

Comparative Effectiveness FDA Activities

Robert J. Temple, MD

Associate Director for Regulatory Policy

FDA/Center for Drug Evaluation and Research

NHPF

Comparative Effectiveness

Sept 26, 2008

Overview

Legal requirements for comparative effectiveness data

When comparative data are required

- Practice

- Clinton-Gore Reinvention

Practical Considerations

- NI Studies

- Superiority studies

- Studies in non-responders or intolerants

Limitations

- Equivalence VERY hard to show

- Multiple drugs rarely studied by drug companies: NIH does (ALLHAT, CATIE)

- Cross-study (indirect) comparisons and observations from sources other than RCTs are treacherous

Overview - Summary

We recognize the high level of interest in comparative data. After basic S&E, there is no area of greater interest to patients, physicians, and payers.

FDA's ability to require comparative data is relatively limited, but we are very interested, and fairly experienced in how to do it, and aware of potential errors.

Law

1. Safety

Demonstration of safety, shown by “all tests reasonably applicable” to show whether drug is safe when used as labeled. (Unchanged since 1938).

“Safe” generally understood to mean benefits outweigh risks.

2. Effectiveness

“Substantial evidence” that it will do what is claimed.

Substantial evidence means convincing evidence from “adequate and well-controlled studies” (generally > 1 , but with exceptions).

3. Labeling

Adequate directions for use and not false or misleading.

There is NO requirement (and this is explicit in legislative history) that a new drug be superior to, or even as good as, existing therapy; i.e., no “relative efficacy” requirement. It could be asked, however, whether in some cases adequate directions for use could include comparative information. This is not a settled issue.

Regulations

Regulations do not really add much to what must be shown

- Define adequate and well-controlled studies.
- Get us all data with complete submission of and access to detailed patient data.
- CRF's for deaths and adverse drop outs.
- Must assess effectiveness and safety in demographic (age, sex, race)

Guidance, Practice, Interpretation

Certain Comparative Data Are Required

Although placebo-controlled studies are often the easiest to interpret, and it can be useful to add a third, comparative, arm. There are situations in which they cannot be done and would not answer the relevant question.

ICH E-10 states:

In cases where an available treatment is known to prevent serious harm, such as death or irreversible morbidity in the study population, it is generally inappropriate to use a placebo control.

These situations overlap substantially with situations in which you need some degree of comparative data because lesser effectiveness would be dangerous, a SAFETY issue, and NOT a relative effectiveness issue (because we do not have authority for those).

Lesser Effect Is Unsafe

Historically, anti-infective trials, many cancer trials (where you're not adding a treatment) and, more recently, cardiovascular trials in settings where favorable outcome of the standard is known, have been non-inferiority trials, comparing the new and standard treatment [Almost always more valuable to do an add-on study

AB vs B (A is the new drug)

where that is reasonable, because you advance treatment but it won't usually work if drugs are in the same pharmacologic class and sometimes you add too much toxicity.]

So in certain areas, NI studies, which are comparisons, are the norm.

Lesser Effect Is Unsafe

Clinton-Gore Reinvention – Late 1990's paper – describes FDA policy on comparisons. It states clearly that we ask for this information where lesser effectiveness would represent a safety concern. It cites oncologic settings particularly but, as noted, there are clearly others.

Apart from cases where decreased effect is plainly unacceptable (survival), recent cases have broadened this view and this has been noted in the press.

Non-Lethal Settings

Certain psychotropic drugs

We have taken the position that marked inferiority (not very well-defined yet) of anti-psychotics and antidepressants to standard treatment is “not safe.” In these cases, by the time it is appreciated the drug didn’t work (2 weeks or so), bad outcomes could have occurred.

Somewhat controversial and we do appreciate need for a range of drugs with different properties and recognize the BELIEF that different patients respond to different drugs, even of the same class. This belief, I should note, fits well with the “individualization” enthusiasm but to date has rather little evidentiary support (lots of anecdote, however).

