Who Studies Whom and How:
Why Clinical Trials Fail in Mood And Anxiety Disorders

Boadie W. Dunlop, MD

Director, Mood and Anxiety Disorders Program
Professor
Department of Psychiatry and Behavioral Sciences
Emory University School of Medicine

Dr. Boadie Dunlop
Personal/Professional Financial Relationships with Industry within the past year

External Industry Relationships *	Company Name(s)	Role
Equity, stock, or options in biomedical industry companies or publishers**	None	
Board of Directors or officer	None	
Royalties from Emory or from external entity	None	
Industry funds to Emory for my research	NIMH, Compass, Otsuka, Usona, Boehringer-Ingelheim, AIFRED	Multiple clinical trial projects
Other	Myriad Neurosciences, Dept of Defense, Sage, Otsuka, Cerebral Therapeutics, Biohaven	Consultant

*Consulting, scientific advisory board, industry-sponsored CME, expert witness for company, FDA representative for company, publishing contract, etc.
**Does not include stock in publicly-traded companies in retirement funds and other pooled investment accounts managed by others.

A patient comes to see you with a problem. How do you know what to do?

1. Evidence Base of the Literature
   – Internal Validity of a Clinical Trial
   – External Validity (Generalizability) to populations outside the trial
2. Your clinical experience, acumen, knowledge, skills, and wisdom.

*** Published evidence base should INFORM, not dictate, clinical practice ***
EBM ≠ EDM

Outcomes in Depression Trials

- <u>Continuous Measures of Effect Size:</u>
  - Absolute reduction in HAM-D score
  - Cohen's *d* or Hedges' *g* to compare across studies using different scales (Standardized Mean Difference, SMD)
- <u>Response</u>
  - significant improvement but not necessarily complete relief of symptoms
  - often measured by ≥50% decrease from baseline HAM-D score
- <u>Remission</u>
  - minimal or no symptoms
  - return to functional normality
  - often measured as HAM-D ≤ 7

Frank E et al. *Arch Gen Psychiatry*. 1991;48:851-855; Rush AJ, Trivedi MH. *Psychiatr Ann*. 1995;25:704-709.

Standardized Mean Difference (SMD)

- SMD is a summary statistic in meta-analysis when the studies *all assess the same outcome* but <u>measure it different ways</u> (e.g depressive symptoms with HAMD or MADRS).

- It is necessary to standardize the results of the studies to a uniform scale before they can be combined.

- The standardized mean difference expresses the size of the intervention effect in each *study relative to the variability* observed in that study. (A difference in means, not a mean of differences.)

- Thus, studies for which the difference in means is the same proportion of the standard deviation (SD) will have the same SMD, regardless of the actual scales used to make the measurements.

- <u>Note:</u> the method assumes that the differences in SDs among studies reflect differences in measurement scales and not real differences in variability among study populations. This assumption may not always hold: e.g. Pragmatic versus Phase II trials: pragmatic trials may include a wider range of participants and may consequently have higher standard deviations.

$$SMD = \frac{\text{Difference in mean outcome between groups}}{\text{Standard deviation of outcome among participants}}$$

Effect Sizes

- **Effect size** tells you how meaningful the relationship between variables or the difference between groups is. It indicates the <u>practical significance</u> of a research outcome.
  - There are many ways to report "effect size."
    - Correlations: Pearson's *r*
    - Differences between groups: Cohen's *d* or Hedges' *g* (for small n's), SMD
- In meta-analyses, effect sizes typically, though not always, refer to versions of the SMD.
- "Statistical significance" (e.g., p<.05) only tells the likelihood of the result arising by chance, NOT how important the result is.
- Common thresholds:

Effect size	Cohen's *d*	Pearson's *r*
Small	0.2	.1 to .3 or -.1 to -.3
Medium	0.5	.3 to .5 or -.3 to -.5
Large	0.8 or greater	.5 or greater or -.5 or less

The Usual Plot
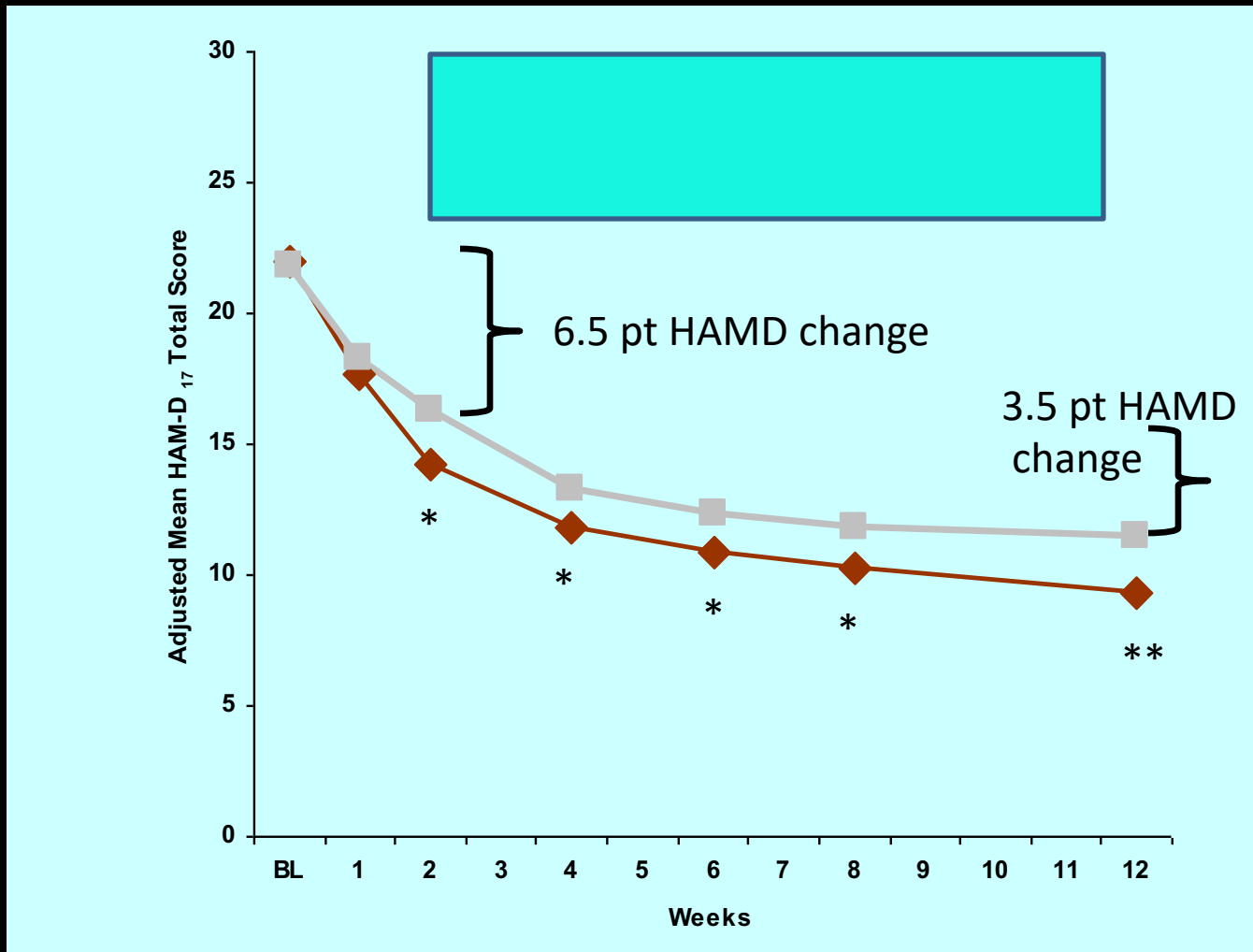
* = Significant!   **= Extra Significant!!

Dunlop et al, J Clin Psychopharmacol, 2011; 31:569-76

Companies Hiding Negative Data?

- 12 antidepressants

- <u>Compared:</u>
- FDA Database (N=74) vs
- Published Literature (N=65)
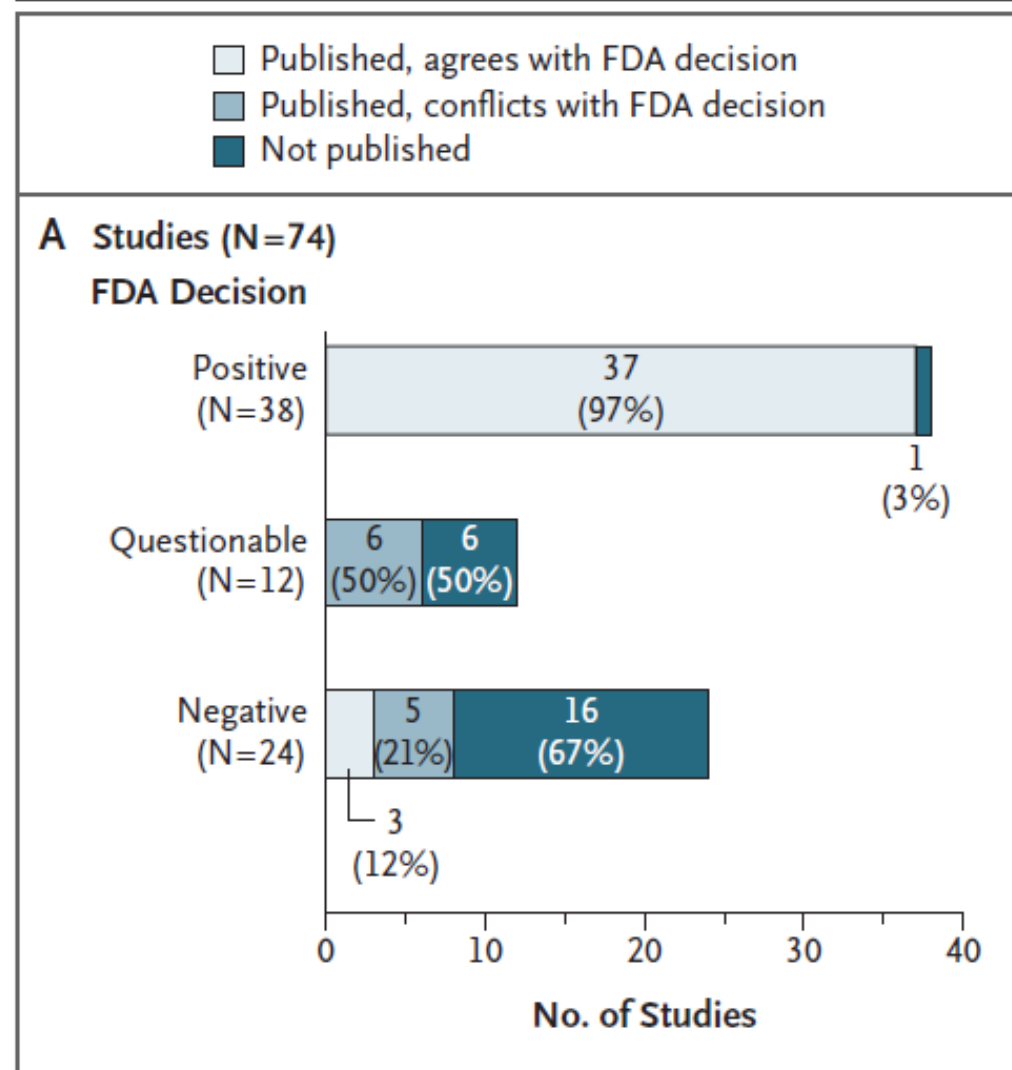
- <u>Results:</u>
Publication vs FDA conclusion:
- Agreement: 54%
- Conflicting: 15%
- Not published: 31%

Overall effect size:
- From Publications: g= 0.41
- All FDA trials: g =0.31

Turner et al, NEJM, 2008; 358: 252-60

Kirsch: Antidepressants = Placebo for Most

Meta-analysis
- 35 studies in FDA database
- 4 antidepressants vs PBO
- a priori "clinically significant"
  - SMD: d ≥ 0.5
  - HAMD: change ≥ 3

Results
- Drug: d = 1.24 (HAMD: 9.6)
- PBO: d = 0.92 (HAMD: 7.8)
- Difference: d = 0.32
  - HAMD: 1.8

- 80% of drug response is captured by PBO response
- Drug only meaningfully better than PBO in in severely depressed (Baseline HAMD=28)

"There seems little evidence to support the prescription of antidepressant medication to any but the most severely depressed patients, unless alternative treatments have failed to provide benefit."

