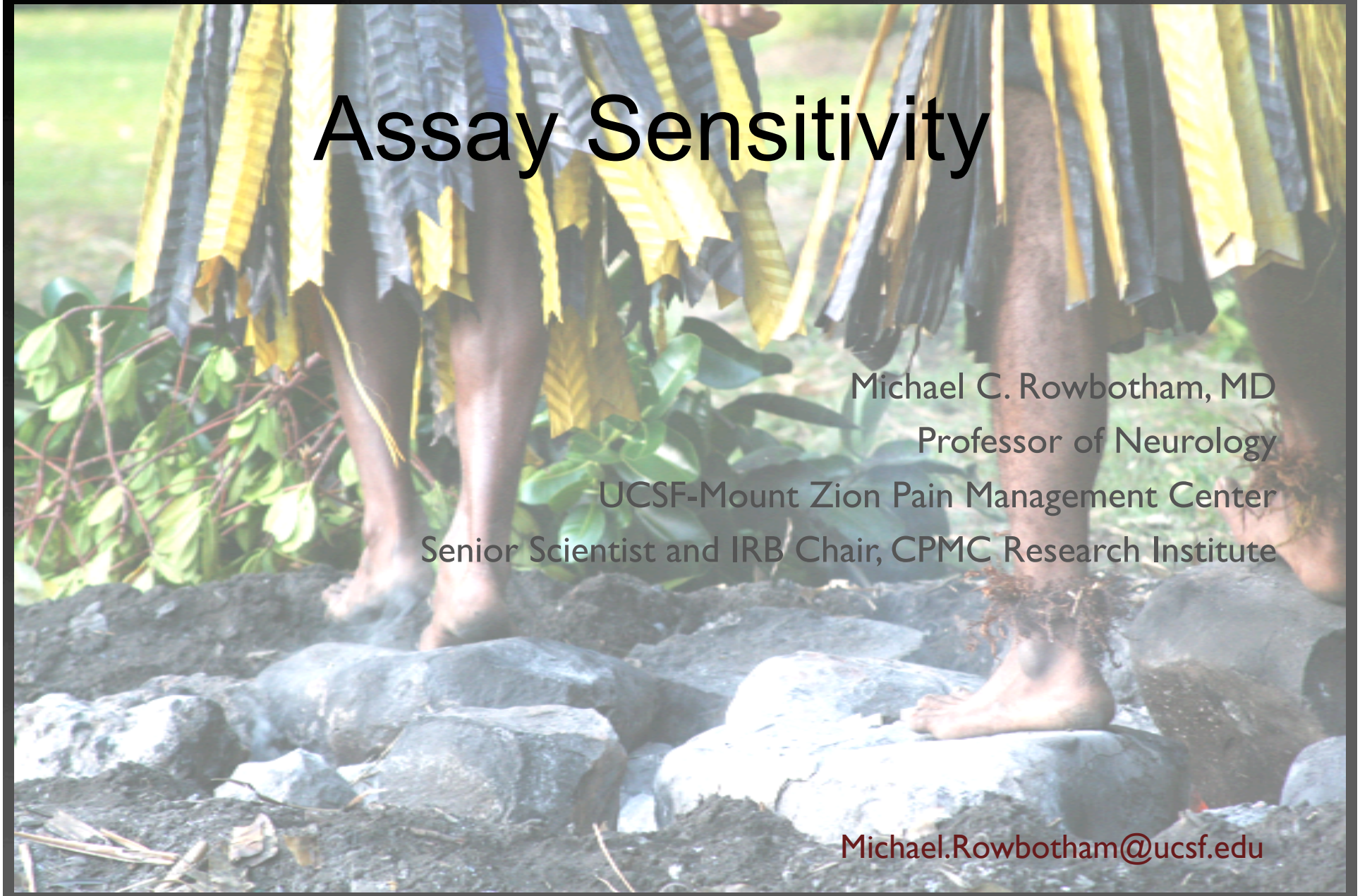
Assay Sensitivity



Michael C. Rowbotham, MD
Professor of Neurology
UCSF-Mount Zion Pain Management Center
Senior Scientist and IRB Chair, CPMC Research Institute

Michael.Rowbotham@ucsf.edu

Outline

- What is assay sensitivity and how can it be quantified?
  - False positive and false negative clinical trials
- Is the standardized effect size the best measure?
  - Effect of selective publication
- Can variability and error variance be reduced in a way that increases assay sensitivity?
- Role of sample size
- Comparing drugs and diagnoses
- Role of active comparators
- Use of concomitant analgesics
- Effect of titration
- **E**ffect of clinical trial research designs, including number of treatment arms

Assay Sensitivity

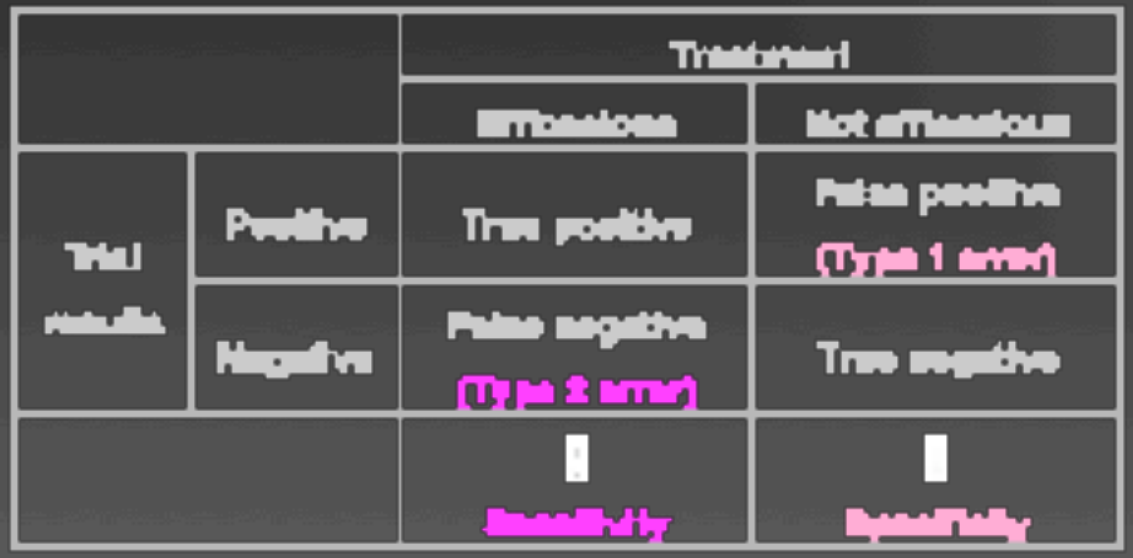
ICH E10 Guidance for Industry on Choice of Control Group and Related Issues in Clinical Trials