Practical Considerations

1. Limitations of Non-Inferiority Studies and Comparative Studies Generally
Superiority is relatively easy to interpret. Non-inferiority (NI) is a real challenge.

- NI, in most cases, does not really show similarity in the usual sense of the word, unless the drugs have very large response rates (antibiotics for UTI, strep throat).
- NI trials are a long story but basically you need to show that the new drug is not so inferior that all of the effect of the control is lost. E.g., if a statin reduces AMI plus CV death by 30%, you must not see that the new drug is 30% worse. In fact, because NI studies are used mainly where drug effect is valuable, we generally ask that a loss of more than half the control effect be ruled out. If event rates are small, say 2% per year, showing this needs a big study. But even this showing doesn't really prove "equivalence."

NI Studies (cont)

Consider what was needed to gain approval of a new TPA, the first NI case I recall.

We:

- Calculated (from a meta-analysis of many trials of streptokinase) the effect of a thrombolytic as mean 25% reduction in mortality, with a 90% CI lower bound of 22%.
- We said prove half (50%) of that (11%) is preserved, requiring a 15,000 patient study [Ad Com said preserve 75% but that needed a 50,000 patient study.]
- The proof of effect is that the upper bound of a 95% confidence interval of C-T (control minus test drug, i.e., the advantage of the control drug) is $< 11\%$ (pretty similar to $p = 0.05$).

This assures (rules out less than) 50% retention of effect. You can also look at the point estimate, which usually is very similar for the two treatments. But is that what people mean by equivalence (at least 50%)?

NI Studies (cont)

The reality is, for CV studies, or oncology studies, NI means assurance that at least half of the critical effect is preserved. Of course in a study of drug vs placebo, with significance shown at $p = 0.05$, all you know is that the effect is > 0 . Again we look at point estimates.

To really show “equivalence” (e.g., the generic bioavailability standard of 80-125%, studies would need to be vast.

You can do NI studies only where drugs are regularly superior to placebo, leaving out most symptomatic conditions. Lack of evidence of a difference is not evidence of the absence of a difference if trial lacks assay sensitivity (i.e., could have distinguished drug from placebo), which cannot be assumed for most symptomatic conditions.

Superiority

Sometimes one can indeed show superiority of one drug to another. But it's very hard because the drug-drug differences are smaller than the drug-placebo differences and often need large studies.

Some examples:

- PPIs vs H2 blockers
- LIFE
- PROVE-IT
- Candesartan

LIFE

RCT: The Losartan Intervention for Endpoint Reduction in HT

Losartan vs atenolol in 9193 hypertensive patients

Events

	Losartan N (%)	Atenolol N (%)	RR	P
Primary: CV death, AMI, Stroke	508 (11)	588 (13)	13%	0.02
Stroke (F/NF)	232 (5)	309 (7)	25%	0.001
AMI (F/NF)	198 (4)	188 (4)	-7%	0.491
CV mortality	204 (4)	234 (5)	11%	0.206
CHD	125 (3)	124 (3)	-3%	0.839
Stroke	40 (1)	62 (1)	35%	0.032

Won on primary endpoint but all effect on stroke.

PROVE-IT

RCT: Pravastatin or Atorvastatin Evaluation and Infarction Therapy

Pravastatin 40 mg vs atorvastatin 80 mg in 4162 patients with ACS within 10 days.

	Pravastatin 2003	Atorvastatin 2009	
LDL cholesterol	95 mg %	62 mg %	
All death, AMI, Stroke, unstable angina, revascularization	26.3%	22.4 %	P = 0.005

Studies in Non-Responders

Studies in non-responders are a very attractive design. It should give your new drug an edge (they've failed the other) and has allowed approval of drugs otherwise too toxic

- Captopril (thought to cause agranulocytosis) was superior to diuretic, reserpine, hydralazine (triple therapy) in patients failing triple therapy.
- Bepridil (a CCB) superior to diltiazem for angina in diltiazem failures.
- Clozapine superior to Thorazine in standard therapy failures.

