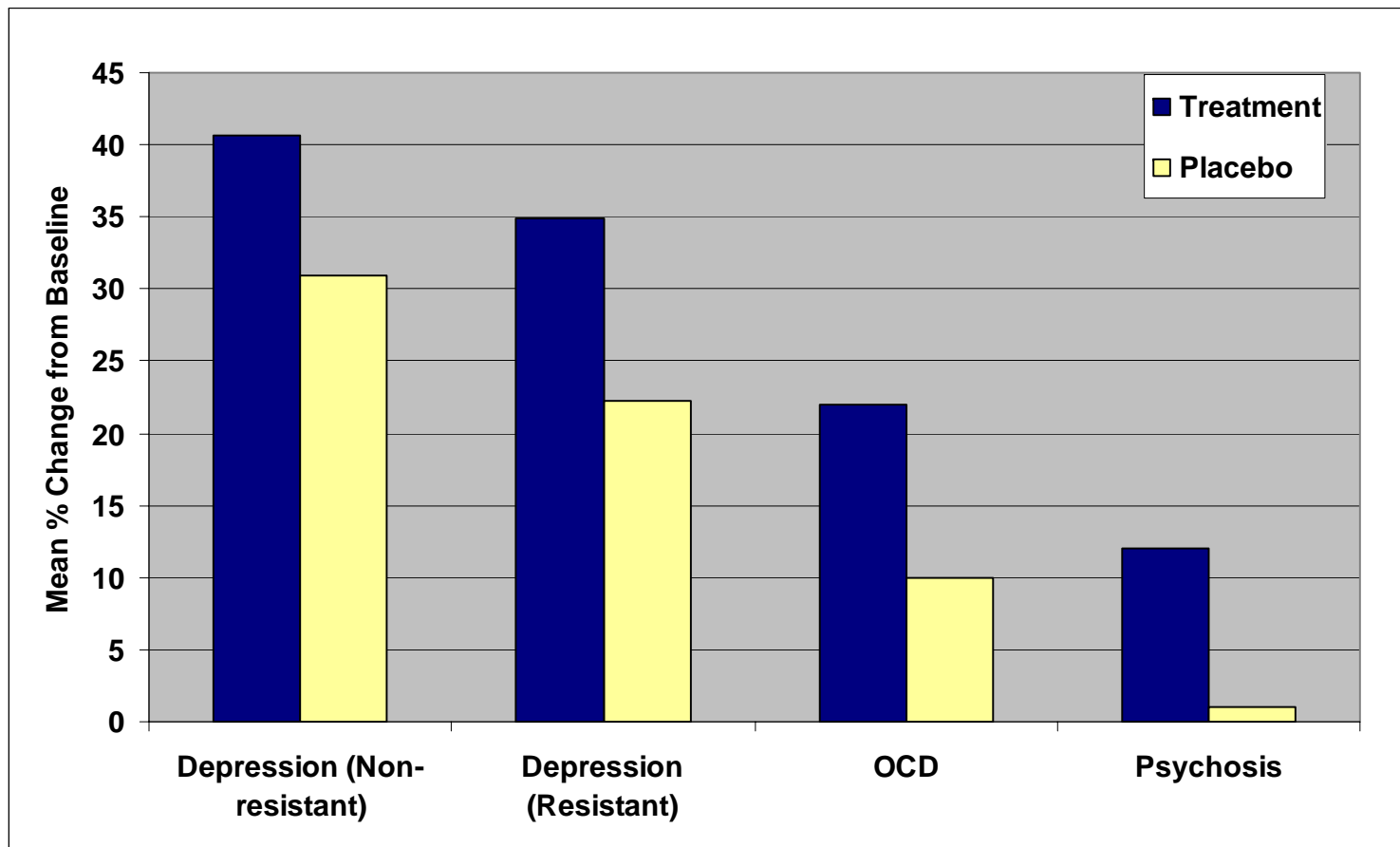
- Measurement error associated with a clinical rating scales can be large in a multisite clinical trial setting
 - The difference ranged up to 21 points in total score
 - Actual power of the study likely overestimated
 - Short term rater training appeared to have little influence on the intra-class measures observed
 - “Garbage in, garbage out”
 - Most psychiatric rating scales are methodologically “unruly” measures
-
-

- Drugs typically used early in treatment, while devices, usually later in algorithm
 - It is likely that devices will therefore be studied in more complex, treatment-resistant forms of the illness
 - This fact has implications for placebo response expectation, and interpretation of study outcome
-
-

Why is the Placebo Effect Important?