- *Assay sensitivity* is a property of a clinical trial defined as the ability to distinguish an effective treatment from a less effective or ineffective treatment
- If a trial intended to demonstrate efficacy by showing *superiority* of a test treatment to control lacks assay sensitivity, it will fail to show that the test treatment is superior and will fail to lead to a conclusion of efficacy
  - False negative outcome
- If a trial is intended to demonstrate efficacy by showing a test treatment to be non-inferior to an active control, but lacks assay sensitivity, the trial may find an ineffective treatment to be non-inferior and could lead to an erroneous conclusion of efficacy
  - False positive outcome

Assay Sensitivity

ICH E10 Guidance for Industry on Choice of Control Group and Related Issues in Clinical Trials

- The question of assay sensitivity, although particularly critical in non-inferiority trials, actually arises in any trial that fails to detect a difference between treatments, including a placebo-controlled trial and a dose-response trial. If a treatment fails to show superiority to placebo, for example, it means either that the treatment was ineffective or that the study as designed and conducted was not capable of distinguishing an effective treatment from placebo



Is the standardized effect size the best measure of assay sensitivity?

- *effect size*
  - Based on a standardized measure of effect (such as *r*, Cohen's *d*, and odds ratio)
  - Standardized effect size measures are typically used when:
    - the metrics of variables being studied do not have intrinsic meaning (could 0-10 pain intensity rating fit this definition?)
    - results from multiple studies (using different scales) are being combined
    - want to convey the size of an effect relative to the variability in the population.
    - Conducting a meta-analysis, as an overall summary

Standardized effect size options

- Cohen's *d* is defined as the difference between two means divided by a standard deviation for the data

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

- Odds ratio (OR) is appropriate for case-control and retrospective studies when comparing binary outcomes, such as proportion achieving 50% pain reduction or proportion achieving outcome of 'mild' or less pain severity
  - (active #success / active #failure) / (control # success / control # failure)
- Relative Risk (RR) is appropriate for binary outcomes in RCTs.
  - (active # success / total active) / (control # success / total control)

SPECIAL ARTICLE

Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy

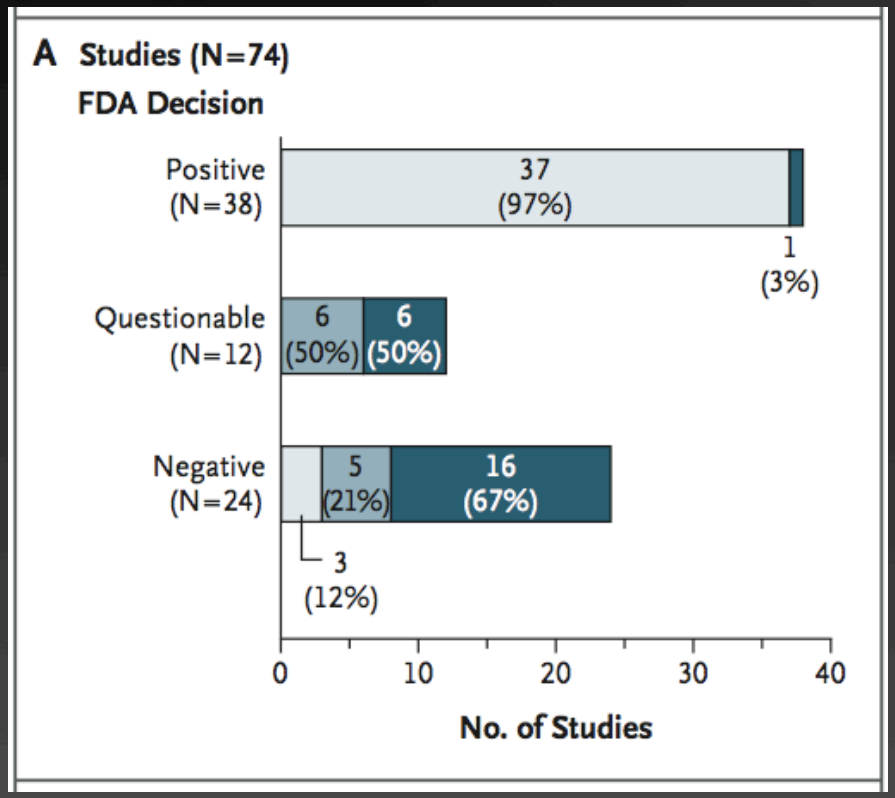
Erick H. Turner, M.D., Annette M. Matthews, M.D., Eftihia Linardatos, B.S., Robert A. Tell, L.C.S.W., and Robert Rosenthal, Ph.D.