Kirsch et al, PLOS One, 2008; 5 (2): e45
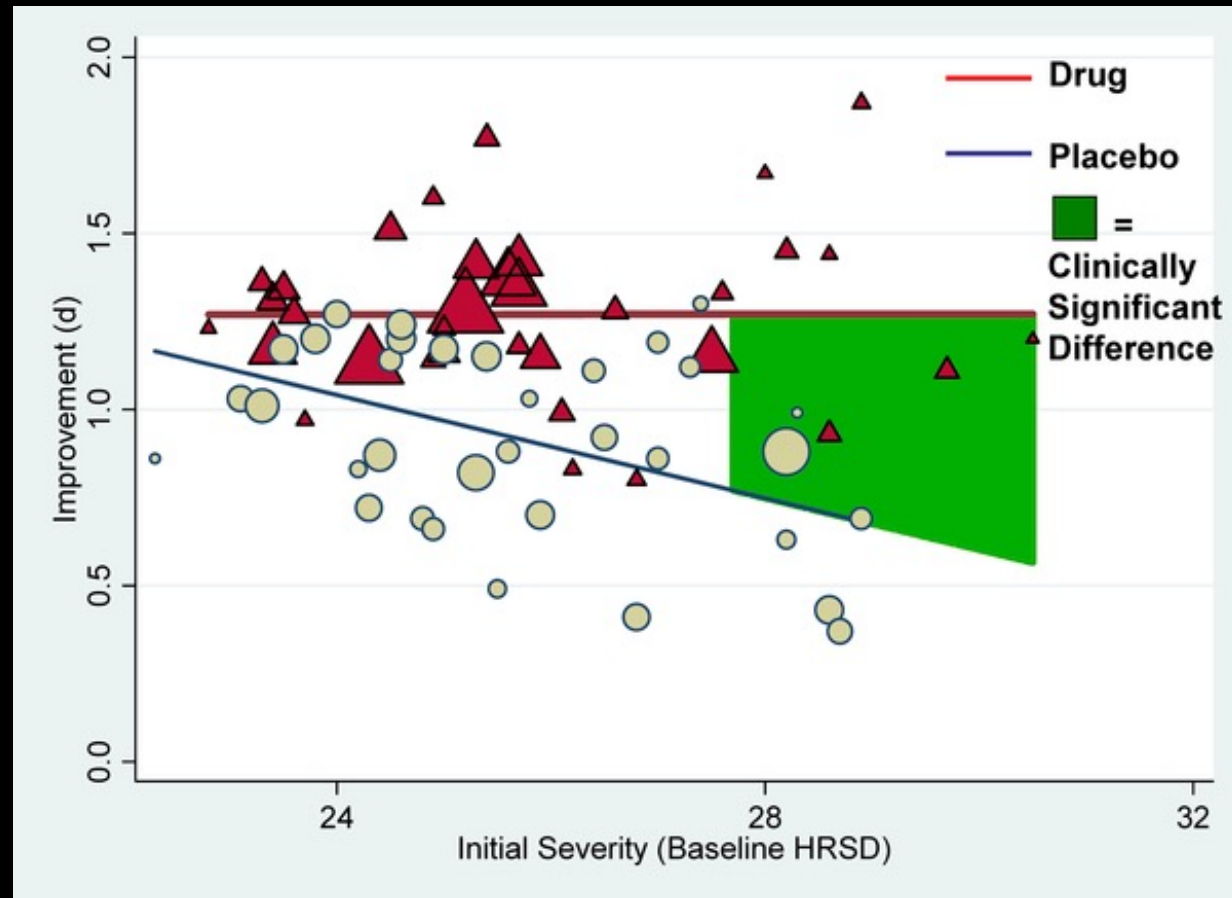
Change in HAMD Following Treatment With Antidepressant vs Placebo
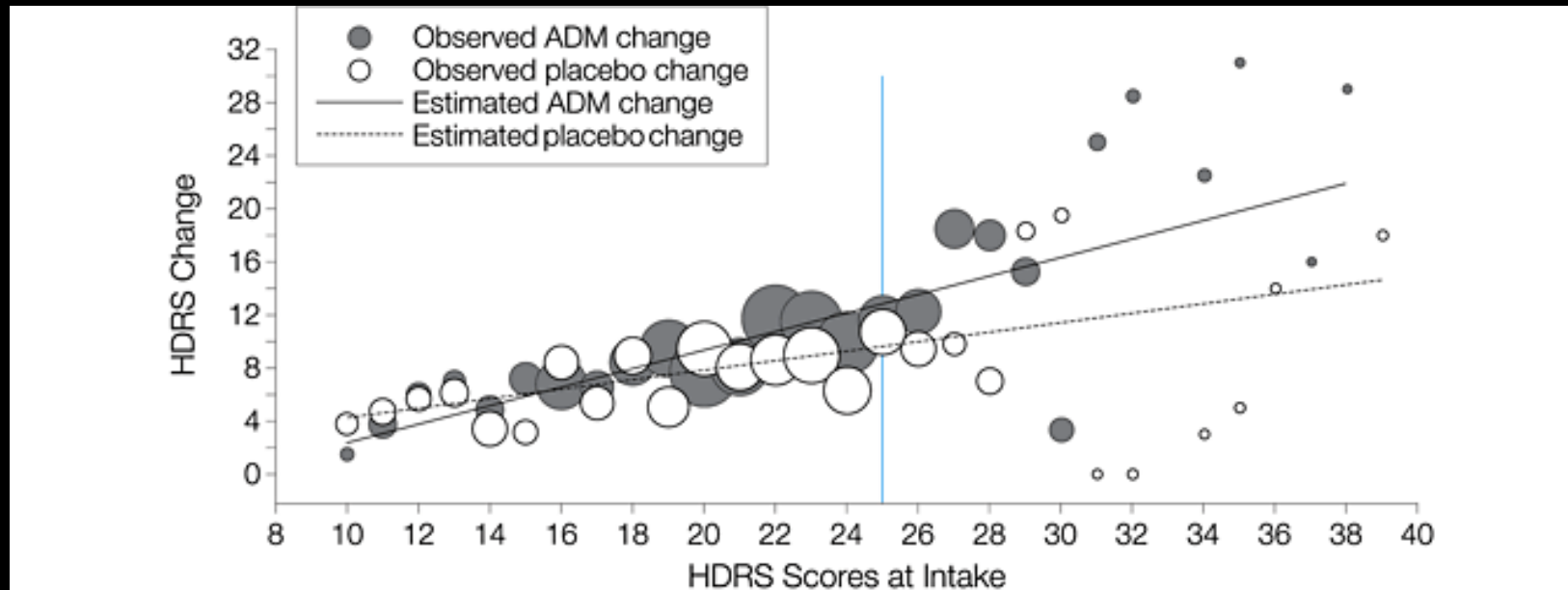
- Subject-level meta-analysis of 6 studies (718 patients) that did not use a placebo lead-in.
- 3 Point HAMD difference at Baseline HAMD=25

Effect Sizes for Antidepressants and Placebos in Depression

	PBO SMD (d)	Drug SMD (d)	Difference (d)	% Drug improvement "explained" by PBO
Kirsch et al. 2008	0.92	1.24	0.32	74%
Dunlop et al, 2012	1.15	1.46	0.31	79%
Rief et al, 2011 *	1.69	2.50	0.81	68%

• Rief et al included studies of dysthymia and minor depression.
For MDD studies, SMD for Placebo: d=1.83

The Crisis in Antidepressants

BAD MEDICINE

HALF OF ALL CLINICAL TRIALS AREN'T PUBLISHED, LEAVING DOCTORS AND PATIENTS IN THE DARK

Newsweek

11.21.2010

Newsweek

HEALTH

AUTHOR

Sharon Begley

The Depressing News About Antidepressants

Jan 28, 2010 7:00 PM EST

Studies suggest that the popular drugs are no more effective than a placebo. In fact, they may be worse.

How can ADM work if low serotonin doesn't cause MDD?

Daily Mail health .com

Depression 'is NOT caused by low serotonin levels': Study casts doubt over widespread use of potent drugs designed to treat chemical imbalance in brain

- Depression sufferers have been urged to still continue taking their medication

By JOHN ELY SENIOR HEALTH REPORTER FOR MAILONLINE
PUBLISHED: 20:00 EDT, 19 July 2022 | UPDATED: 20:06 EDT, 19 July 2022

Molecular Psychiatry

SYSTEMATIC REVIEW OPEN

The serotonin theory of depression: a systematic umbrella review of the evidence

Joanna Moncrieff[1,2], Ruth E. Cooper[3], Tom Stockmann[4], Simone Amendola[5], Michael P. Hengartner[6] and Mark A. Horowitz[1,2]

© The Author(s) 2022

"This review suggests that the huge research effort based on the serotonin hypothesis has not produced convincing evidence of a biochemical basis to depression."

icshades, Towie, Wallis And Futuna, 2 months ago

Before getting a prescription, make sure the people around you aren't the problem.

Click to rate ⬆ 1742 ⬇ 38

Study finds depression is not likely caused by a chemical imbalance or low serotonin levels (fox5atlanta.com)

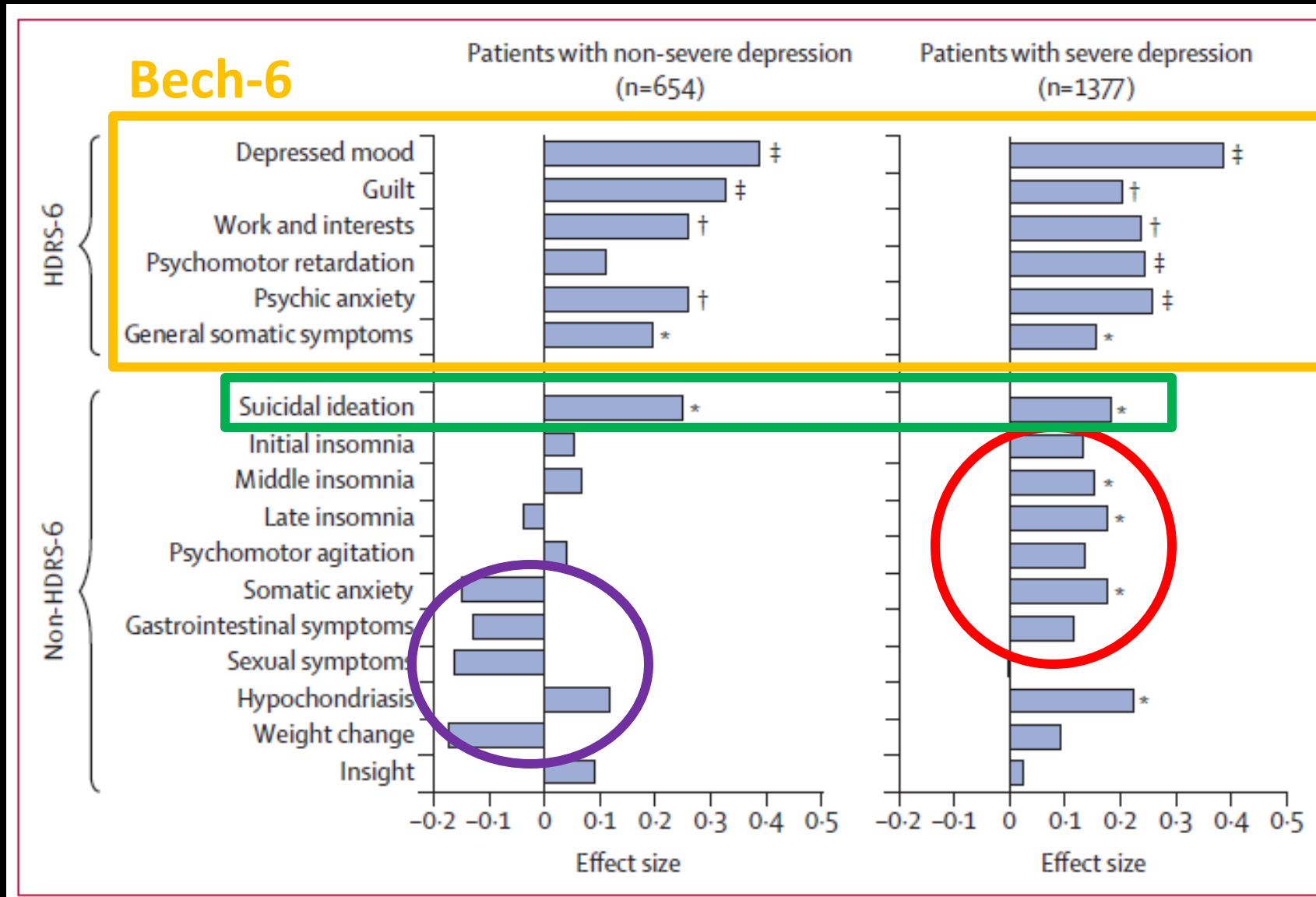# Antidepressants DO work across range of severity

- Patient-level meta-analysis of 1st 6 weeks of treatment with fluoxetine, venlafaxine or PBO
- Children, adult, geriatric studies (N=41)
- <u>HAMD Change:</u>
  - Overall: Med: -11.8   - PBO: -9.3   - Difference: 2.6
    - 27.7% greater improvement for drug
  - Low Severity: Med: -9.4   - PBO: -7.2   - Difference: 2.2
  - High Severity: Med: -12.9   - PBO: 10.1   - Difference: 2.8
- <u>Response Rates:</u>
  - Overall: Med: 58.4%   - PBO: 39.9%   - OR: 2.11 (NNT=5.4)
  - Low Severity: Med: 54.8%   - PBO: 37.3%   - Difference: 17.5%
  - High Severity: Med: 57.7%   - PBO: 40.5%   - Difference: 17.2%
- <u>Remission Rates:</u>
  - Med: 43.0%   - PBO: 29.3%   - OR: 1.82 (NNT=7.3)

