

# Comparative Effectiveness Research and FDA

Robert J. Temple, M.D.

Deputy Center Director for Clinical Science

Center for Drug Evaluation and Research

U.S. Food and Drug Administration

Nat'l Comparative Effectiveness Research Summit

Sept 14, 2010

# Current Status of CE at FDA

## I. Legislative history

The 1962 amendments require for approval substantial evidence that a drug will have its claimed effect, with the source of that evidence being adequate and well-controlled studies.

The Senate report made it very clear that there was no relative effectiveness requirement – a new drug need not be better than, or even as good as, available therapy.

With an exception I'll note, we have followed this practice and, indeed, we usually do not have direct comparative data for most drugs.

# CE at FDA (cont)

## II. An Important Exception

If inferiority to available treatment is a matter of life and death (or major morbidity) we can and do consider it.

This was stated clearly in April 1995 in a “Reinventing Regulation of Drug and Medical Devices” document by President Clinton and VP Gore, which was also set forth in the August 1, 1995 FR as an FDA “position,” signed by Bill Schultz, Deputy Commissioner for Policy.

# CE at FDA, Exception (cont)

The 1995 notice was not written primarily to assert a new concern or requirement. Rather, it was actually intended to “address concerns about a comparative effectiveness standard” that had been raised by drug and device manufacturers. They had been suggesting that we had moved to a “de facto” comparative standard.

It said that in general, “for most new drugs and devices intended to treat serious illness or provide symptomatic relief, effectiveness is shown by comparing the drug to placebo; i.e., there is no comparison to another active treatment.

“In certain circumstances, however, it may be important to consider whether a new product is less effective than alternative therapies, when less effectiveness could present a danger to the patient or to the public,” e.g.

1. When the disease is life-threatening or cause irreversible morbidity, or
2. Contagious disease

# CE at FDA, Exception (cont)

So the 1955 notice said that we care about CE in those special cases. But it is also true that in those cases one cannot ethically carry out a placebo-controlled trial so that the only choice is a comparative trial (active control trial).

The notice implies that we want the new drug to be as good as or better than existing treatment. BUT these trials infrequently show superiority of the new treatment. Moreover, they don't really show "equivalence" or "as good as." They are usually non-inferiority (NI) studies, seeking to show that the difference between the control established therapy and the new drug is not too large, i.e., small enough to allow a conclusion that the new drug has some effect.

In practice we seek more than "some effect" (as the effect is critical), often asking for evidence that no more than 50% of the control effect is lost.

# Comparative Effectiveness

In fact, we see relatively few serious attempts to assess comparative effectiveness (show actual equivalence). We often see, as noted, NI trials Uncommonly we do see attempts to show superiority to an active control.

A fair number of trials do have active controls as well as placebos; this is common in studies of pain, depression, and many other symptomatic conditions, but the active control is there to establish assay sensitivity, i.e., to show that the study is capable of detecting the effect of an active drug vs placebo (highly relevant if the test drug does not beat placebo), and the trials are rarely sized for a valid comparison of the active drugs.

[They could be sized that way; they just aren't. The active drug groups would need to be very large to be able to show a small difference, e.g., a difference of 25% of the drug-placebo difference.]



# CE and FDA, Exception (cont)

The 1995 notice specifically noted one kind of comparative claim of interest. It said we are willing to approve drugs for subpopulations who do not respond to, or don't tolerate, available therapy, a kind of comparative effectiveness special case that should be of more interest than it has been.