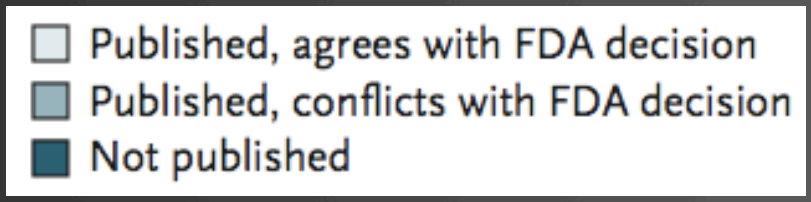
N Engl J Med 2008;358:252-60

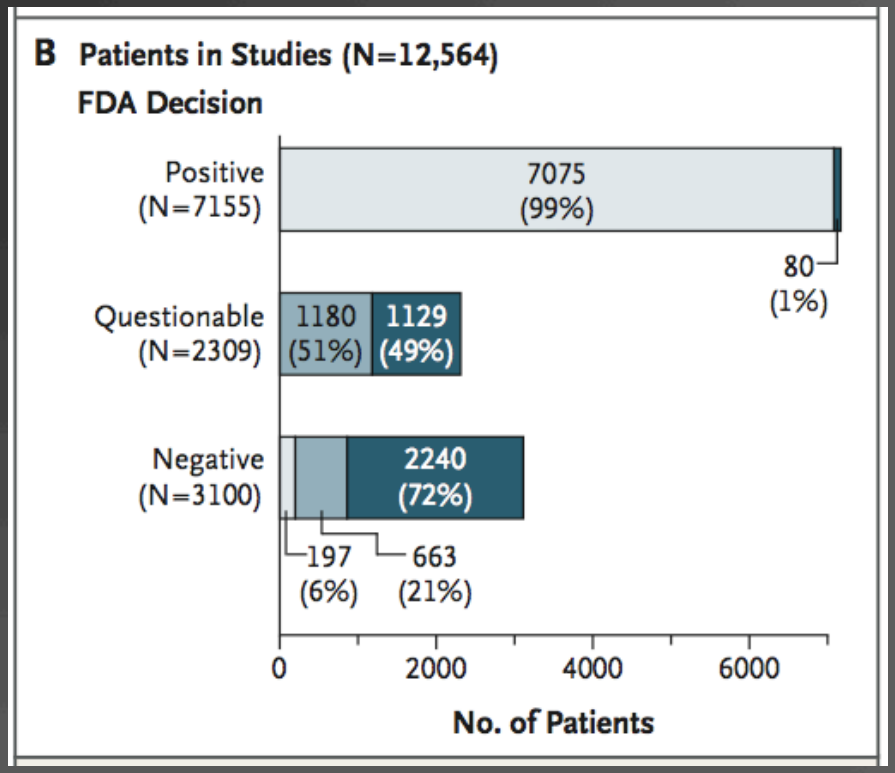
Table 1. Overall Publication Status of FDA-Registered Antidepressant Studies.

Publication Status	No. of Studies (%)	No. of Patients in Studies (%)
Published results agree with FDA decision	40 (54)	7,272 (58)
Published results conflict with FDA decision (published as positive)	11 (15)	1,843 (15)
Results not published	23 (31)	3,449 (27)
Total	74 (100)	12,564 (100)