Gibbons et al, Arch Gen Psychiatry, 2012

Sensitivity of HAMD
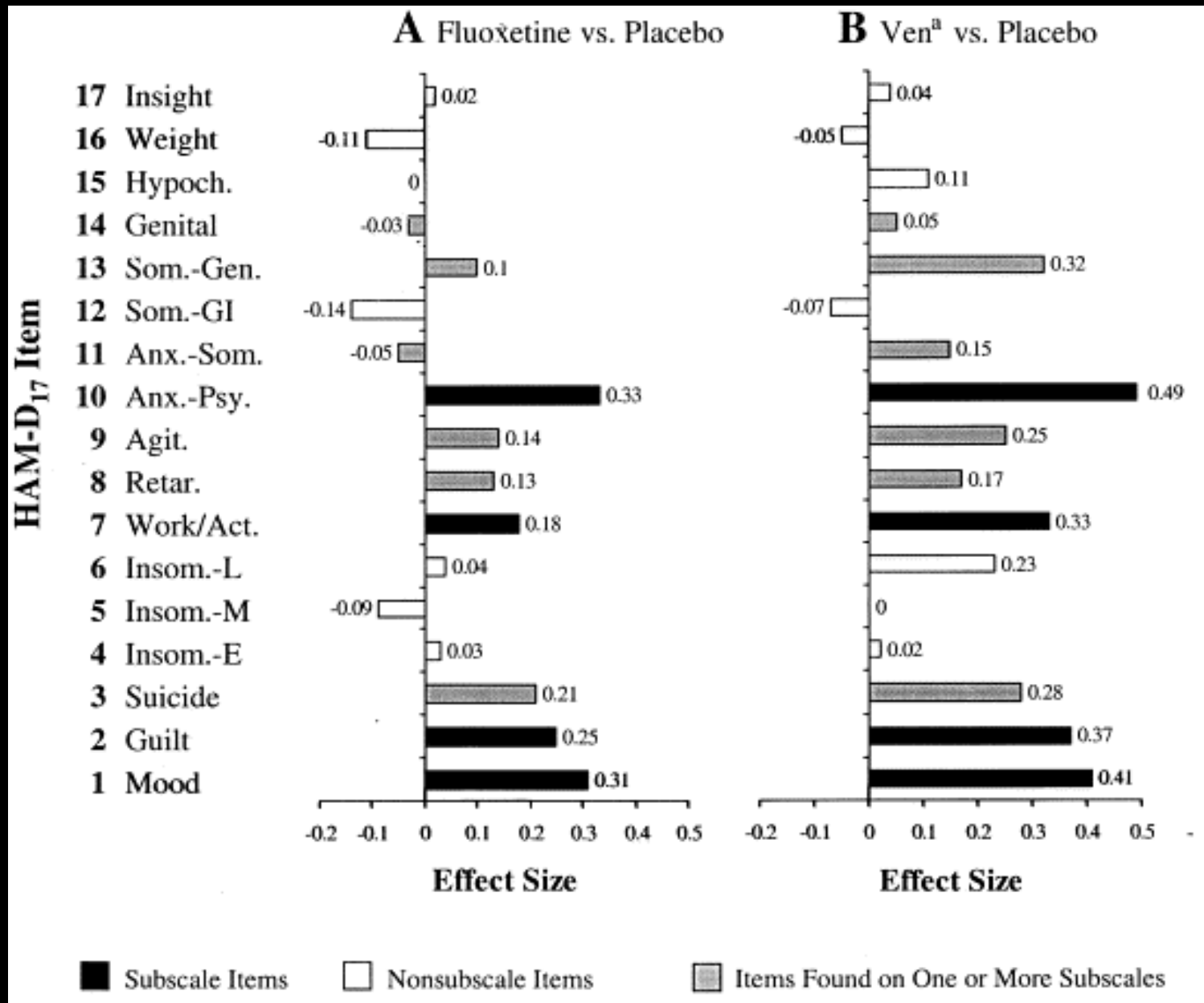
Entsuah et al, J Psychiatr Res, 2002; 36:437-48

15

SSRIs in Severe vs Non-Severe MDD
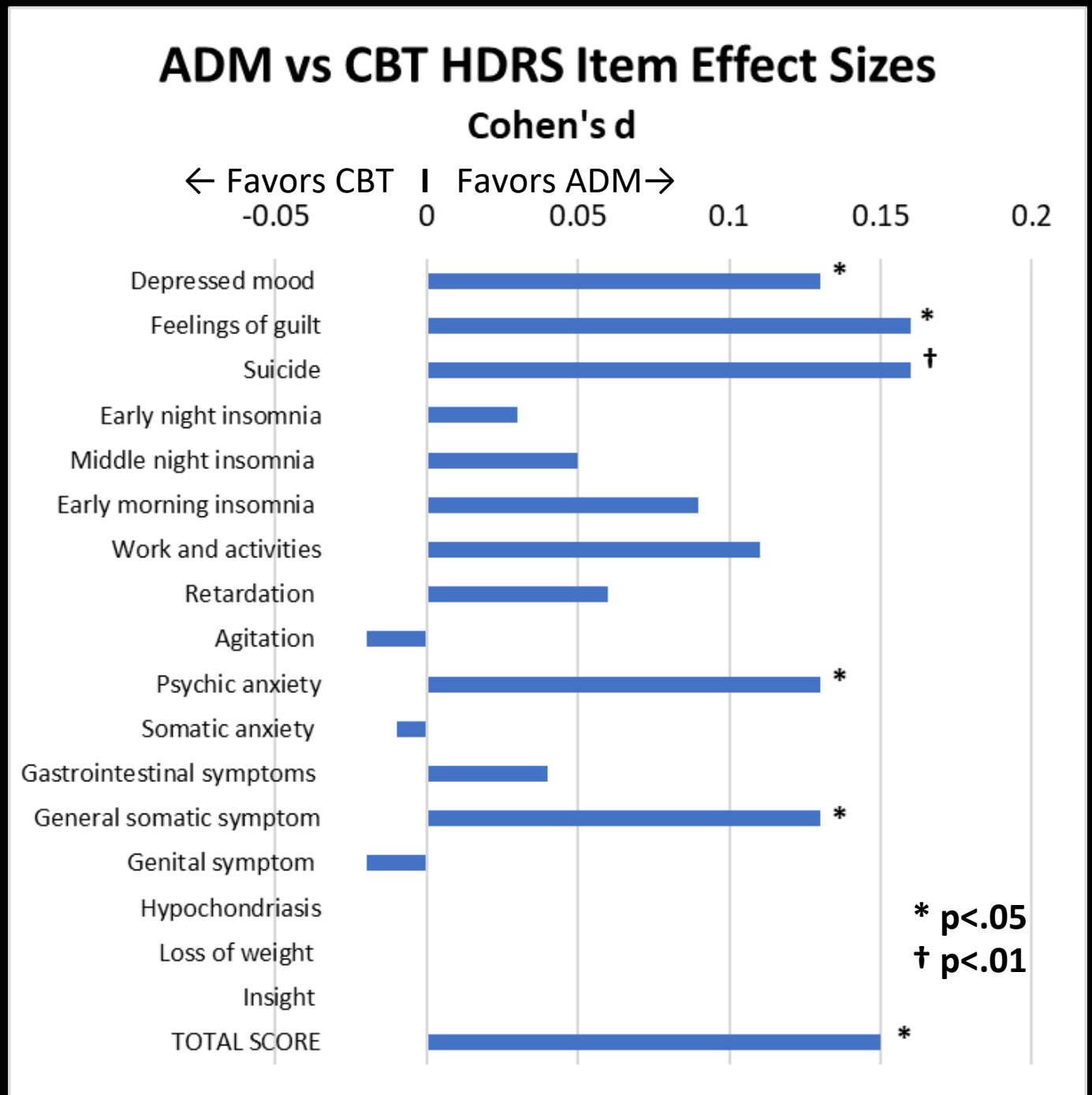

Hieronymus et al., Lancet Psychiatry, 2019

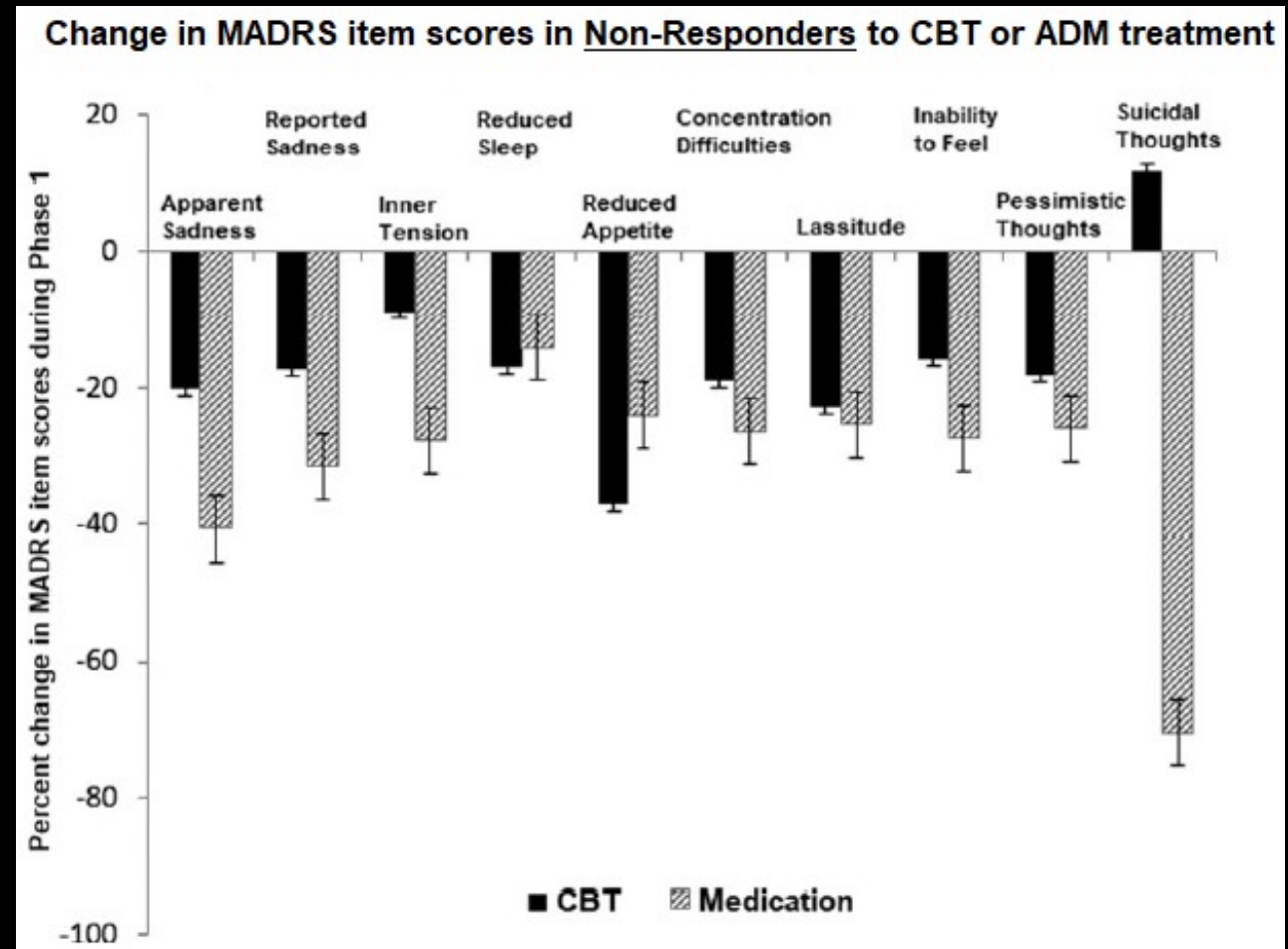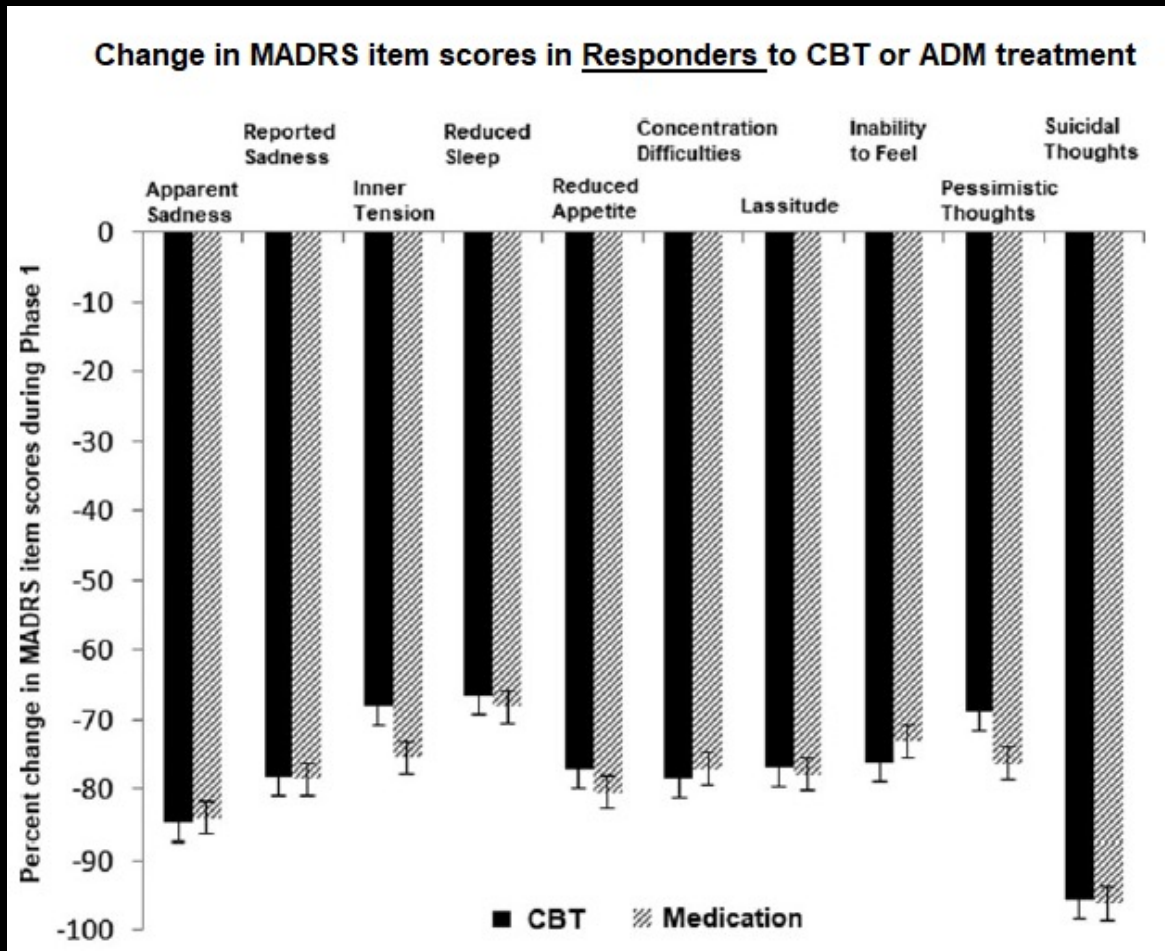# HAMD Item score change in CBT vs ADM treated MDD patients

N=17 RCTs of ADM vs CBT

N=1,856 patients

Note: These were all patients willing to be randomized to CBT or ADM, which may impact generalizability