- Magnitude of symptom reduction with placebo *drives the active treatment effect* and in turn clinical trial outcome
- Treatment resistant depression is associated with *lower placebo effect* and hence, lower active treatment response
- Low placebo effect *increases the confidence in the validity of the study* result

Magnitude of Placebo Effect Drives the Active Treatment Effect



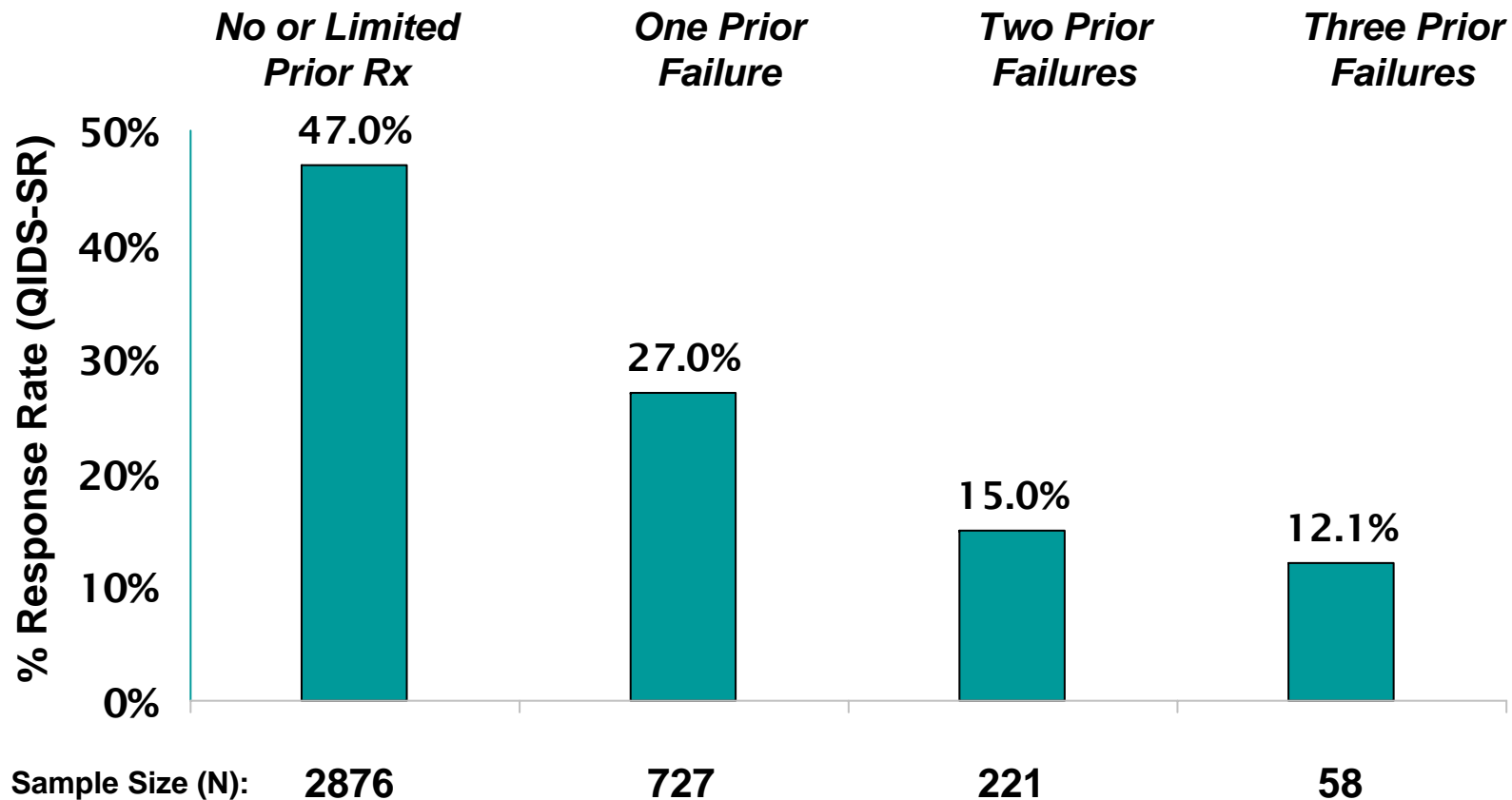
Influence of Medication Resistance on Response to Treatment

ECT response rate (%)	Total Group (N=100)		Patients Who Were Medication Resistant (N=65)		Patients Given Inadequate Pharmacotherapy (N=35)	
	N	%	N	%	N	%
Immediately after ECT	73	73.0	41	63.1	32	91.4
One week post-ECT	57	57.0	31	47.7	26	74.3

Prudic, et al (1996)