Evidence-Based Medicine: Selective Publication

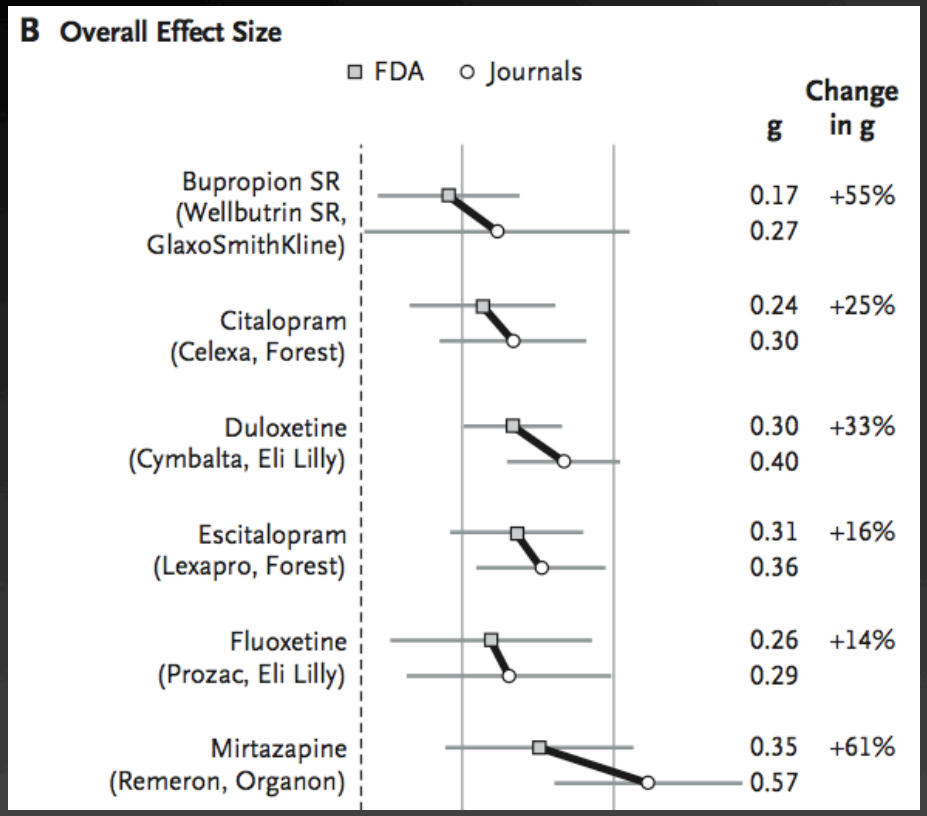


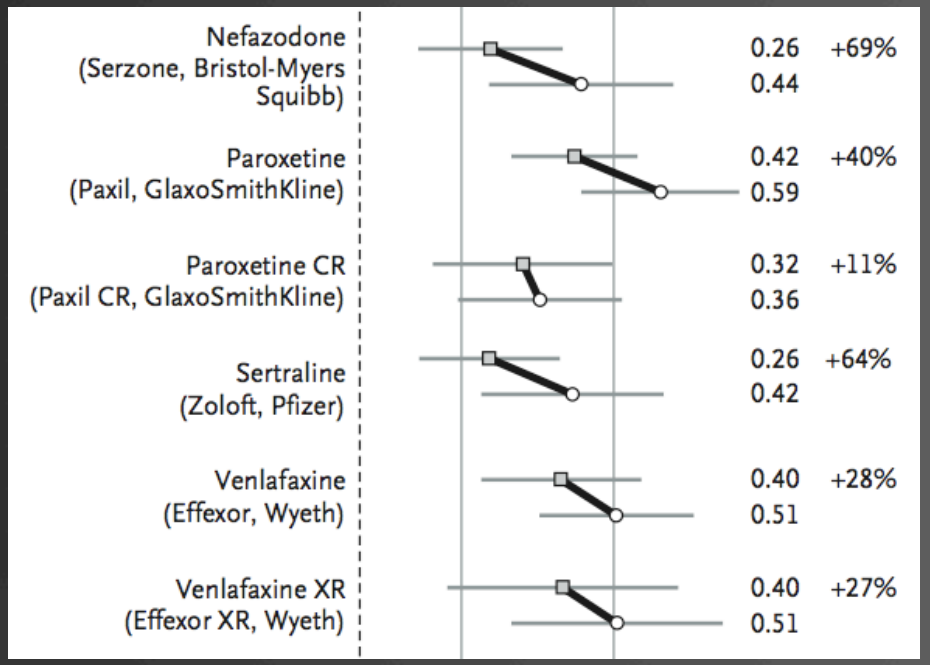




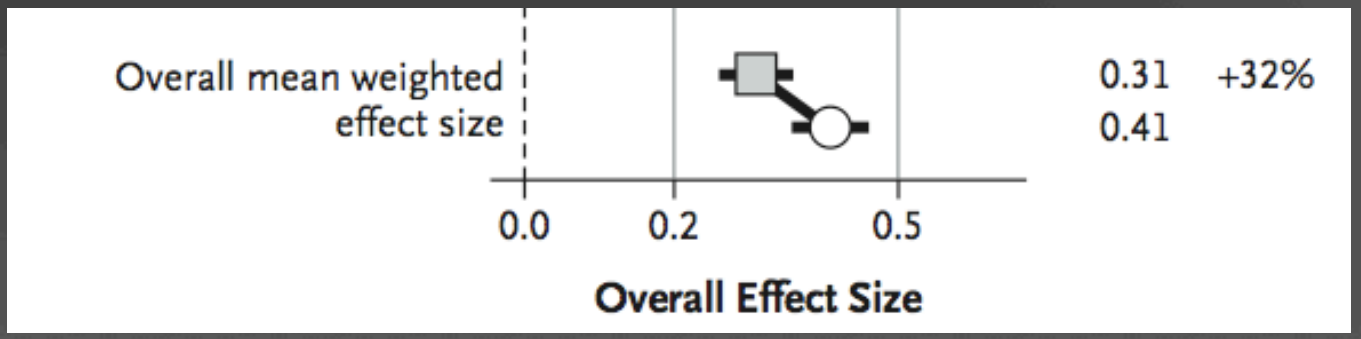
"Not only were positive results more likely to be published, but studies that were not positive, in our opinion, were often published in a way that conveyed a positive outcome."

Selective Publication: Impact on Effect Size





Greater Clinical Effect



FDA Journals

Selective Publication: Does this extend to early phase trials and investigator-initiated studies?

- Phase II and III studies
  - Only after a drug is approved will the FDA publish the SBA containing details of all studies conducted
  - Few compounds progress to Phase III
  - Fewer still result in a successful NDA (<10% for pain)
  - Many ‘negative’ trials, even large ones, go unreported or appear only as posters at meetings - the data becomes unavailable
  - Individual investigators can only independently publish data from their study site
- Investigator-initiated studies
  - Generally smaller and have limited power
  - Journals reluctant to publish negative studies
  - Investigators feel pressure to publish only positive results
  - Failure to prespecify analysis scheme on clinicaltrials.gov allows HARKing (AKA ‘data dredging’)

What is variability?

- **Variability** describes how spread out or closely clustered a set of data is
  - May be described by range, standard deviation, interquartile range, mean difference, etc
  - 95% confidence interval helps by providing boundaries for the reliability of an estimate
  - Responses of individuals in a pain trial may not be normally distributed (bell shaped). An extreme example would be a U shaped distribution in which few subjects cluster around the mean response; the rest are either clear successes or clear failures

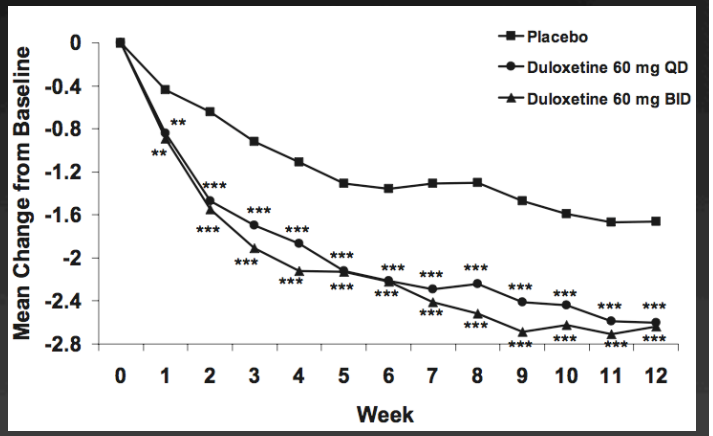
Can variability and error variance be reduced in a way that increases assay sensitivity?

- The variation observed might be *intrinsic* to the phenomenon of pain: distinct members of a population differ greatly and selecting subjects with consistent levels of pain may not be of benefit
  - Patient with unchanging “10/10” pain

Can variability and error variance be reduced in a way that increases assay sensitivity?

- Real time electronic data capture reduces measurement error due to the 'elevator effect'
- Averaging over time is the typical way of masking day-to-day within-subject variability in pain severity
  - Can replace repeated measures analysis with paired t-test
- Composite measures approach (such as combining pain and relief scales into one measure) assume that the components assess related but not identical components of improvement
- End of treatment global impression of change category assessments replace weeks of daily data with a single endpoint
  - Eliminates within-subject variability

Does adding additional pain assessment questions improve assay sensitivity?