Boschloo et al, World Psychiatry, 2019;18:183–191

ADM vs CBT HDRS Item Effect Sizes
Cohen's d

Symptom-level effects of CBT vs ADM in Responders vs Non-Responders

Polychroniou, et al., Psychol Med, 2018
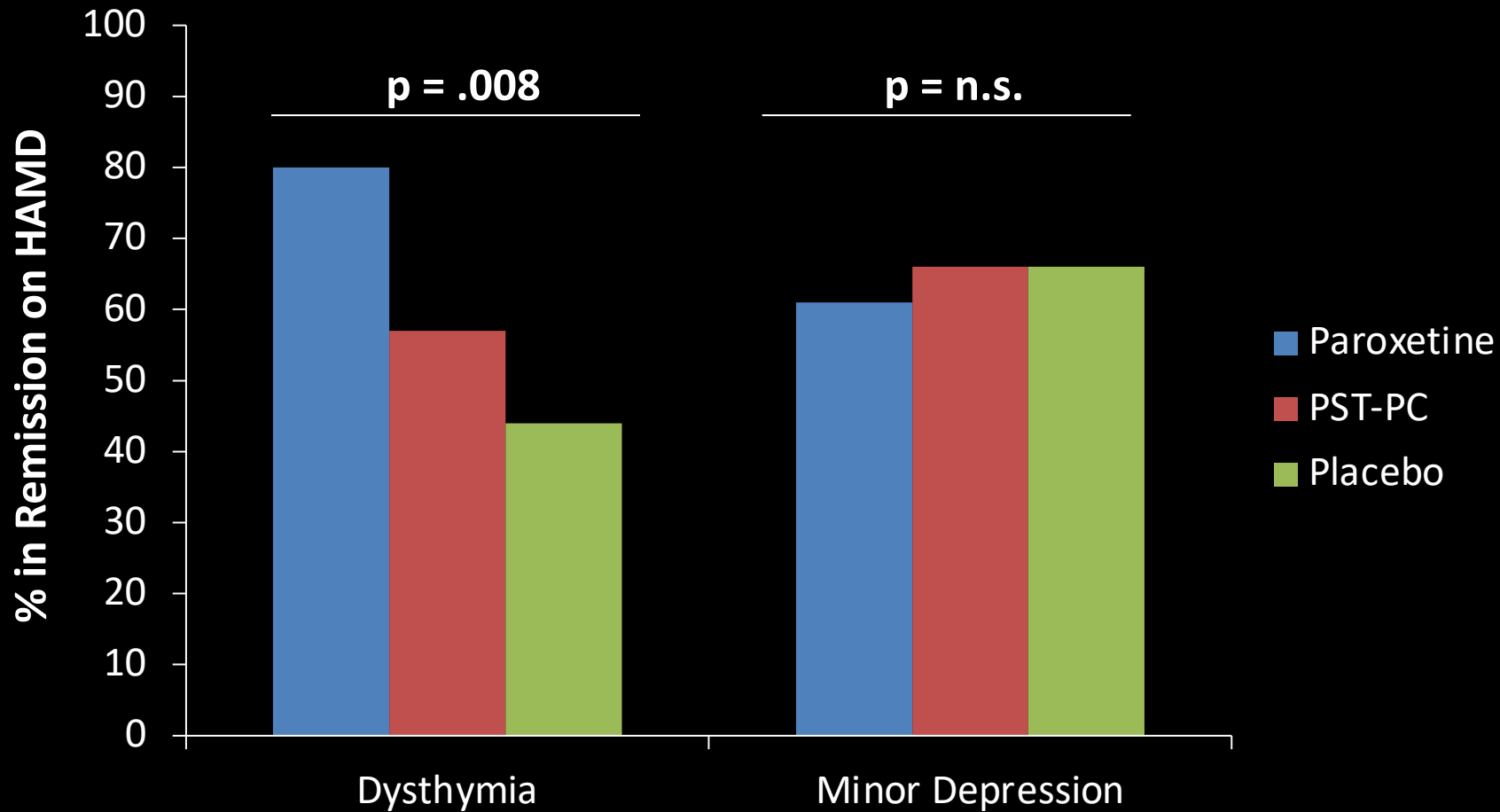
Cochrane Collaboration Review:
Medication vs Placebo for Dysthymia

OUTCOME	N Trials	N Subjects	Risk Ratio	95% CI
All Dysthymia Trials				
Risk Ratio for Response	15	1,992	1.52	(1,37, 1.67)
Risk Ratio for Remission	3	836	1.47	(1.33, 1.64)
Pure Dysthymia or Double Depression *Risk Ratio for Response*				
Pure Dysthymia	7	1009	1.61	(1.35, 1.92)
Double Depression	6	862	1.54	(1.39, 1.69)

- **NNT = 4** for response or remission

de Lima, et al., Cochrane Database of Systematic Reviews, 2005

Medication vs Psychotherapy:
Dysthymia vs Other Psychiatric Disorders

Diagnosis	SMD (95% CI)
Schizophrenia (n = 92)[59,a]	-0.56 (-0.98 to -0.14)
Major depressive disorder, acute (n = 1662)[67]	-0.05 (-0.24 to 0.13)
Major depressive disorder, relapse (n = 231)[64]	-0.71 (-0.40 to 1.01)
Dysthymic disorder (n = 874)[49]	-0.47 (-0.75 to -0.18)
Panic disorder (n = 375)[68]	0.08 (-0.13 to 0.28)
Generalized anxiety disorder (NI)[66]	0.33 (-0.02 to 0.67)
Social phobia (n = 208)[68]	0.15 (-0.12 to 0.43)
Bulimia (n = 237)[65]	0.52 (0.20 to 0.84)

Favors Drug — Favors Psychotherapy

SMD (95% CI)

Huhn et al., JAMA Psychiatry. 2014;71(6):706-715

Phases of Treatment for Depression

Kupfer DJ. *J Clin Psychiatry*. 1991;52(suppl 5):28-34
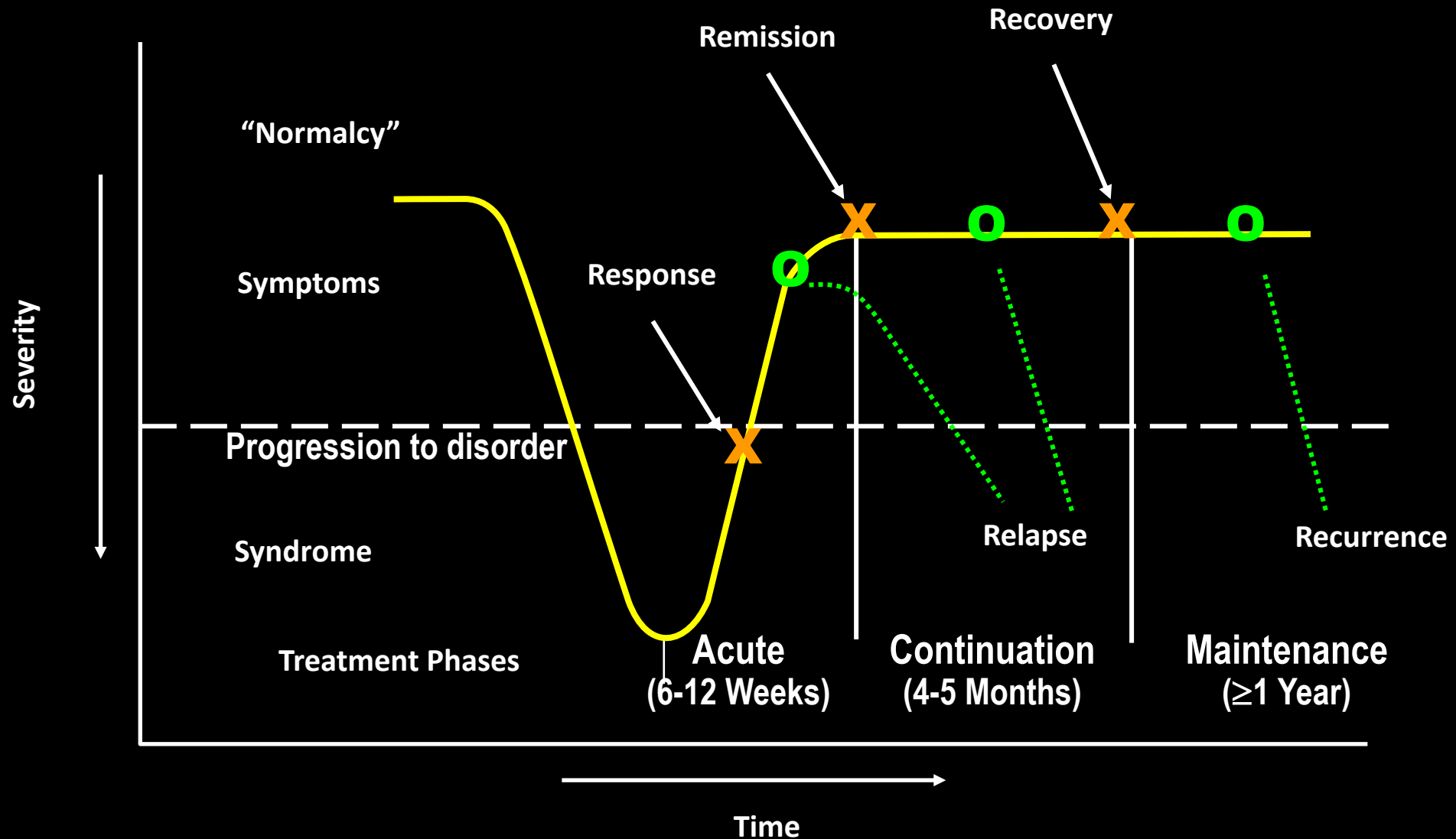
Antidepressant Prevention vs Relapse

Study G: A representative study
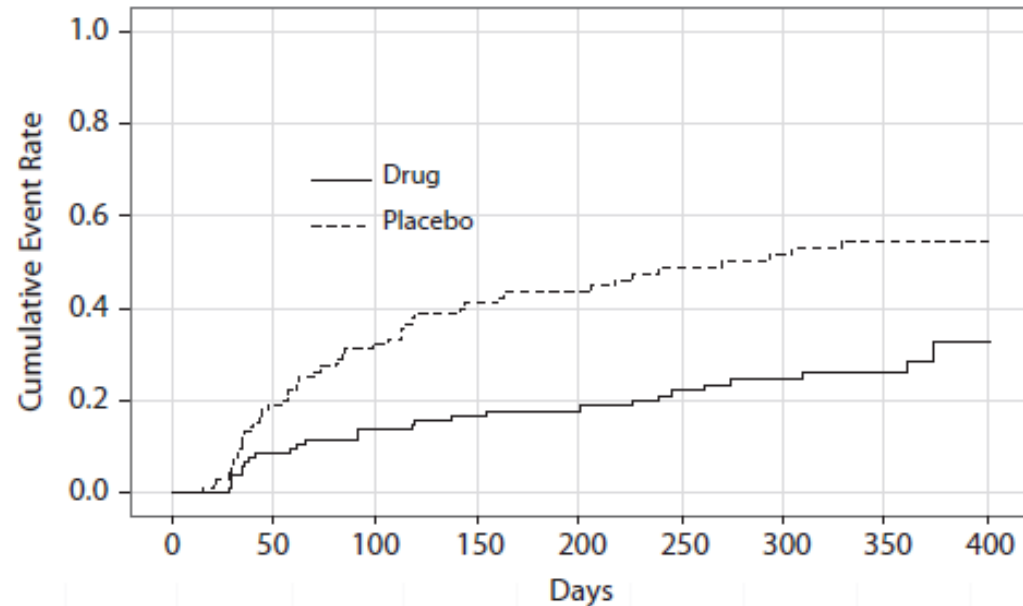
Figure 3. Relapse Rate Differences Between Drug and Placebo Arms at Each Double-Blind Phase Time Point in Antidepressant Maintenance Studies[a,b]
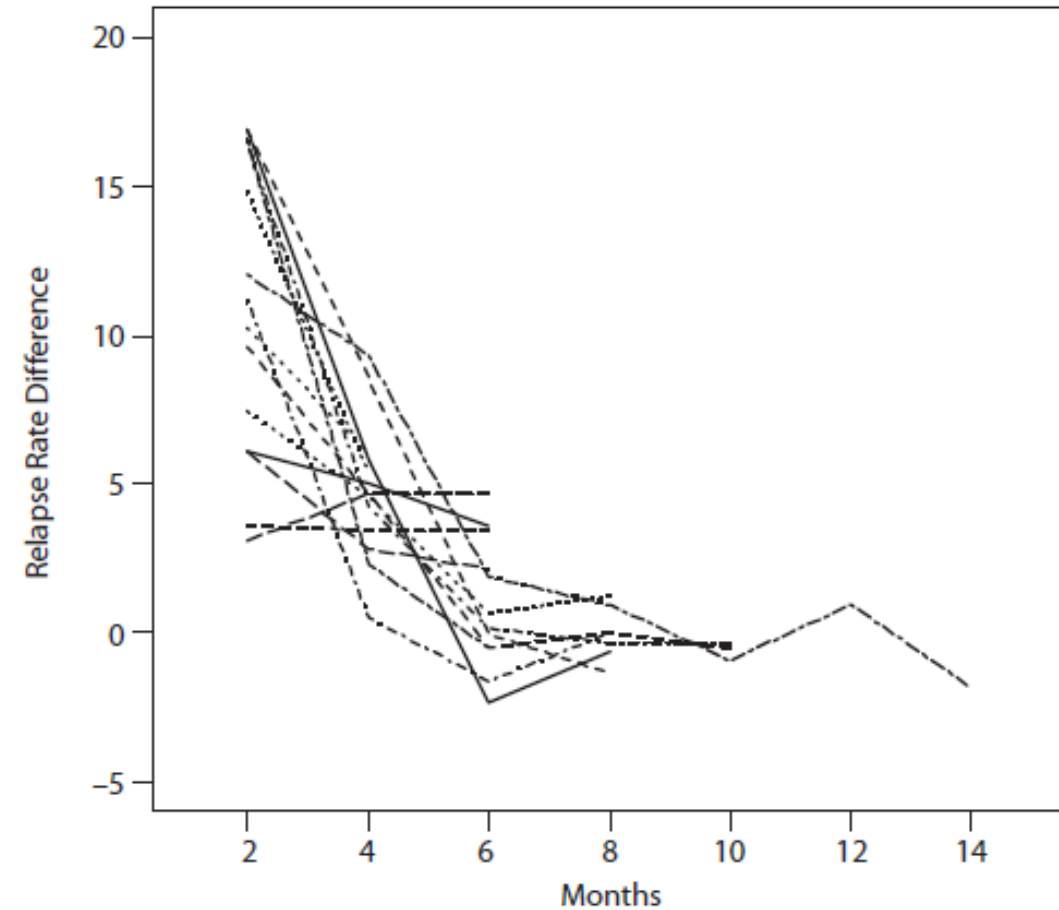
[a]Each curve represents a study.