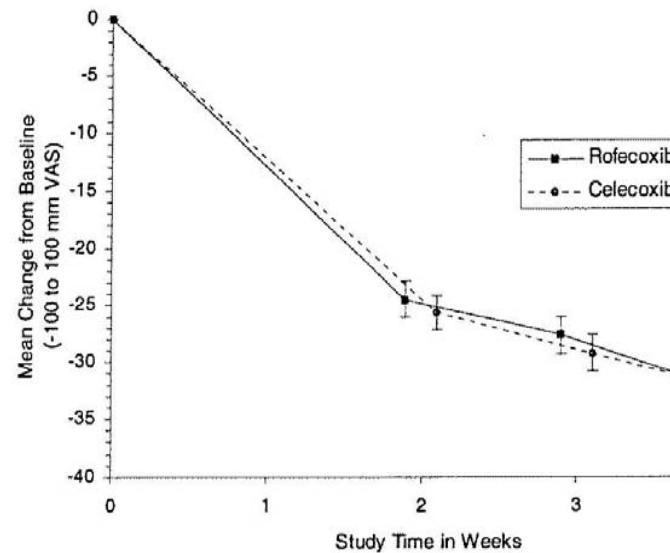
# Superiority in a Subset (Non-responders)

A very attractive study design, so attractive it seems almost unfair (how could you lose?), is to study a drug in failures on another therapy or in people who cannot tolerate other therapy. After all, if people really do respond differently to alternatives, surely this is the population that could show that. Strictly, this is not really a study to show comparative effectiveness, but it is very useful to know whether a drug works in non-responders or intolerants to other therapy. Oddly, such a showing is rarely even attempted properly. A study that can really show an effect in non-responders or intolerants requires randomizing patients back to the failed or poorly tolerated treatment as well as to the new drug.

I'm aware of only 4 attempts to rigorously show an effect in non-responders, 3 successful – clozapine, bepridil, and captopril, and these drugs had toxicity concerns that allowed approval only because they had the data to show this benefit, and one total failure, rofecoxib in celecoxib non-responders.



Figure 1  
Pain at Night while in Bed (WOMAC)  
Mean Change from Baseline  $\pm$  SE by Timepoint  
(Primary Per-Protocol Approach)



Note that without a celecoxib control, rofecoxib would have appeared VERY effective in this NR population.

# Superiority in a Subset (cont)

There have been surprisingly few attempts to show better tolerability of a drug in people who had adverse effects on another drug, even though this would seem to represent an attractive opportunity. In this trial it is again critical to randomize patients back to the poorly tolerated drug and the new drug. The historical poor tolerance may not show up in a second exposure. In such studies:

- It was clearly shown that losartan did not induce cough in patients who reliably coughed on lisinopril.
- Wellbutrin was shown not to affect female sexual function in patients whose function was impaired with SSRI's.

If there are more of these I'm not aware of them.

# Superiority in a Subset (cont)

It would seem possible to use genetic information to identify patients who would do better on one drug than another, and such cases are actively sought. An easy case would be to study people who do not form the active metabolite of a drug because they lack the CYP450 enzyme needed to convert the drug (e.g., such drugs as tamoxifen or clopidogrel) to its active metabolite. They could then compare the drug having such a problem with a similar drug that did not, either in the overall population or, preferably, in the subset that does not make the metabolite. At some point, of course, these trials become ethically problematic.

Not really a superiority or comparative study but a very informative, and very frequent kind of study, is one that adds a new treatment to established therapy (add on study), showing an additive effect. These have been done for a wide variety of treatments: for heart failure, hypertension, CAD, pain, etc. Usually, to succeed, the added drug would need to be pharmacologically distinct.

# CE at FDA, Exception (cont)

So, in practice, we see occasional examples of attempts to show that one drug is better than another, either overall or in a subset. But what we see most is active control studies intended to show “non-inferiority.” These are active control comparative studies, but their goal is to provide evidence of effectiveness, i.e., an effect greater than placebo and not “too” much less than the control. We would expect such evidence of effectiveness for a new drug to that life-threatening or debilitating illness, such as cancer or serious infectious diseases.

Apart from that expectation, a comparative study is the only one possible because it would be unethical to do a placebo-controlled trial. Such a trial would mean denying patients a known effective therapy that would prevent death or serious disability (clearly stated in ICH E-10).

So, in those cases non-inferiority (NI) studies are carried out to show effectiveness. It is important to appreciate what NI studies are, and are not.