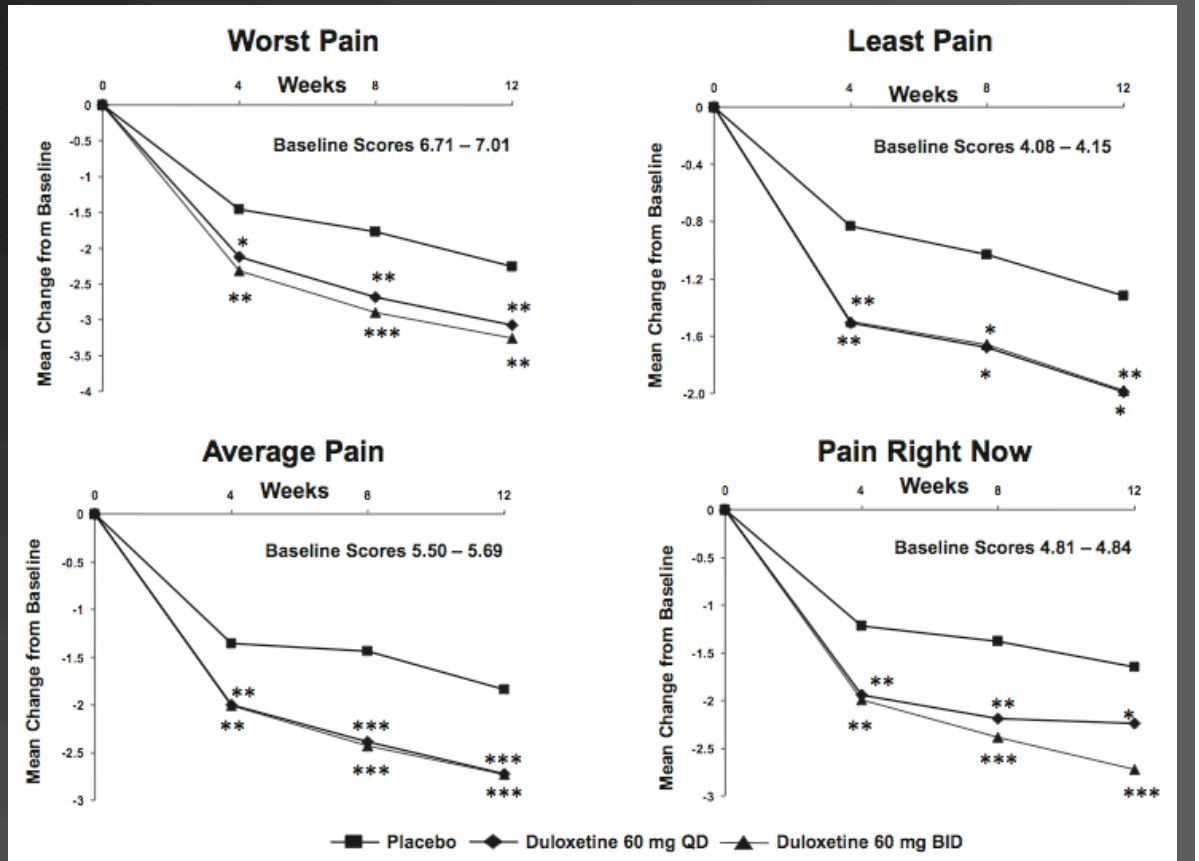


Figure 2 Brief Pain Inventory individual severity score. *$P \leq 0.05$ versus placebo; **$P \leq 0.01$ versus placebo; ***$P \leq 0.001$ versus placebo.

Raskin et al, Pain Medicine, 2005

Role of sample size

- “The random play of chance has large effects when the number of events is small” (Moore et al, 2010)
  - Smaller trials more prone to bias, especially publication bias
- If larger trials can reliably detect smaller differences between treatments, they have greater assay sensitivity but leave open the question of clinically meaningful differences between treatments

PAIN® xxx (2010) xxx–xxx

In Press

PAIN®

www.elsevier.com/locate/pain

Topical review

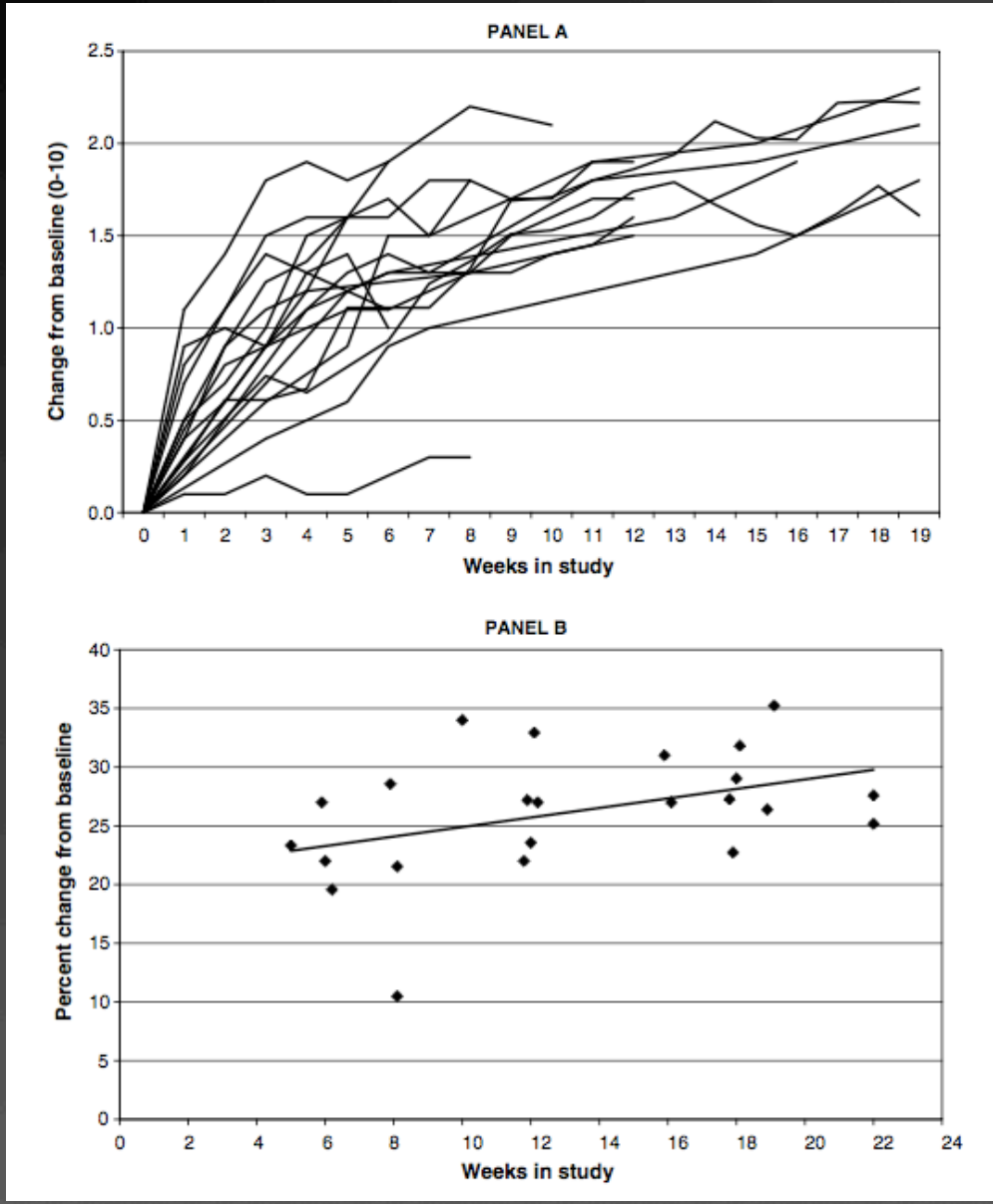
"Evidence" in chronic pain – establishing best practice in the reporting of systematic reviews