FDA review of 15 Double-blind discontinuation trials with at least 6 months' follow-up

Borges et al., J. Clin. Psychiatry, 2014; 75:205-214   23

Considerations for Placebo Effects

Components of Placebo Response

Placebo Effect   (effects of PBO pill + trial interactions)

Natural Course   (spontaneous recovery/worsening)

\+   Regression to Mean

_____

Placebo Response

Response to Placebo is Increasing

Walsh et al, JAMA, 2002; 287:1840 -7
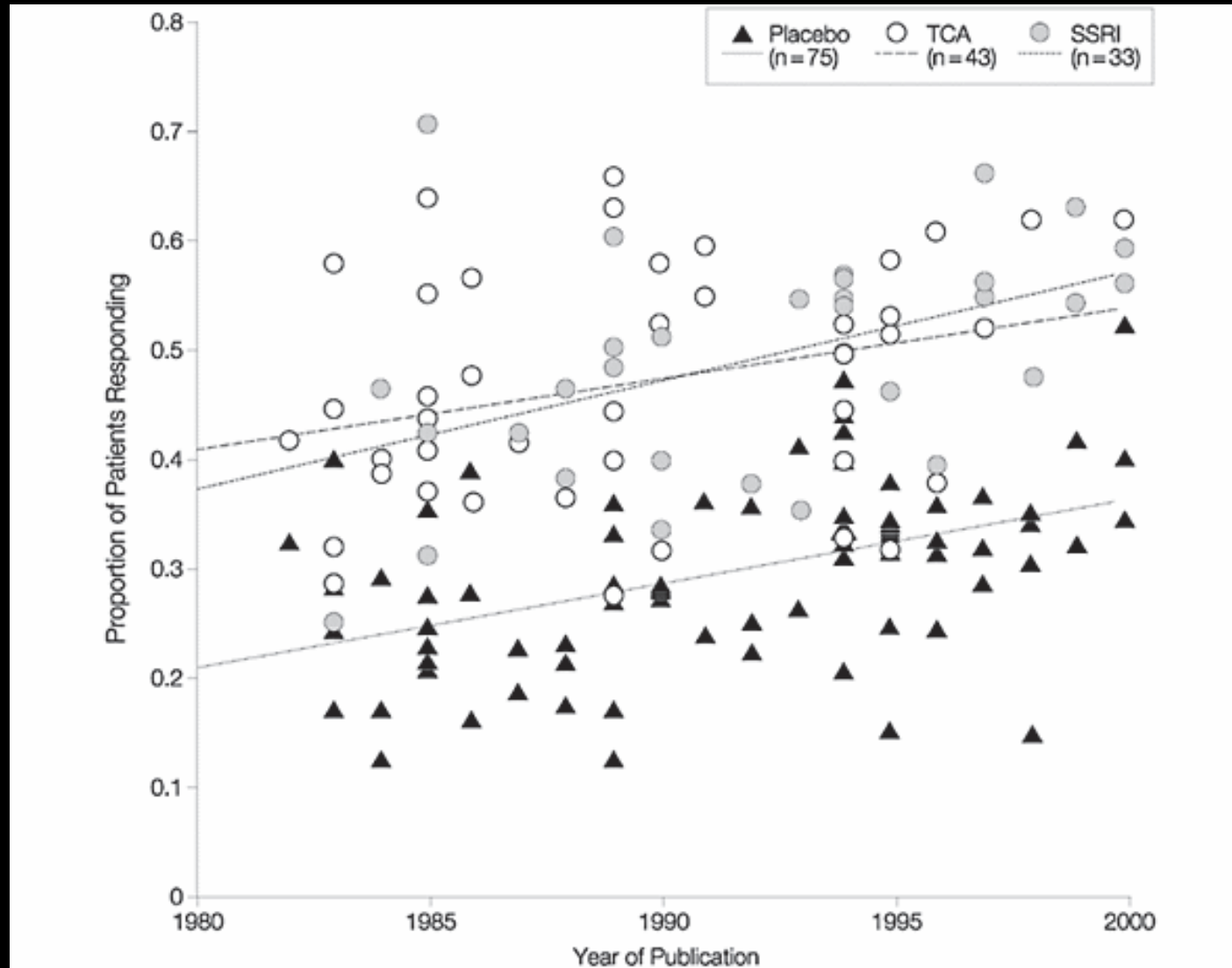
Increasing Placebo Response is Reducing Effect Sizes in Trials

• Meta-analysis:
30 Wyeth/Pfizer trials of Venlafaxine or Desvenlafaxine

•Very similar trial designs
   • 6-12 weeks
   • 33 or 50% PBO risk

◆ Standardized difference in means (Placebo only)
■ Standardized difference in means (Drug vs placebo)
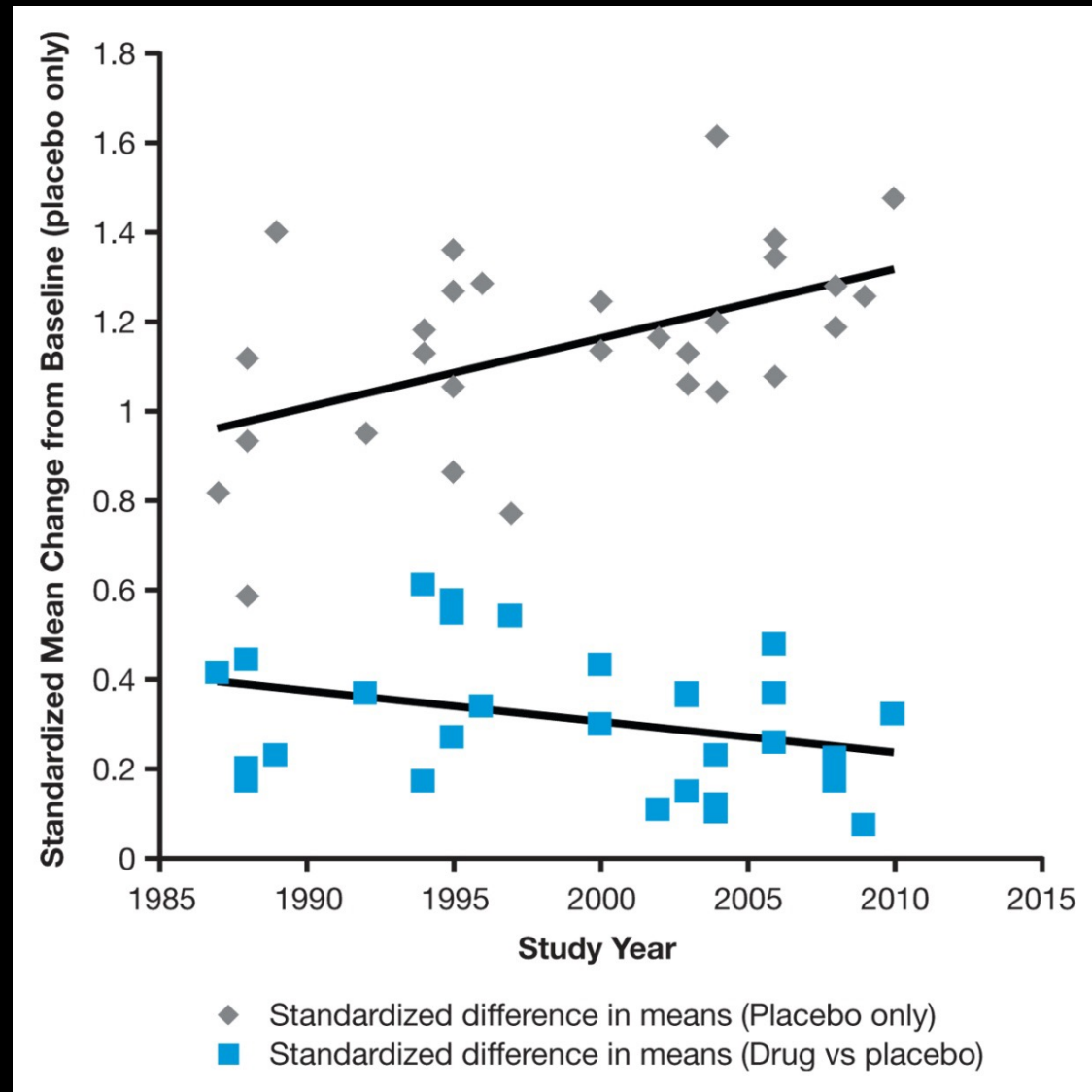
Dunlop et al, Neuropsychopharmacol, 2012

Drug-PBO Differences Across Classes & Time

Measures	All drugs	TCAs	MAO inhibitors	SRIs	SNRIs	Atypicals	Early (1983–1997)	Late (1998–2010)
Trials (*n*):	124	31	5	47	30	11	57	67
Responder Rate Difference								
Pooled RD	16.30%	21.40%	12.10%	14.60%	16.40%	11.90%	20.70%	13.40%
Improvement Relative Difference (% Difference in Rating Scale score)								
Pooled RD	12.50%	16.20%	16.00%	11.50%	9.80%	12.80%	16.80%	9.80%
Number Needed to Treat								
	8	6.2	6.2	8.7	10.2	7.8	6	10.2
95%CI	7.1–9.1	5.2–7.5	3.0–102	7.0–11.5	8.0–14.0	5.7–12.2	5.2–6.9	9.8–13.9

WHO Studies WHOM and HOW?
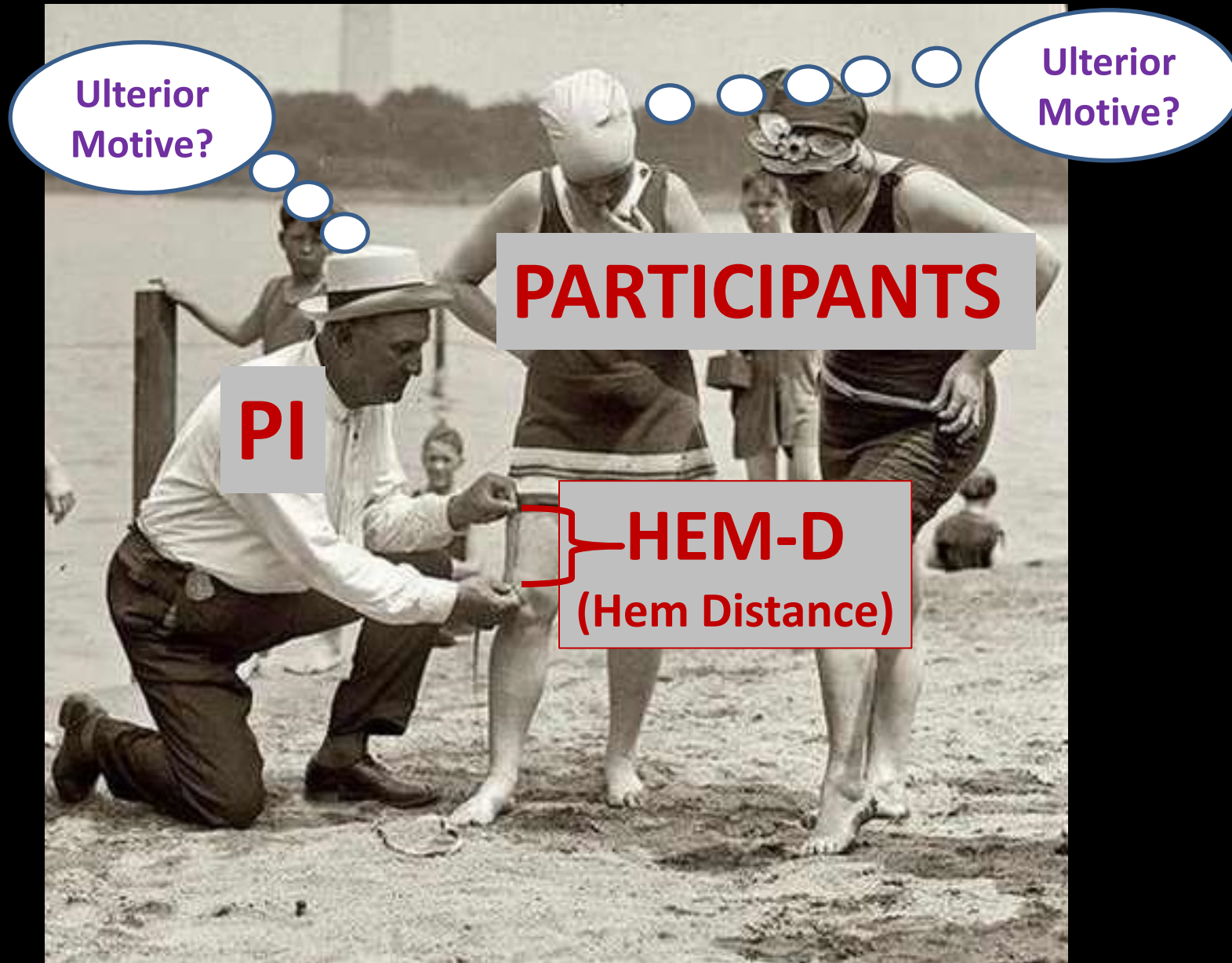
Add-on Sites to Meet Recruitment Targets