# CE at FDA, Exception (cont)

NI studies show effectiveness by showing, at a minimum, that a new drug is not worse than the active control by an amount (the NI margin) equal to the whole effect of the active control; i.e. that some of the effect is preserved. The effect the active control has in the new study (based on part performance) is called  $M_1$ , the largest possible NI margin because a difference between the new drug and control  $> M_1 =$  all effect lost.

BUT, if you must ethically do an NI study, the control effect is of perceived great value, so we usually ask that some fraction of the control effect be preserved (typically 50% although usually more for antibiotics), so the study needs to rule out loss of  $M_2$  (the largest fraction of the whole effect of the active control that can be lost).

This IS a kind of relative effectiveness requirement.



# CE at FDA (cont)

## III. Comparative Effectiveness Claims

There are 2 kinds of claims to consider:

- Similarity/equivalence
- Superiority

When an NI study succeeds in showing non-inferiority (i.e., that at no more than 50% of the control effect has been lost), that does not really show “equivalence,” or support a claim of equivalence. That would be a much higher standard, not really defined but plausibly represented by the 80-125% CI we demand for bioequivalence for generics.

I can't think of any cases where we've accepted such a claim, other than for some topical products (skin, eye) that did indeed meet the 80-125% standard.



# Superiority Claims

There are several possible kinds of superiority that could be shown.

1. Overall superiority in effectiveness in the general population.
2. A safety advantage in the general population.
3. Advantages in subsets
  - Greater effectiveness in non-responders to another drug
  - Better tolerability in people with an adverse effect on another drug
  - Effectiveness in a genomically or proteomically defined subset
  - Other: better compliance (o.d. dosing) leading to better outcome

# Comparative Claims (cont)

Overall superiority claims have been sought and our standard has been the approval standard: adequate and well-controlled studies (usually more than 1). Moreover, the studies must be fair, as discussed in ICH E-10 [Choice of Control Group and Related Issues in Clinical Trials, 2001]. A comparison could be unfair if:

- Low dose of the comparator was used.
- The patient population had previously failed the older drug (but note that although this does not show superiority, as noted above, it is a very useful study).

# Superiority Claims (cont)

It is not easy to get such a claim, but there have been successes.

- Two large studies showed that candesartan had a larger blood pressure effect than losartan (in labeling).
- LIFE study (losartan vs atenolol) showed superiority of losartan vs stroke, but in only one trial. Losartan got stroke claim, but not a direct comparative claim.
- Prasugrel was more effective than clopidogrel in decreasing the rate of heart attacks in people with acute coronary syndrome (it caused more bleeding too).
- PPIs have claims vs H2 blockers.
- Anastrozole is superior to tamoxifen as adjuvant Rx post surgical treatment of breast Ca, especially in ER positive.
- Irbesartan delayed decline in renal function in type 2 diabetes; it was superior to amlodipine, which had no effect.

# Superiority Claims (cont)

We thus use the legal effectiveness standard for what is, in fact, a claimed effect, just as the law demands. It is a high standard, but it is not easy to see how a lesser standard would fit the law nor (my opinion here) whose interest such a standard would serve.

And we can be certain that people, will, if given the opportunity, use lower quality data to make such claims. We know that before there was an effectiveness standard, the effectiveness of thousands of drugs and more thousands of claims were unsupported and proved unsupportable. We know that claims for dietary supplements, unencumbered by any requirement for controlled studies, are rarely supported by such trials. It is not easy for me to see a public interest in a proliferation of comparative effectiveness claims based on data known to be unsuited to the purpose.

# Comparative Claims (cont)

## IV. Current Interest in CE

As anyone can see, there is a large and growing interest in CE, and, of course, there should be.