R. Andrew Moore [a,*], Christopher Eccleston [b], Sheena Derry [a], Phillip Wiffen [c], Rae F. Bell [d], Sebastian Straube [e], Henry McQuay [a], for the ACTINPAIN writing group of the IASP Special Interest Group (SIG) on Systematic Reviews in Pain Relief and the Cochrane Pain, Palliative and Supportive Care Systematic Review Group editors

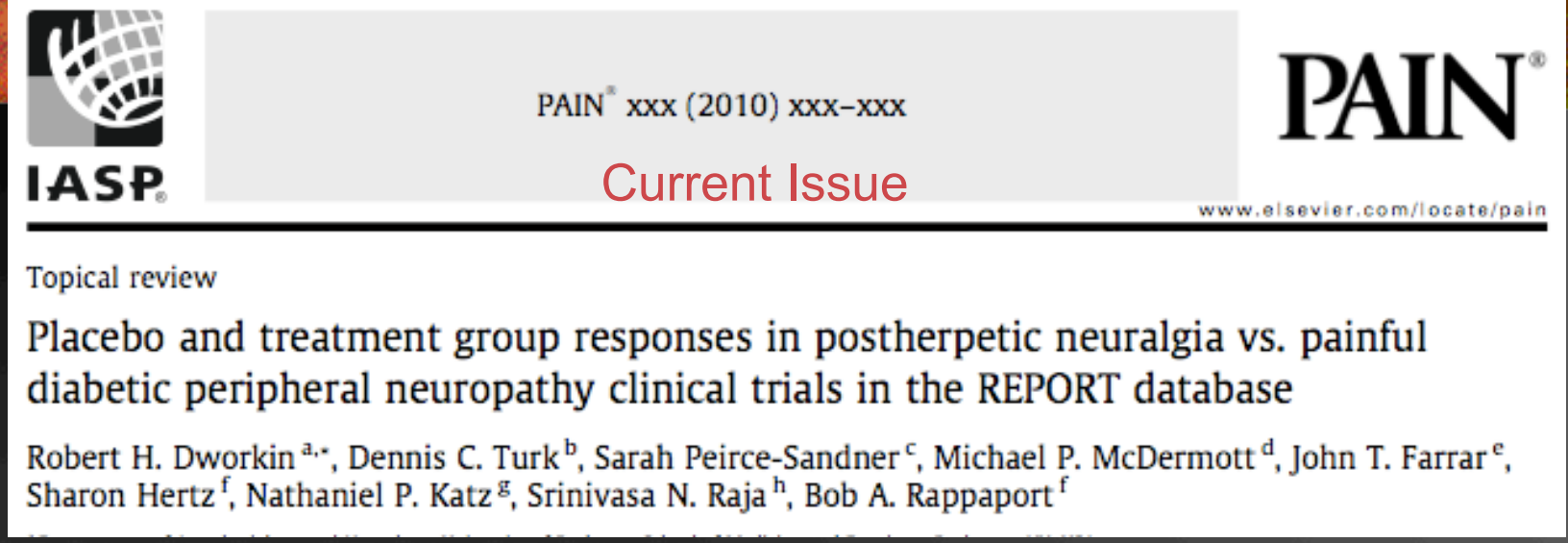
Response to placebo

- Often blamed for negative trial results
- Unpredictably variable, but seems to be a random effect
- Placebo rarely beats active, but magnitude usually tracks active
- Pregabalin DPN trials (Lyrica SBA)
  - 5 trials (3 successful) of similar design with total n=1,413
  - Placebo group pain reductions of 13, 18, 21, 29, 30%
- Riluzole NP trial (Galer et al, *Neurology* 2000)
  - N=43, crossover design
  - Placebo treatment pain reduction under 2%

Study Duration - Placebo Response



Quessy and Rowbotham
PAIN, 2008



IASP

PAIN® xxx (2010) xxx–xxx

Current Issue

PAIN®

www.elsevier.com/locate/pain

Topical review

Placebo and treatment group responses in postherpetic neuralgia vs. painful diabetic peripheral neuropathy clinical trials in the REPORT database

Robert H. Dworkin [a,*], Dennis C. Turk [b], Sarah Peirce-Sandner [c], Michael P. McDermott [d], John T. Farrar [e], Sharon Hertz [f], Nathaniel P. Katz [g], Srinivasa N. Raja [h], Bob A. Rappaport [f]

- Although PHN and painful DPN trials differed significantly in the percentages of studies with positive vs. negative outcomes in only one of the four analyses, there were consistent trends favoring a greater likelihood of a positive outcome in PHN trials, with 80-94% of PHN trials reporting positive outcomes compared to 60-71% of DPN trials
- Data suggest that PHN trials might have a greater likelihood of demonstrating efficacy
- Consistently greater mean treatment effects and standardized effect sizes as well as lower pooled variability of the endpoint pain values and difference scores in PHN than in DPN trials
- In analyses of covariance that controlled for treatment duration, mean placebo group response in the DPN trials remained significantly higher than that in the PHN trials

DPN vs PHN

- DPN is more common but doesn't have validated surrogate measures and has a higher trial failure rate
  - EMG/NCV and skin biopsy aren't pain measures but are measures of disease severity
- Possible to conduct a DPN trial entirely within the USA
- PHN is much less common but has validated sensory testing measures as surrogates and provides more consistent results in clinical trials
  - Area of pain is consistent across long time periods
  - Allodynia severity correlates with pain severity
  - Improvement is allodynia correlates with improved pain
  - Capsaicin response test may help stratify patients
- Not possible to conduct a PHN trial entirely within the USA unless the experimental treatment is very attractive to subjects and inclusion/ exclusion criteria are generous
- Both disorders have the advantage of well-defined dx criteria

Trial Design Options

- Most are randomized, placebo controlled, parallel design
- Number of groups depends on number of dose levels under study
  - More than 5 or 6 study arms may be associated with lower likelihood of success
- Randomization varies between 1:1 or 2:1 (total active : total placebo)
  - Greater skew towards active may reduce study power and increase response to placebo