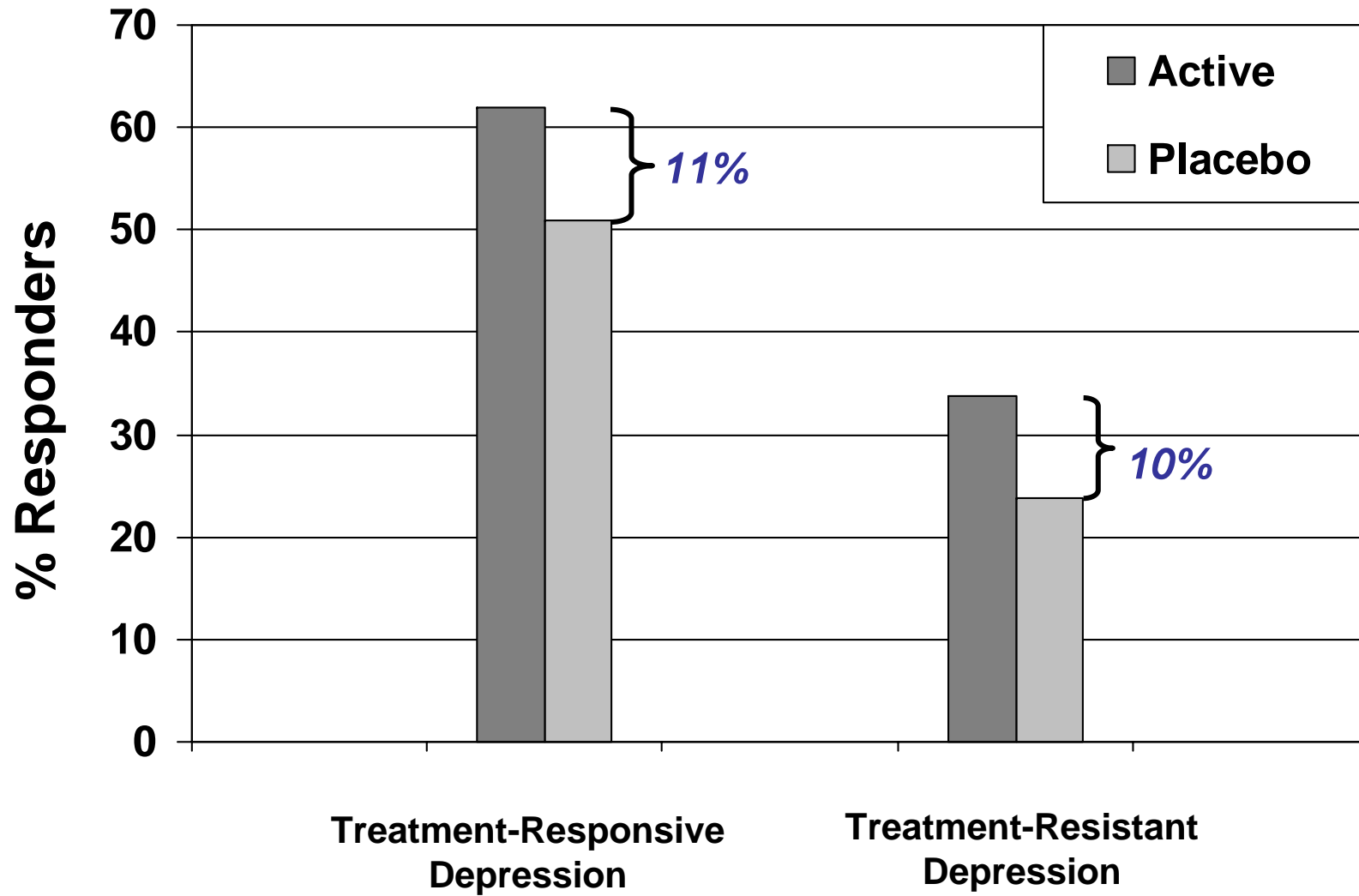
Categorical Outcome (Response) Decreases with Number of Prior Treatment Failures

*STAR*D Study Results*



Trivedi et al. (Am J Psychiatry, 2006); Rush et al. (NEJM, 2006);
Fava et al (Am J Psychiatry, 2006); McGrath et al (Am J Psychiatry, 2006)

Treatment Difference Between Active and Placebo Constant WRT Prior Treatment Resistance



Drugs vs Devices

Topic	Drugs	Devices
Training	<i>Rating scales</i>	<i>Rating scales</i> <i>Device performance</i>
Site Selection	<i>Minimal space needs</i> <i>'Basic' staffing configuration</i>	<i>Additional space, staff</i> <i>On-call/weekend coverage</i>
Outcome Measures	<i>'Tailored' for drugs</i>	<i>No device-specific scales</i>
Study Monitoring	<i>Document review</i>	<i>Document + Procedure review</i>
Study Design	<i>Time course well understood</i>	<i>Time course not always obvious</i>
Clinical Application	<i>Early in treatment</i>	<i>Later in treatment</i>
Regulatory environment	<i>Long history of experience, precedent</i>	<i>Most cases still first of kind</i>
Choice of comparator	<i>Many available options</i>	<i>Fewer options</i>
Study blinding	<i>Rarely examined</i>	<i>D---ed if you do...</i>

Thanks