# Comparative Effectiveness

## The excitement is palpable. . . And why not?

Despite the paucity of comparative trials, they are very important. Clinically, after knowing a drug works and is safe (which FDA takes care of) most of the important questions about drugs are comparative, i.e., deciding which drug to choose in general or for specific patients

- Does it work better than alternatives? Does it work faster? Does it have less of a serious or annoying adverse effect?
  - In all patients
  - In a subset
- Can you add it to other treatments?
- Does it have some additional benefit in some or all patients?
- Does it work when others fail?
- Is it about as good, but cheaper?

But there usually isn't much of such data

- Drug companies historically have not done proper comparisons except for the NI trials where active control trials are ethically necessary, but those NI trials are not really useful comparative trials in most cases.
- Trials almost never have  $> 1$  comparator; usually interest is in comparing all members of a class
- Trials rarely compare across classes
- Trials usually are too small to give definitive answers



# Comparative Effectiveness

So the medical need for comparative data is great and apparent.

We also need to acknowledge a major interest in costs of therapy. All of us, payers too, as I understand it, are willing to pay more for a treatment with an advantage

- maybe after other therapy fails
- maybe it depends on how much advantage

but there is great reluctance to pay more for the same effect. So a major interest of payers is showing whether there is an advantage. (Could they just agree to pay only when one is shown?)

But wanting comparative data does not necessarily mean we know how to get comparative data of high quality with reasonable effort and at acceptable cost.

And it seems to me it must be of high quality. Mistakes will greatly undermine the credibility of the effort, not to mention the harm they could do, and there is already concern about people (other than the personal MD) deciding on choice of therapy. Imagine the response if the choice (decision not to pay) is based on less than credible data.

# Comparative Effectiveness Is Not the Only Need

I realize there is current enthusiasm for comparative effectiveness, but we need to keep our balance. If there is to be funding for trials, and funding is not infinite, there are other critical issues too. For example:

1. Do our physical therapy and non-pharmacologic psychiatric interventions work at all? Many are untested. They cost a lot.
2. How can we improve compliance/persistence with vital chronic therapy (lipid-lowering, BP, diabetes control, smoking cessation, weight reduction)? Could cluster-randomized trials help?
3. How low should we push LDL, BP, BS; is it the same for everyone? How many anti-platelet treatments should we give in ACS and after PCI and how long should we give them? There is work on some of this, of course, but there is a long way to go.

The right determination of what to study is the value of what we'd learn, not whether it is comparative. The best study may be a comparison of an added (to standard regimen) treatment to no added treatment. The IOM list of 100 important CER issues is very consistent with this.

# Comparative Effectiveness Issues

Once we've decided in a given case that comparative effectiveness is an important question, studying it raises a host of issues, all of them interesting and most of them matters of long FDA and personal interest, including

1. How we can obtain credible evidence of comparative effectiveness and safety: role of randomized trials, meta-analyses of trials comparing the agents of interest, cross-study comparisons, observational data.
2. Often (? usually) you're interested in comparisons with multiple drugs in a pharmacologic class, not just one, and frequently with drugs in different pharmacologic classes as well. How to compare multiple treatments is challenging and doing it is costly.

# Comparative Effectiveness Issues

3. Even if there is funding, there are major challenges in doing comparative effectiveness trials:

- Differences between effective treatments will, at most, be small, so that
  - Trials will need to be very large to show them
  - Nothing but an RCT (or conceivably pooled RCTs) directly comparing the treatments will be credible
- Showing there is no (or not much) difference between treatments, often the goal of the comparison, is also very hard, will often need a placebo group to assure assay sensitivity (supporting the validity of a finding of no difference), and again, trials may need to be very large, depending on the size difference to be ruled out
- Efficiency and simplification are critical

# Comparative Effectiveness

## You Need Randomized Trials (Maybe Meta-analyses)

With rare exceptions, differences between drugs, if any, will be small, considerably smaller than the whole effects of the drugs, which themselves are often small. And the difference you want to rule out is also small. Consider outcome studies.