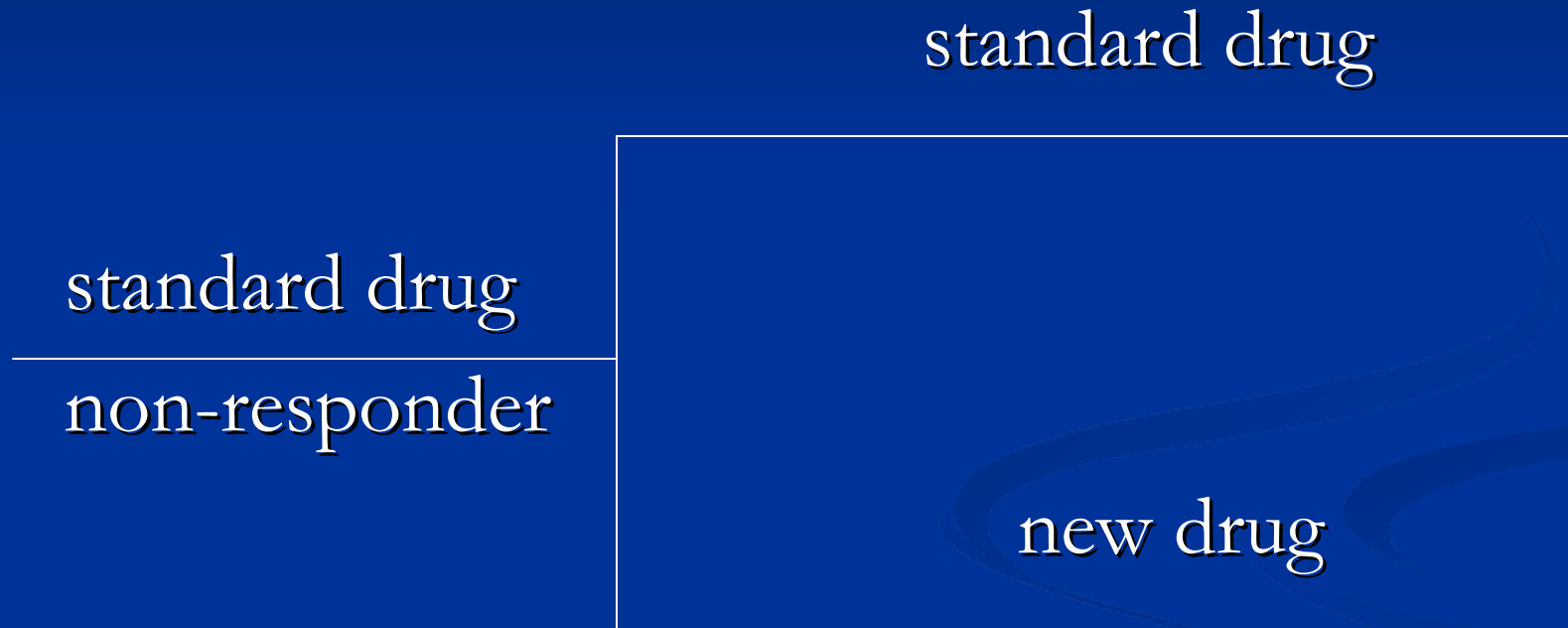
The design must randomize to failed and new drug.

Candasartan

Astra-Zeneca engaged in a determined effort to compare candasartan to losartan, the first AIIB. Initial efforts (in advertizing) were rejected (compared with less than full dose of losartan) but eventually conducted 2 studies, leading to labeling:

The anti HT effects of candasartan and losartan at their highest recommended doses administered once daily were compared in two randomized, double-blind trials. In a total of 1268 patients with mild to moderate HT [on no other anti HT therapy], candasartan 32 mg lowered systolic and diastolic BP by 2-3 mmHg on average more than losartan 100 mg [at peak or trough effect].