‘Positive Control’ Designs

- In addition to a placebo group (negative control), some studies use a reference compound
  - For neuropathic pain
    - reference is most often either gabapentin or pregabalin
    - Opioids sometimes used as comparator
    - Antidepressant such as duloxetine could also be used
    - NSAIDS/COX-2 drugs would not be appropriate
- Several drawbacks to positive controls
  - Reduces the number of subjects providing safety data with the new compound unless the trial is made proportionally larger
  - Positive controls may fail to show the expected effect because of study-to-study variability and insufficient power in the positive control arm
  - Still need a placebo group
  - Must assure the positive control group is not enriched in some way
  - NET EFFECT: Adds considerable expense and time needed to complete enrollment

Crossover Designs

- A few examples in the neuropathic pain literature of 3 period (Raja) and 4 period (Gilron) designs
- Failure to return to baseline at the end of each treatment period complicates analysis
- Asymmetry in response, such as drug effective when given first and not second, also complicates analysis
- Risk that only the first period can be used to show efficacy and study ends up lacking power
- Subject attrition risk, unblinding risk
- Subjects like crossovers because they are assured of exposure to the new compound, but open label follow-on treatment accomplishes the same goal
- Not suitable for P3 trials

Enriched Enrollment

- Gain homogeneity in subject population at expense of generalizability
- Genetic screens
- Pain phenotype
- Pharmacologic probes:
  - Relief from intravenous lidocaine infusion to enter study of selective Nav blocker
  - Intravenous fentanyl challenge prior to entering opioid trial
  - Capsaicin response test prior to entering study of a TRPV1 antagonist
- Eliminating ‘responders’ after placebo run-in period is probably useless, maybe even counterproductive

EERW Designs
(enriched enrollment randomized withdrawal)

- All subjects receive experimental compound first, then ‘responders’ are randomly assigned under double-blind conditions to either continue on drug or switch to placebo
- Underlying assumption is that worsening of pain with drug withdrawal is as good a measure of efficacy as pain reduction in a simpler randomized single period parallel study
- Should not be combined with traditional placebo-controlled RCTs in meta-analyses

Multiple exposures: variations on ‘n of 1’ designs

- Five period enriched enrollment design (Fedele et al, PAIN 1989)
  - Initial cycle of dysmenorrhea treated with placebo (n=152, single-blind)
  - Responders to placebo (n=55) randomized to receive NSAID or placebo for 4 subsequent treatment cycles
- Cycle 1 placebo 84% NSAID 96%
- Cycle 2 placebo 29% NSAID 83%
- Cycle 3 placebo 16% NSAID 87%
- Cycle 4 placebo 11% NSAID 83%

Phase 1 and Phase 2
Incorporating Experimental Pain Models

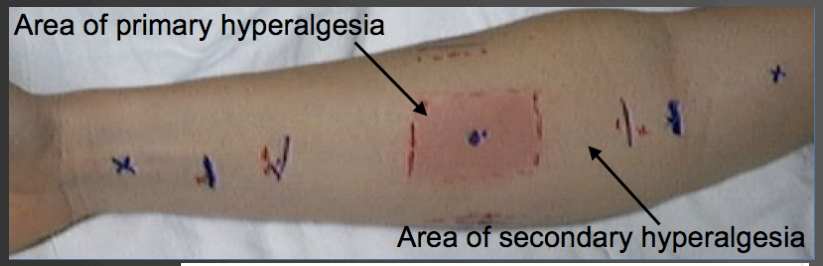
- Standardized activation of the nociceptive system
- Sample size ~ n = 20-25 in P1 studies
- Multiple session cross-over design possible
  - Determine dose with max analgesia and tolerable side-effects
  - Dose-response studies possible
  - Drug combinations can be tested
- Validated positive/negative controls
- Low additional cost when incorporated into P1 trials
- Must be non-invasive and low risk to use in P2
- Must also be simple to use in P3