A blockbuster outcome study in CHF, hypertension, CAD will reduce event rates by 40%. Far more commonly, it will be more like 20%. If that is the whole effect of the drug, i.e., an HR of 0.8, a complete loss of that effect ( $1 \div 0.8$ ) would give an HR of just 1.25 for the comparison of a new drug vs the standard; i.e., it would be only 25% worse.

But between-treatment differences of interest, or the difference to be ruled out, will not be the whole drug effect, but something smaller: suppose you wanted to detect a loss of half of the 20%, a 10% difference. In that case the HR for the inferior drug, the upper bound of the CI for new/old, would be just 1.125, i.e., very hard difference to detect.

In terms of risk, that means you're trying to detect a risk ratio of 1.1-1.2 at most. This is possible in large ambitious RCTs, but you cannot reliably detect such differences in anything but randomized trials (or conceivably, very well done pooled analyses).



# Comparative Effectiveness

## You Need Randomized Trials (cont)

Symptomatic conditions usually pose at least as great a problem (and one might ask how important it is to rule out or document small differences).

Trials of antidepressants fail about 50% of the time (cannot distinguish drug from placebo) and a typical effect size is 3 HamD points (drug-placebo). Trials these days are 100-200/arm.

A large between-drug difference could conceivably be 1.5 HamD points (that would be a very large difference and, usually, the less effective agent would have had difficulty beating placebo). Far more likely would be a difference of 1.0 HamD point or less.

Trials to show such differences would be enormous. Moreover, failing to show a difference would be meaningless without a placebo group to assure assay sensitivity (ability of the study to detect effects).

Most symptomatic conditions are like this, except where effects are huge (Tysabri vs interferon, a difference so large it is obvious in cross-study comparisons).



# Comparative Effectiveness

## You Need Randomized Trials (cont)

It is not insulting to observational/epidemiologic approaches to say that they are generally unreliable when trying to detect risk ratios of  $< 1.5$ , and certainly when looking for risk ratios of 1.2 and less. It is not a lack of power. What makes such approaches tempting is in fact their huge power and speed.

But those advantages do not make up for potential bias and confounding. There are many sobering examples. Let me give two:

Hormone replacement therapy

Calcium channel blocker toxicity

The incorrect results of epidemiologic studies in these cases, unfortunate at best, quite harmful at worst, did not usually arise from obvious methodological flaws or foolishness. The methods are just not reliable for small differences, usually because without randomization you cannot assure the needed close similarity of the groups receiving each treatment and because of substantial, usually unidentified, multiplicity.

# Hormone Replacement Therapy

Although observational studies did not give uniform results, hormone replacement therapy was thought to reduce coronary heart disease (CHD) by 40-50%. The Women's Health Initiative randomized > 16,500 post-menopausal women 50-79 to HRT (0.625 oral equine conjugated estrogens + 2.5 mg medroxyprogesterone acetate) or placebo.

Despite favorable effects on LDL and HDL cholesterol and triglycerides, coronary heart disease effects were adverse

	HRT	Placebo	HR	95% CI
	8506	8102		
CHD	188	147	1.24	1.00-1.54
NFMI	151	114	1.28	1.00-1.63
Fatal CHD	39	34	1.10	0.70-1.75
CHD, revasc, angina	369	356	1.00	0.86-1.15

# HRT

HRT has obvious short-term benefits but the case for CHD prophylaxis, although logical (women have less CHD than men while producing hormones and catch up with men after menopause) and epi-supported, was not only not made, but CHD harm was strongly suggested.

There were also increases in breast Ca, thrombophlebitis, pulmonary emboli.

# Calcium Channel Blockers

The full CCB story deserves a book, not a few slides. Over the course of several years, roughly 1995 through 2002, cohort and case control studies, almost all of them comparing CCB's with other antihypertensive drugs, suggested that CCB's:

1. Increased the rate of AMI (Psaty, et al, JAMA, 1995).
2. Increased mortality (Furberg and Psaty, Circulation, 1995) actually a subset of a meta-analysis of nifedepine).
3. Increased mortality (Pahor, et al. J Am Geriatrics Society, 1995, a cohort study). Oddly, verapamil was protective; diltiazem, nifedipine AND ACEIs all gave RR's of 1.5-1.9.