Recruitment Over Time
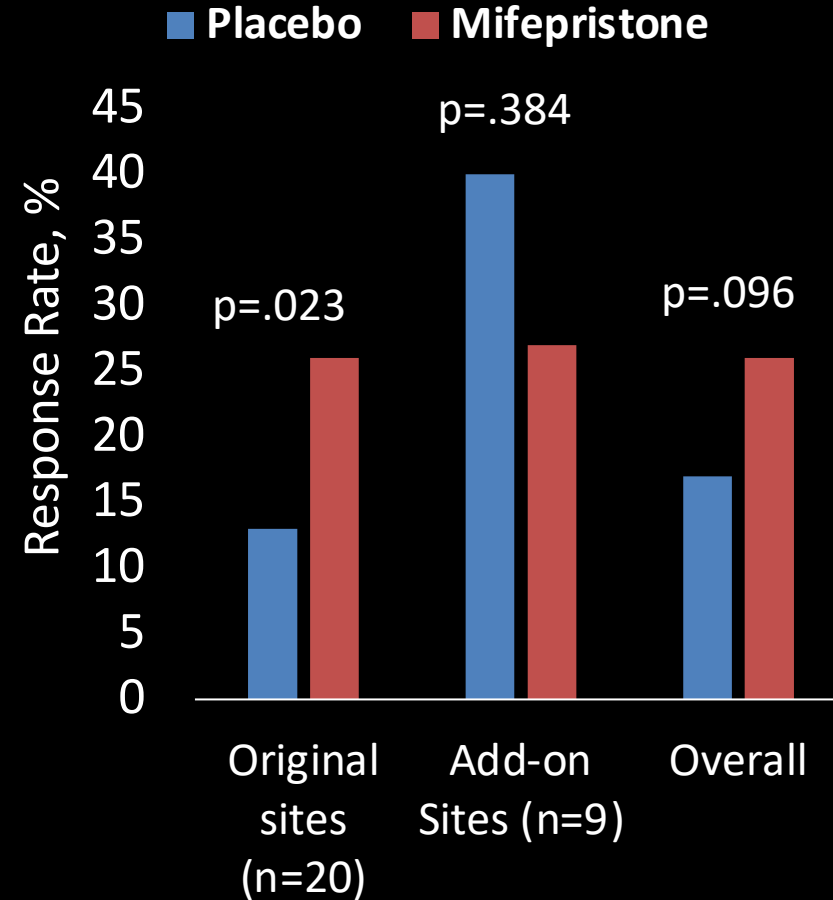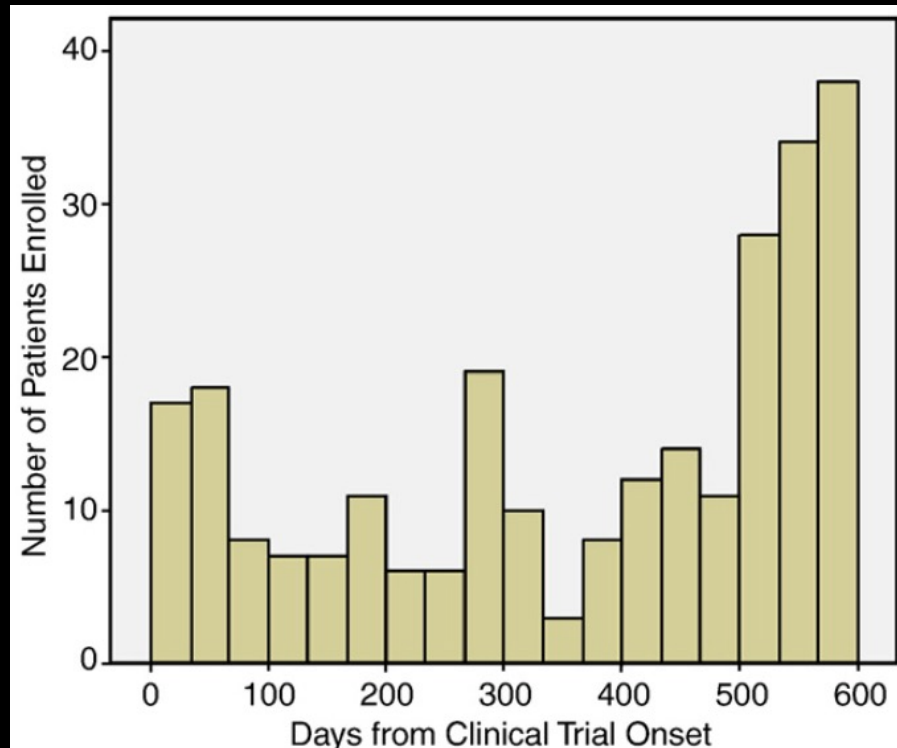
Mifepristone for MDD with psychotic features

Blasey et al, Contemp Clin Trials 2009; 30:284-88

Regression to the Mean

- Reduction of severity score resulting from:
    1. Reduction in distress
        - Taking action; Education; Instillation of hope
        - May be particularly salient for patients recruited from advertising vs MD's own treatment clinic
    2. Removal of incentive bias of researcher once subject is randomized
        - Selective score inflation

Selective Score Inflation and Regression to the Mean
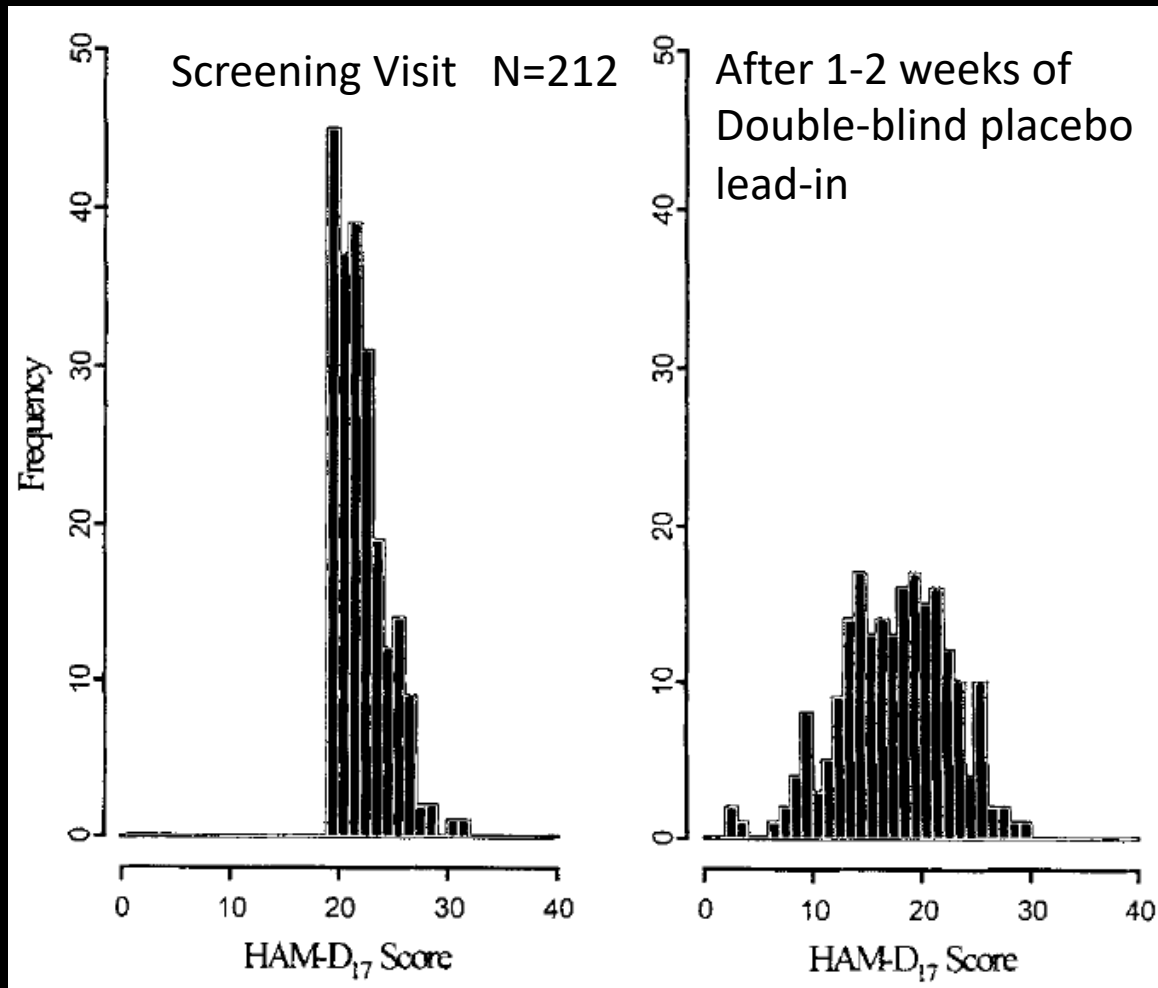
PBO Change	Drug Change	Treatment Difference
-5.09	-5.07	-.02

≥ 10% score inflation, N=136
PBO chng: -5.63
Drug chnge: -3.83
Trmt Diff: -1.80   p=.159

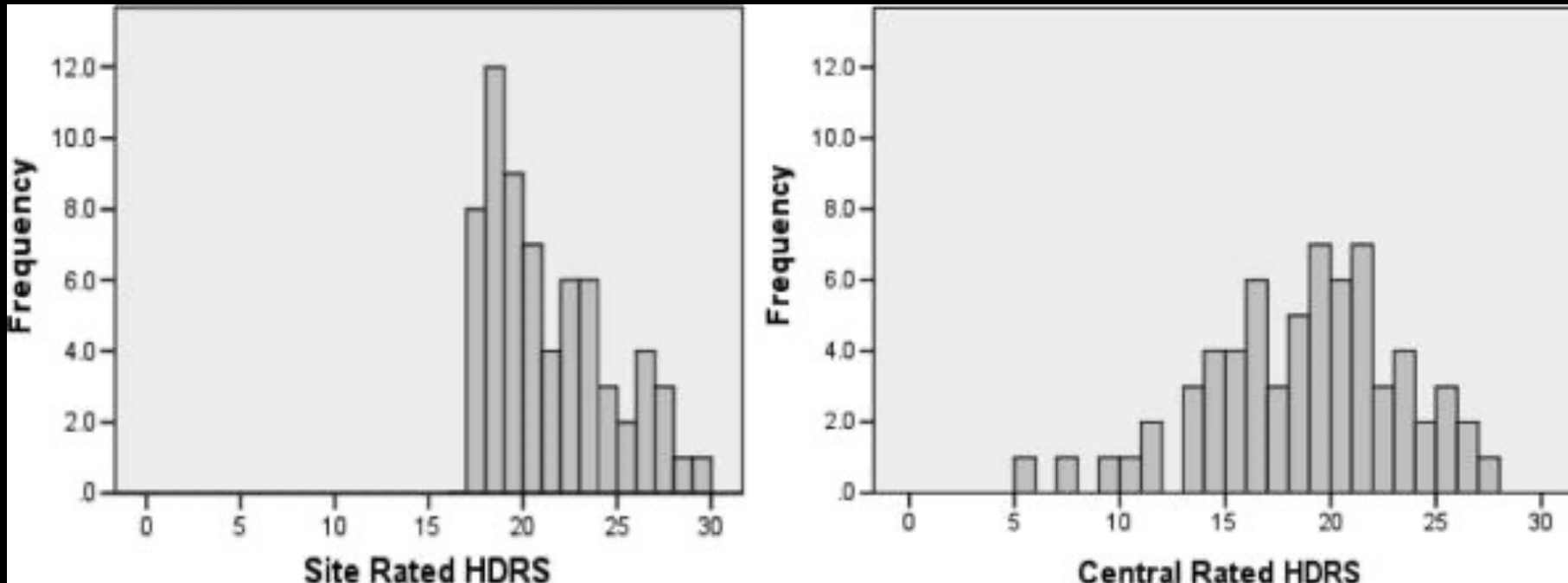**<10% score inflation,** N=76
PBO chng: -4.23
Drug chnge: -7.39
Trmt Diff: +3.16   p=.078

Landin et al, Biometrics, 2000; 56:271-8

Centralized Raters may improve severity measurement validity

Baseline Depression Severity Ratings

- HAMD (HDRS) ratings twice at BL visit: 1)Site and 2) Videoconference with Central Raters
- N=81 MDD patients
- Recruited from UCLA and MGH