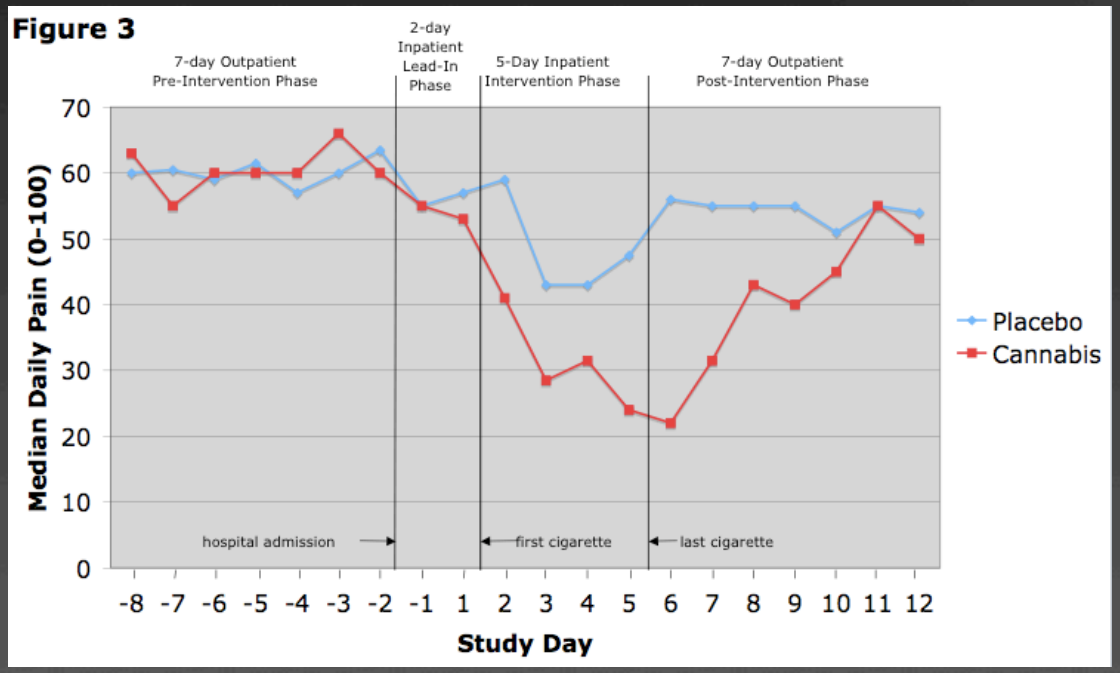
Experimental Pain Models

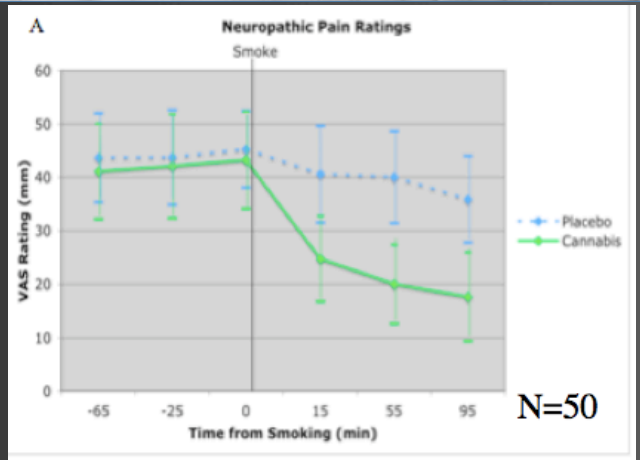
Preclinical		*Human*
Acute nociception ▪ Hot plate, tail flick	→	Acute nociception ▪ Pain threshold ▪ Suprathreshold stimulus pain ▪ (Cold pressor)
Inflammation ▪ UVB, carrageenan	→	Inflammation ▪ UVB
Neuronal sensitization ▪ Capsaicin, formalin	→	Neuronal sensitization ▪ Capsaicin, heat, electrical
Nerve injury ▪ SNL, CCI, SNI	⊗	Nerve injury ▪ Unethical to induce

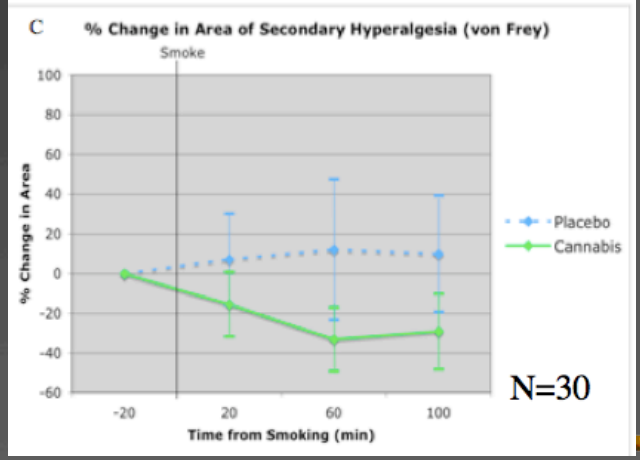
Pain models: Smoked marijuana for HIV-neuropathy pain

- 55 HIV+ subjects, 7-day inpatient parallel DB RCT
- Smoking cigarettes x 3 daily
  - 3.56% delta-9-THC vs placebo
- Outcome Measures
  - Daily diary VAS (primary)
  - Present pain VAS during smoking
  - H/C sensitization secondary hyperalgesia area









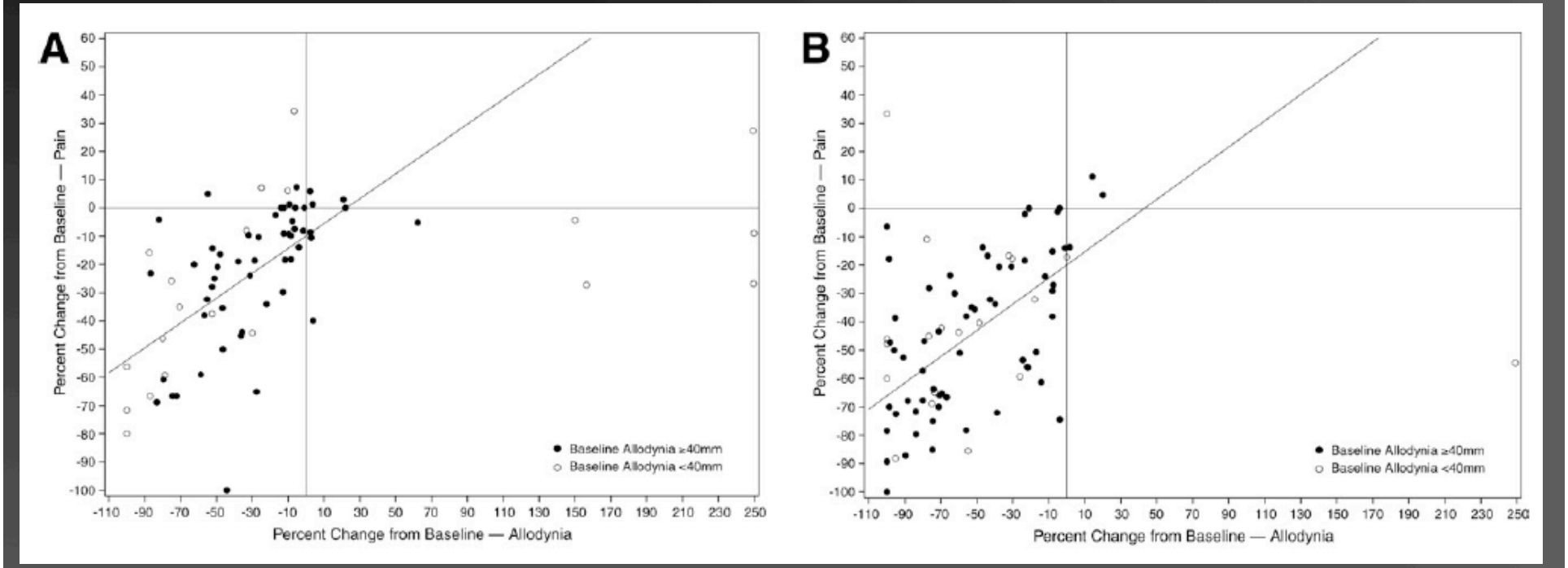
Abrams, Petersen, Rowbotham, et al, Neurology 2007

Allodynia Testing in PHN patients

Pregabalin for Postherpetic Neuralgia: Placebo-Controlled Trial of Fixed and Flexible Dosing Regimens on Allodynia and Time to Onset of Pain Relief

Brett R. Stacey,* Jeannette A. Barrett,† Ed Whalen,† Kem F. Phillips,† and Michael C. Rowbotham‡

J Pain 2008;9(11):1006-1017



Placebo

Flexible Dose

Summary

- Assay sensitivity limited by the largely subjective nature of current primary outcome measures
- Does the problem lie in preclinical models that result in candidate molecules acting on targets that aren't that important?
- Or, does the problem lie in a clinical trial testing paradigm that too often produces false negative results?
- Trial protocols can be made more sensitive and specific, but not without adding to trial cost and complexity