# Calcium Channel Blocker (cont)

4. Increased GI bleeding (Pahor, Furburg, et al Lancet 1996; Kaplan, Furburg, Psaty, Letter to Age and Aging, 2002).
5. Increased risk of all cancer (Pahor, et al Lancet, 1996). Oddly, risk was up for verapamil and nifedipine, not at all for diltiazem.
6. Increased breast cancer (Fitzpatrick, Furburg, et al, Cancer, 1997).
7. Caused suicide (Melander, BMJ, 1998).

# Calcium Channel Blockers

To my best knowledge, none of these findings were confirmed in RCTs (ALLHAT, various CAD trials of verapamil and diltiazem). The findings were discussed, condemned, supported in dozens of papers. A Sounding Board piece (NEJM) in 1997 by Deyo, Psaty, and others described manufacturers' attempts to gain access to Psaty's records related to the 1995 AMI study, as well as many hostile academic (perhaps manufacturer-supported) critiques, citing this as a classic case of attacking scientific results that run counter to financial interests and strongly-held beliefs. That could be part of it, but there were certainly scientifically sound bases for criticism of those studies as well. There was a paper (can't find) comparing industry support for authors supportive and opposed to the CCB findings. Guess which ones had more support. Of course, they were correct.



# Calcium Channel Blockers

People can form their own views as to what all this illustrates. Among other things it shows

1. Inadequate attention to description and presentation of epi results. Epi studies need careful protocols that record all changes, well-described hypotheses made before the study, correction for multiple hypotheses (i.e., all the things we've learned to ask about RCTs). I think they should always be replicated unless risk is very large.
2. Particular risks when an adverse effect is a possible consequence of the disease, where the severity of the condition and the effect of treatments can be confounded. That would always be true for a comparative effectiveness study.
3. RR's  $< 2$  need great care and should be viewed very skeptically (although they can surely generate hypotheses). Comparative effectiveness will almost invariably be about RR's  $< 1.5$  and indeed  $< 1.2$ , a major challenge.
4. Such errors are not benign; they can interfere with important therapy or encourage harmful treatments.

# Calcium Channel Blockers

With recognition of the need to get BP under better control, CCB's must be used in many people. They may even have advantages in some populations. But their use was somewhat marginalized for many years because of these concerns. There is little of that concern expressed in JNC VII (2004), so perhaps the damage has passed.

**WE DO NOT WANT ERRORS.** The questions addressed in comparative studies, especially outcome studies, matter. To get correct answers, the comparisons need RCTs unless differences are very large. They hardly ever are.

# Comparative Controlled Trials - Difficulties

There is not a great deal of experience in doing such trials properly, and the challenges are substantial.

# Comparative Effectiveness - Difficulties

## A. Multiple Drugs of Interest

What physicians really want to know is how all (or at least many) members of a class compare. This is not easy, for many reasons.

1. For many comparisons you need a placebo to assure assay sensitivity, a potential problem for post-approval, often large, studies.

You can sometimes use a NI study design where there is a solid basis for knowing the effect of the positive control, but that would be impossible in depression, anxiety, and most symptomatic conditions; for those you need a placebo to show ASSAY SENSITIVITY, i.e., that you can tell one thing from another, because many studies in those conditions cannot tell active drugs from placebo [You could show superiority without the placebo, but not similarity].

# Comparative Effectiveness – Difficulties

2. Hard to expect a company to study multiple drugs in one study.

Separate comparisons don't really tell you what is needed; you can't usually compare across studies.

Multiple comparisons have been carried out by government: ALLHAT and CATIE

- ALLHAT – chlorthalidone, lisinopril, doxazosin, and amlodipine

Ambitious but results hotly debated; there were design problems (couldn't add diuretic to lisinopril). Meta-analyses and another large trial suggested different answers.