Kobak et al, J Clin Psychopharmacol, 2010; 30 (2): 193-197

Clinical Trial Expectancy Curve
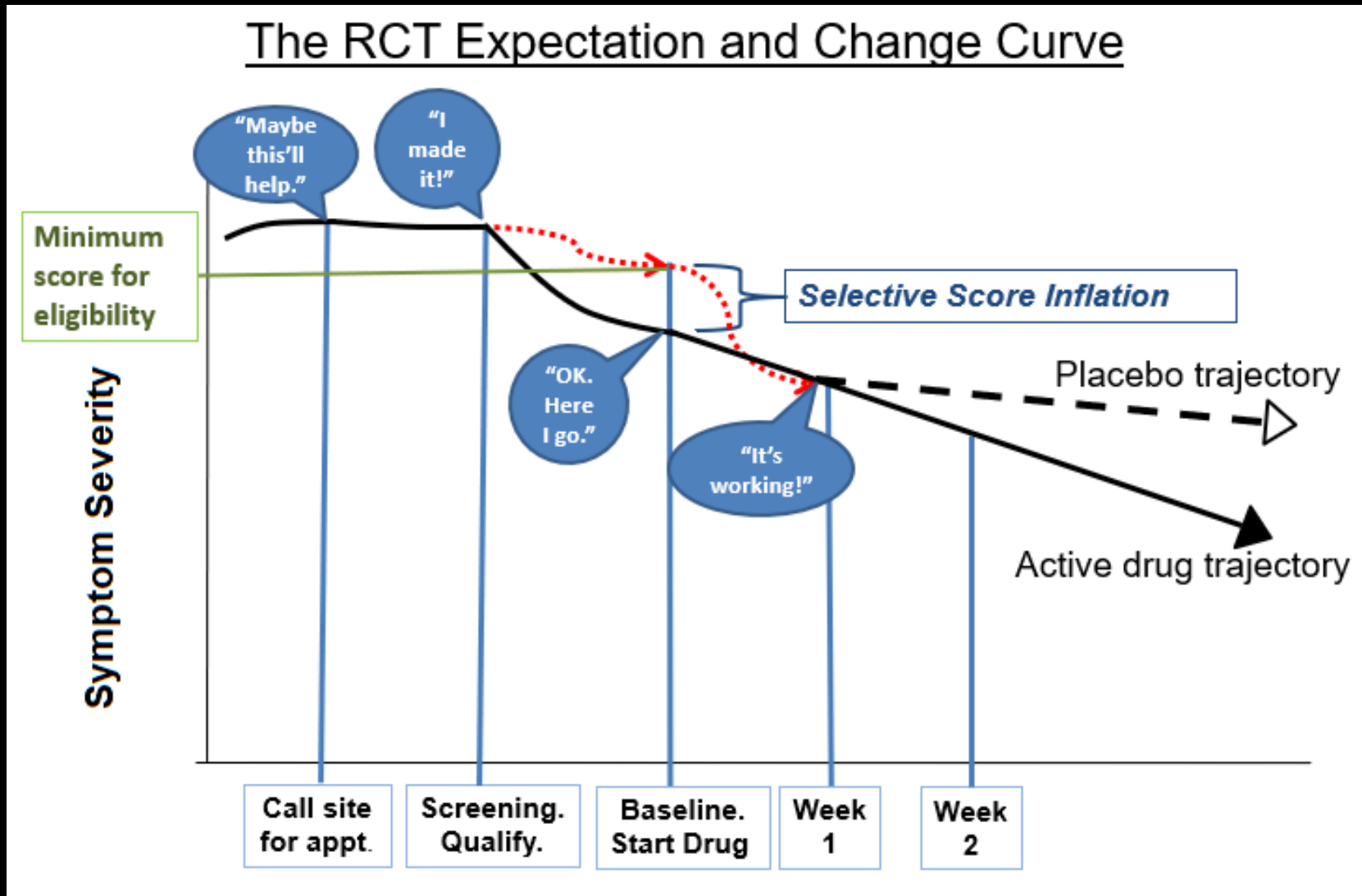
Higher Minimum Severity Entry Criterion does not Improve Signal Detection

- Hierarchical linear regression on HAMD change scores
- 51 Trials in FDA database (11,270 patients)
- $R^2 = .29$; Adjusted $R^2 = .20$; F Change: 4.36

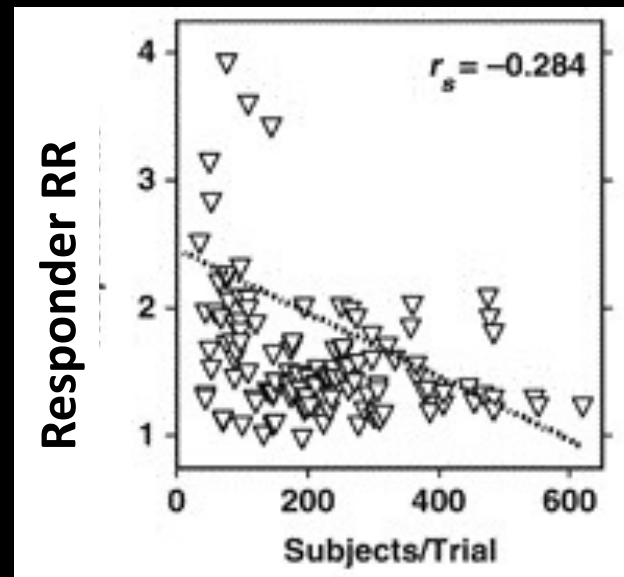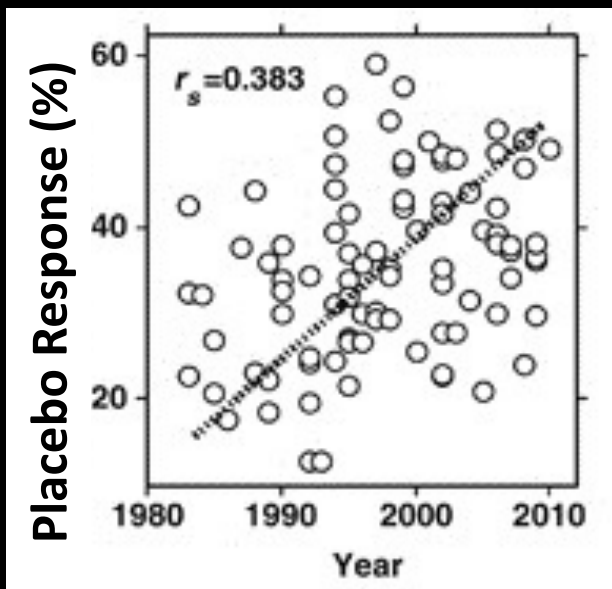| Model	B	SE B	β	p
Trial Length	.24	.22	.17	
Dosing (Flexible vs Fixed)	1.54	.55	.35	<.01
N – Drug	-.004	.02	-.06	
N – Placebo	.01	.02	.12	
Entry Criterion (BL HAMD score)	-.07	.18	-.06	
Pre-Rand HAMD (PBO)	-.61	.37	-.77	
Pre-Rand HAMD (Drug)	.87	.36	1.10	<.05

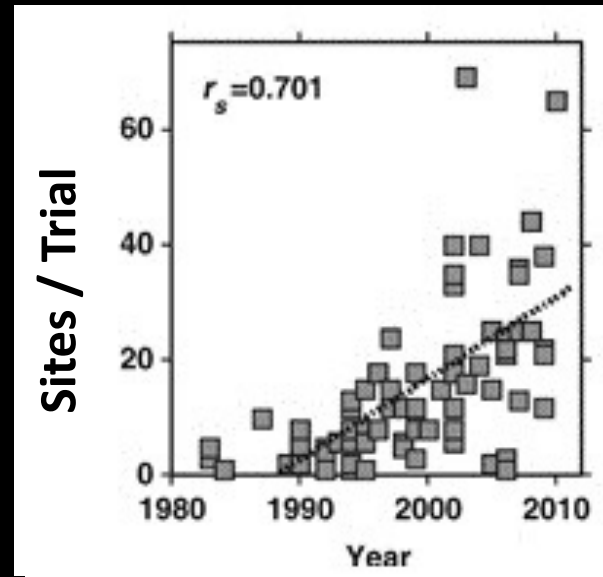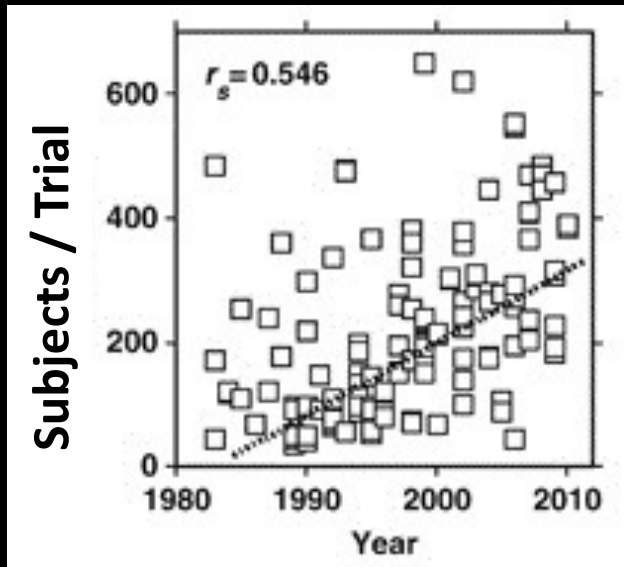Khan et al, Biol Psychiatry, 2007; 62:65-71
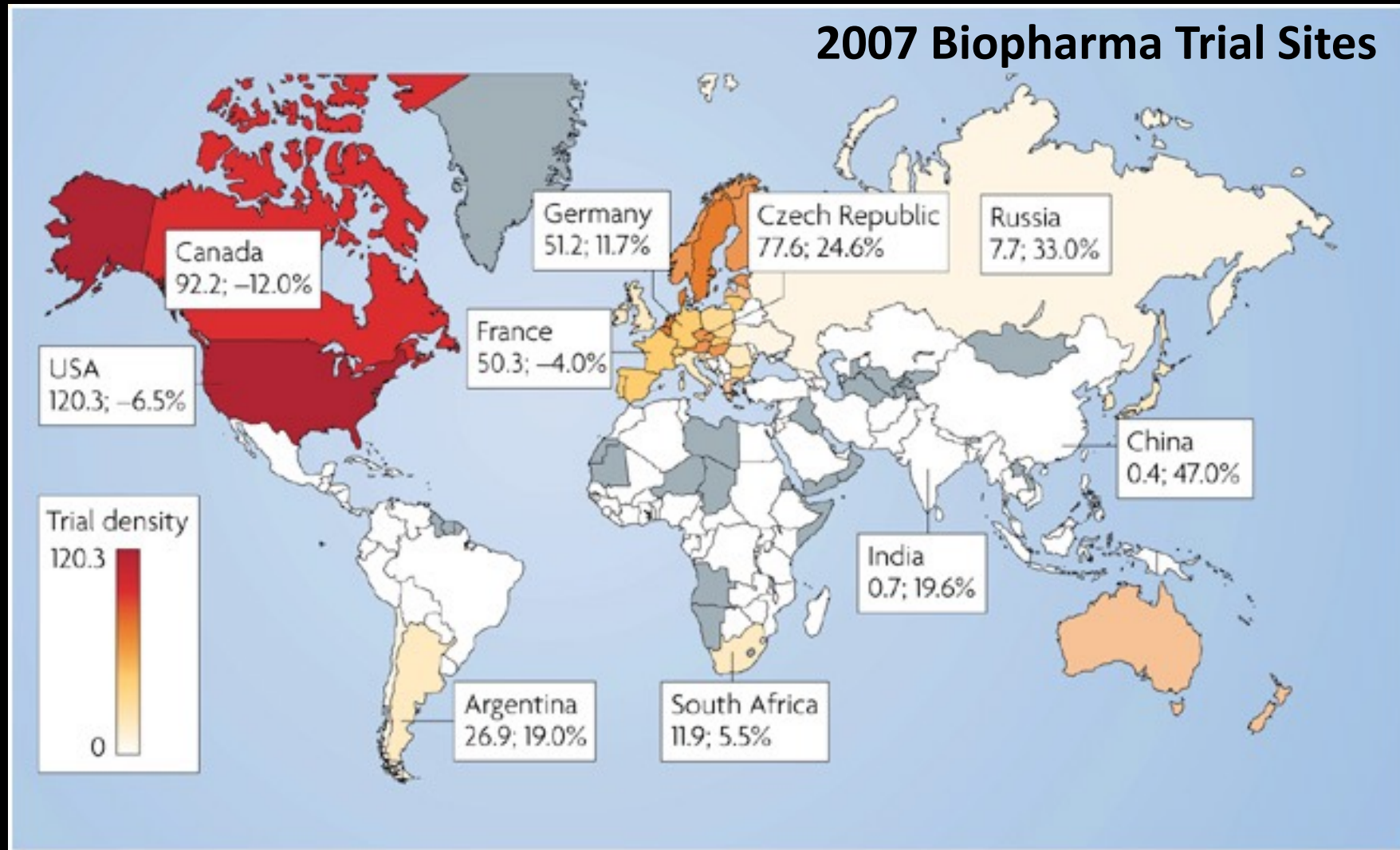
Later Patients Contribute Less
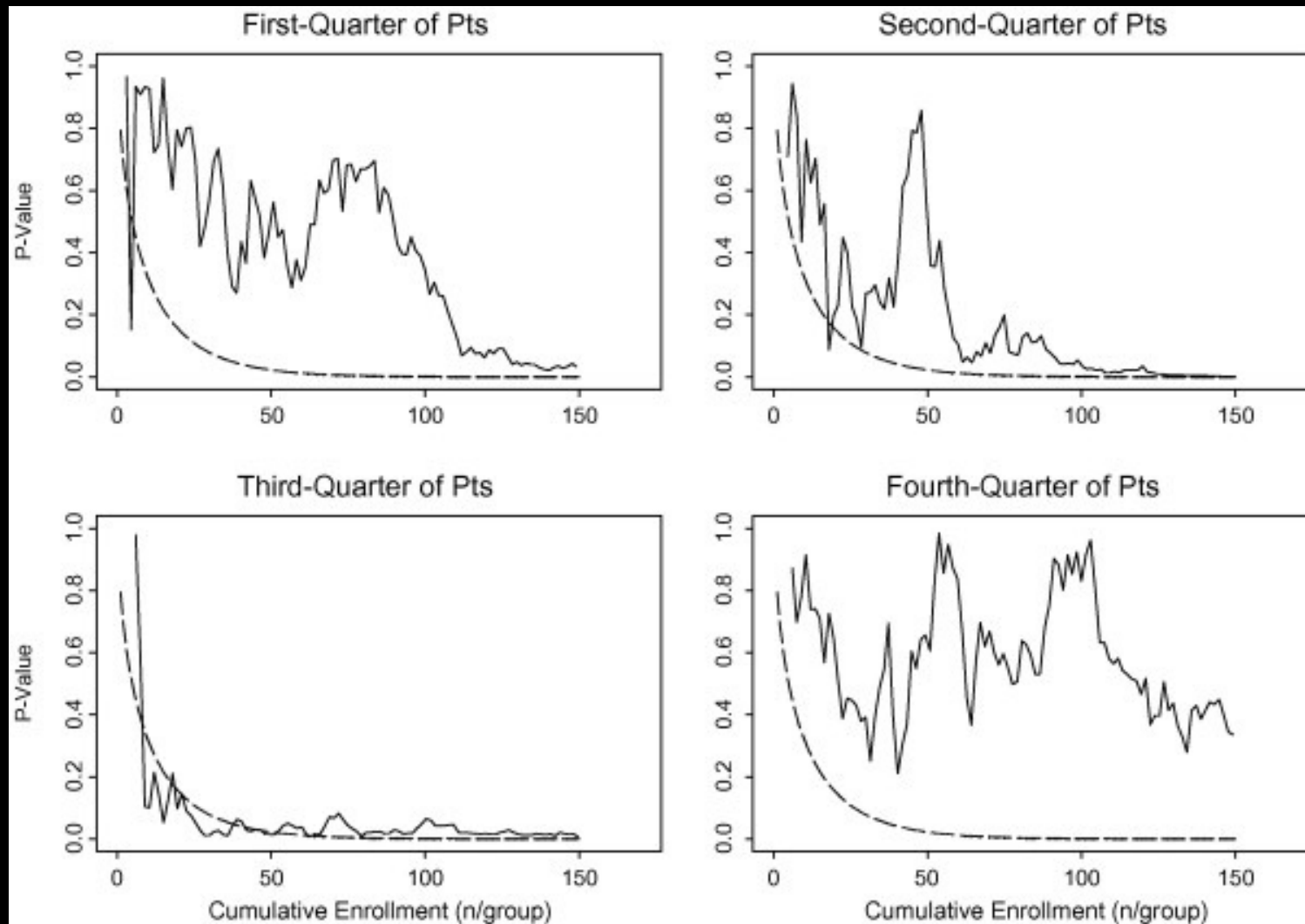
Pooled data from 4 trials of Paroxetine vs Placebo
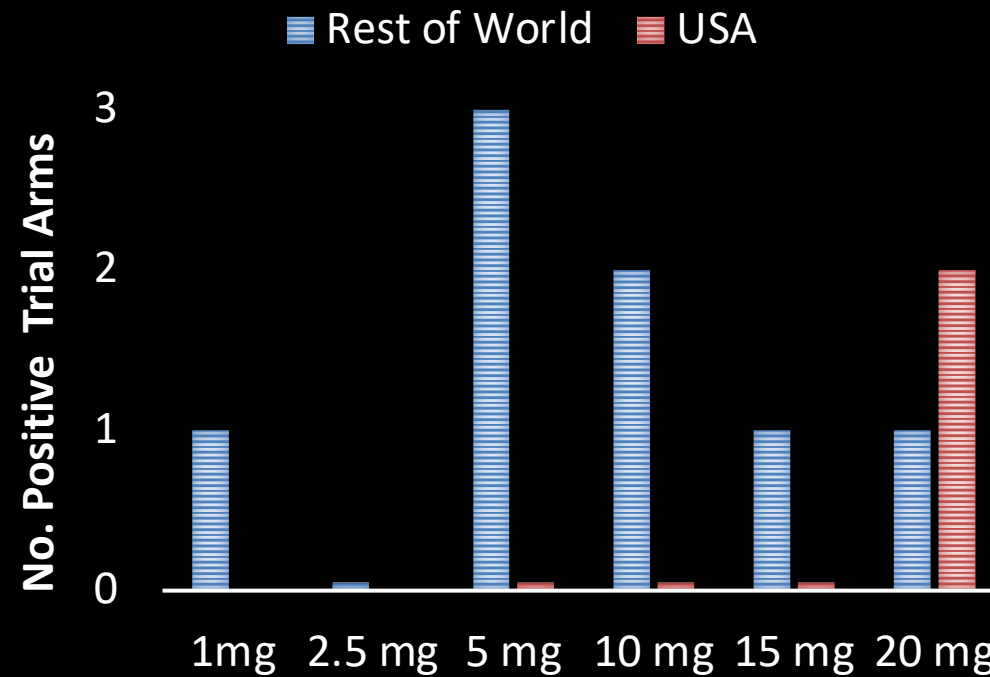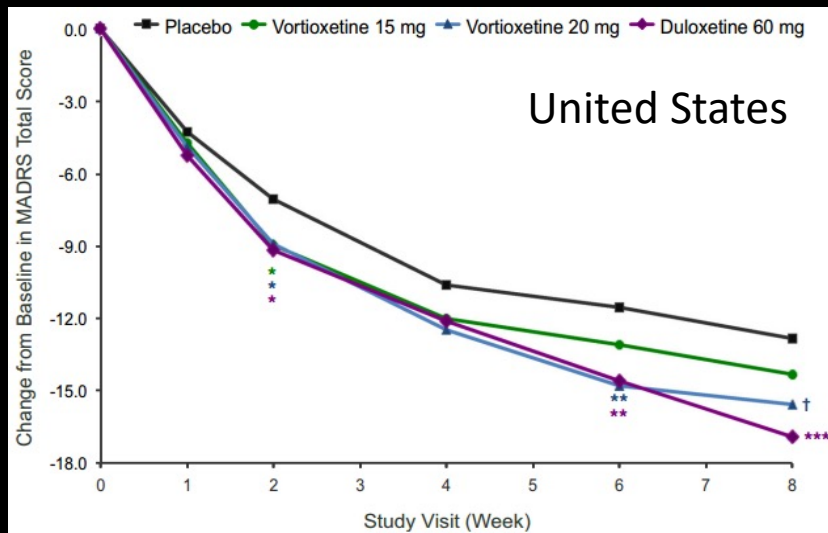
Liu et al, J Psychiatr Res, 2008: 42:622-30

HOW: Larger Trials are not a Solution

WHO: Investigators Everywhere!

2007 Biopharma Trial Sites

Germany 51.2; 11.7%
Czech Republic 77.6; 24.6%
Russia 7.7; 33.0%
Canada 92.2; −12.0%
France 50.3; −4.0%
USA 120.3; −6.5%
China 0.4; 47.0%
India 0.7; 19.6%
Trial density 120.3 ... 0
Argentina 26.9; 19.0%
South Africa 11.9; 5.5%

Vortioxetine Doses in USA vs Rest of World

Europe/Asia/Africa

United States

Rest of World   USA

1mg   2.5 mg   5 mg   10 mg   15 mg   20 mg

No. Positive Trial Arms

Henigsberg et al, J Clin Psychiatry, 2012; 73:953-9
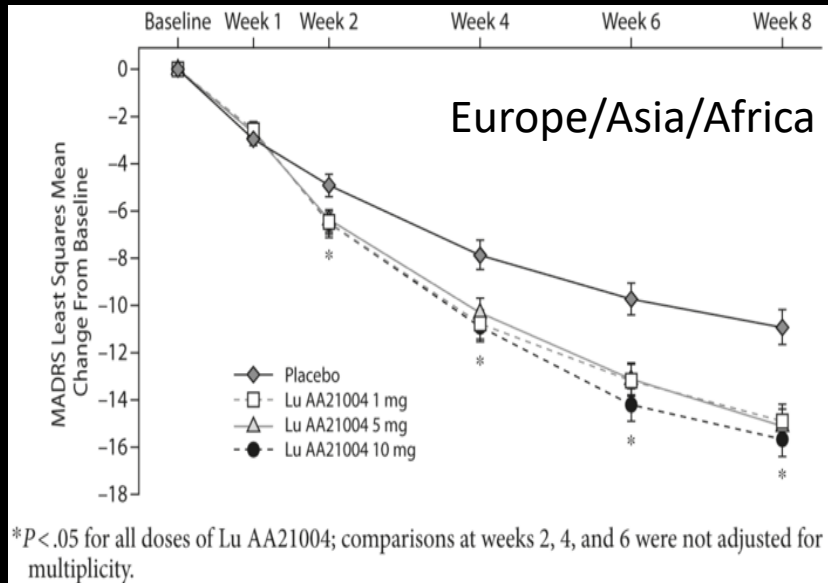
Mahableshwarkar et al, Psychopharmacol 2015; 232:2061-70

Zhang et al, J Clin Psychiatry, 2015; 76:8-14

Studying Professional Study Subjects

Survey for experienced research subjects

- Have you enrolled in more than one study in the past year?
- Have you been in more than three studies in the past three years?

If you answered yes to either of these questions you qualify for the Experienced subject survey.

- Participation involves a one-time interview lasting 60 minutes
- Qualified subjects reimbursed for their time

Call 888-552-5264 and ask for "The experienced subjects study"

100 Subjects participated in ≥2 studies in past year

Boston-based

Annual Income: 57% <$30,000/yr
Mean # Trials: 12   (Range 2-100)
Mean Earnings: $9809   (Range $50 -175K)

51% enrolled in trial for condition they actually had

35% had done Psychotherapy trial

36% Shared information about trial with others

TYPES OF DECEPTION
43% Enrolled in another trial
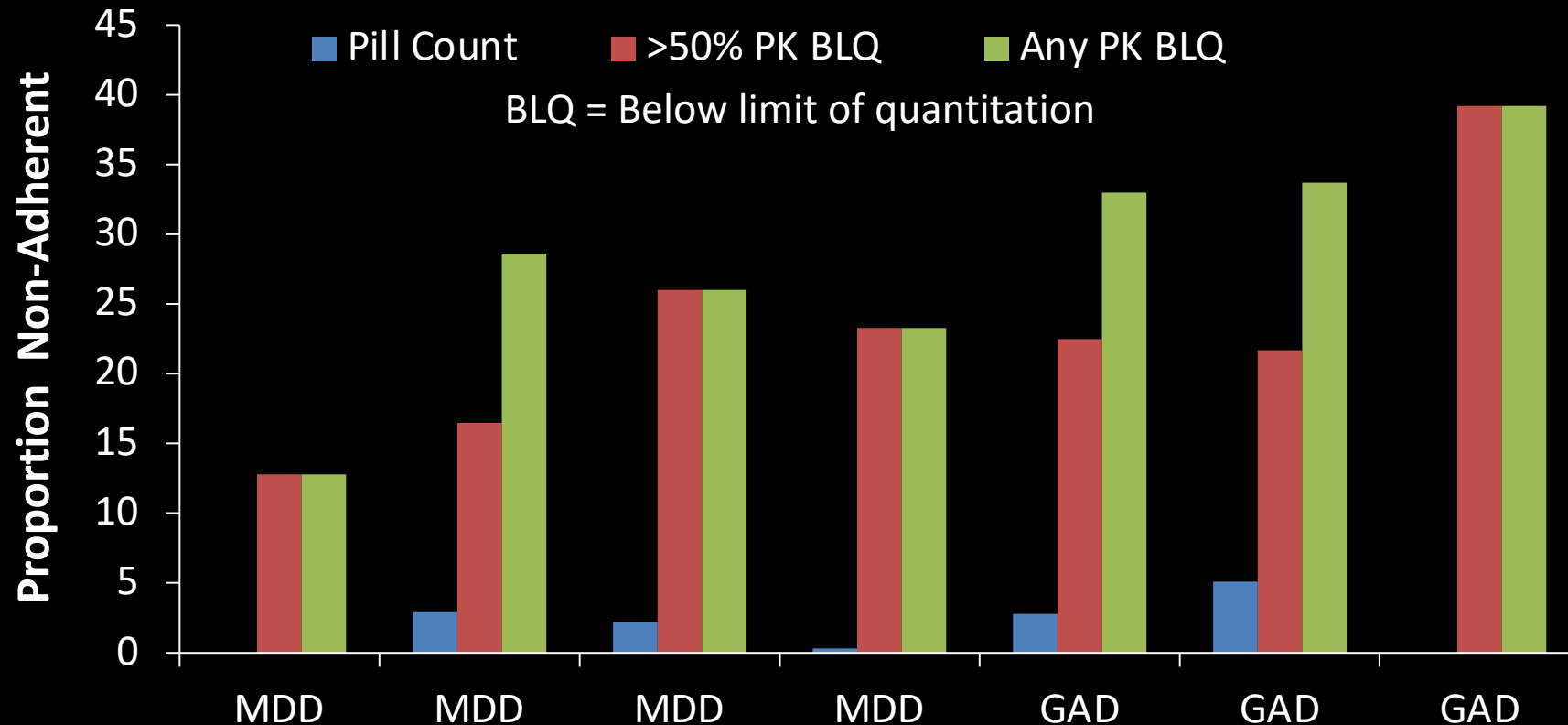32% Hid health problems
28% Hid Concomitant medication
25% Exaggerate symptoms
20% Hid Illicit drug use
17% Hid mental health problem

Devine, et al, Clinical Trials 2013; 10: 935–948

Non-Adherence Detection by Method

7 AstraZeneca trials of MDD or GAD (N=1765 on drug), 2001-2011

- Mean rate Non-Adherence by >50% PK BLQ: **23%!**
- Correlation between N subjects and
  Non-Adherence: **r=0.68, p=.06**

McCann et al, J Clin
Psychopharmacol,
2015; 35:566-73

Why is Signal Detection Decreasing?

- Reduced stigma about depression treatment
  - Patients present earlier in course of illness – more responsive to interventions and less chronic
- Social belief that antidepressants "work"
  - Greater expectancy effects.
- Lower quality of trial conduct
  - Transition of research from academia to for-profit sites
  - Professional patients
  - Recycling patients across trials

All adds up to greater PLACEBO RESPONSE

Summary of Signal Detection Challenges

Rutherford & Roose, Am J Psychiatry 2013

43