ALLHAT clearly did show that cheapest drug (chlorthalidone) was a reasonable start, but drugs have different properties: some treat diabetic nephropathy (ARBs), CHF (ACEI's, BBs, diuretics), angina (CCBs, BBs), or post-infarction (BBs, maybe ACEI's).



# ALLHAT

## Wonderful Intent, Hard Trial

Compared – clorthalidone, lisinopril, amlodipine, and doxazosin.

Some element of interest in cost: “Are newer types of anti-HT, which are currently more costly. . . as good as or better than diuretics in reducing CHD incidence and progression” (abstract, Am J HT, 1996; 9: 342-360).

### Problems:

1. Plainly, ALLHAT was an NI study, but no discussion if NI margin for any endpoint. Doing so would have been difficult because regimens did not match past effective regimens and population (enriched for black patients) was not the same. Did this disadvantage lisinopril? The question is, then, what does failure to see a difference between the treatments mean? It is very hard to know and, to my best knowledge, was not addressed at all.

# ALLHAT

## Problems (cont)

2. No beta-blocker group.
3. Treatments did not get usual accompaniments because you could not add another test drug.

E.g., could not add diuretic to lisinopril. This is particularly critical for black population and for CHF (all CHF studies of ACEI's were added to diuretic). Lisinopril thus had slightly poorer control of BP.

4. ACE inhibitors were superior for CV events in a different study, the Second Australian National Blood Pressure Study (HCTZ, mostly white).
5. Did we learn enough? I'd say yes: main lesson is that it doesn't matter too much how you get the BP down.

# Comparative Effectiveness – Difficulties

- CATIE

NIMH: 4 atypical (olanzapine, risperidone, quetiapine, ziprasidone), one typical (perphenazine) anti-psychotics used in schizophrenia showed olanzapine was most effective (fewest D/C for lack of effectiveness) and least well-tolerated (most D/C for intolerance). CATIE worked because there were differences. Had there been no differences, it would have, absent placebo, been wholly uninformative.

Both ALLHAT and CATIE were very expensive. Perhaps worth it but at those prices can't do too many.

# CATIE

1493 schizophrenics randomized to olanzapine, perphenazine, quetiapine, or risperidone (later ziprasidone).

Endpoint was “discontinuation” of treatment for any cause.

Outcome	Olanz 330	Quet 329	Risp 333	Perph 257	P-value
All DC (%)	64	82	74	75	< 0.001
Lack of E (%)	15	28	27	25	0 signif <
Intolerability (%)	19	15	10	16	0 signif >

# Comparative Effectiveness – Difficulties

## II. Sample size is very large

Suppose you wanted to compare anti-depressants. Current studies vs placebo these days use 100-150 patients per group to show a drug-placebo difference of 3-4 HamD points. You need placebo for assay sensitivity. What HamD difference do you want to rule out?

2 points – no chance it's that large

1 point – sample size for active drug would be many hundreds, perhaps 1000. Is that really feasible?



# Comparative Effectiveness - Difficulties

Given the problems (multiple drugs of interest, small effect sizes) it is tempting to seek alternative data sources, notably meta-analyses and cross-study comparisons. The problem is that in a cross-study comparison patients are not randomized to treatments and patients on one drug may differ from patients on another, making such comparisons treacherous. The problems and potential biases in meta-analyses are well-recognized, but at least potentially, these are well-randomized comparisons.

# Possibilities

The problems I've described can perhaps be overcome, if there is enough interest. Possibilities include

- Doing large studies in treatment environments already collecting data (HMO's, VA), perhaps using internet to enroll, gain consent, follow PRO outcomes. These would not select too much, i.e., we're talking about very pragmatic trials. We know very large trials in Europe (ISIS, GISSI) had reasonable costs.

If patients and doctors were “into” this, maybe it wouldn't cost too much.

- Placebos are, at least now, hard to use in the real world but you don't need one to show superiority. But in symptomatic conditions, absence of a placebo will lead to inability to interpret results if no treatment is superior.