Studies in Non-Responders



Obviously, ability to treat non-responders is very important. You **MUST** randomize to drug and to failed therapy.

Clozapine

Too toxic unless clear clinical advantage

Study in schizophrenics unresponsive to standard therapy

History of poor response to neuroleptics

Diagnosis of schizophrenia, hospitalized

6 week failure on haloperidol

4 week, double-blind comparison of clozapine
vs. chlorpromazine plus benztropine

Results

	Response (%)	
	Clozapine	CPZ
CGI (decrease ≥ 1)	71	37*
BPRS items (dec ≥ 1)		
concept disorganization	60	39*
suspiciousness	64	42*
hallucinations	59	51
thought content	65	40*
CGI and BPRS	15	2*

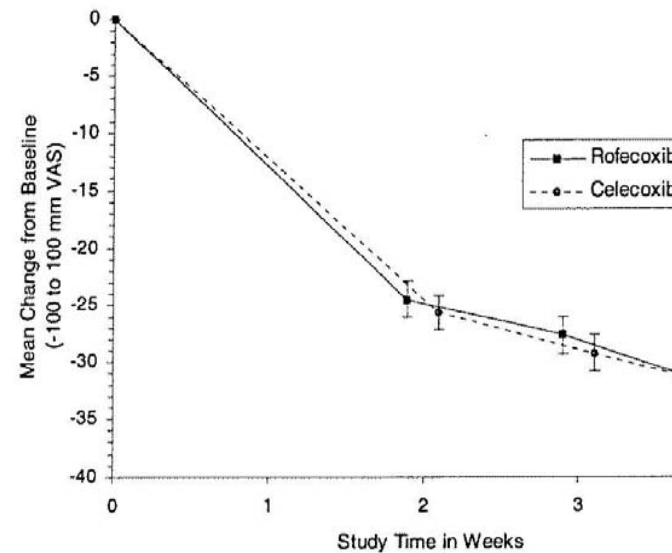
* $p \leq 0.05$

Studies in NRs

It does not always work, though. In discussions of NSAIDs, all arthritis doctors said many drugs are needed because responses are individual. Plausible, but at a COX2 meeting a few years ago I suggested studies in NRs.

Merck did a study comparing rofecoxib 25 mg and celecoxib 200 mg 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.

Study in Intolerants

Chan, et al. J Clin Pharmacol 1997; 37:253-257

Lisinopril (ACE) 10 mg

Metolazone (diuretic) 1 mg

Losartan (angiotensin II antagonist) 50 mg

Patients with ACEI-induced cough

Taiwan and Hong Kong

n=84 elderly hypertensives, non-smokers

Lisinopril re-challenge 8 weeks, at least moderate

Placebo de-challenge 4 weeks, not at all

Randomize to 3 drugs, 10 weeks

Assessment by questionnaire, nurse

Cough

	Lisinopril	Losartan	Metolazone
Cough rate, any	97%	18%	21%

$p < 0.001$

Significant difference in both M,F

Median time to cough 15 days

Studies in NR/Intolerants

These studies are rare, usually because it's the only way to gain approval (clozapine, bepridil). The captopril study was actually designed for the new claim. It took 10 years to do a study showing better sexual function with bupropion than SSRIs, an obvious outcome. Tysabri was never actually compared to interferon in MS.

Perhaps the Merck VIOXX study shows why. The other reason, of course, is that it's unusual for a new drug to be more effective than older therapy.

Superiority

It's not easy

We would generally ask for replication, or a very strong finding, to allow in labeling or promotion.

Comparison must be fair. Conditions should not inappropriately favor a treatment, e.g., by using a less than full dose, an o.d. regimen for a bid drug, patients who had already failed the control therapy. This is considered briefly in ICH E-10.

Difficulties / Limitations

- As noted, real “equivalence” is very hard to show – very large studies.
- Now, with ClinTrials.gov, failure to win will become known.
- I’d strongly recommend studies in non-responders, but then I have been for 20 years to NO effect.
- FDA is not interested in cost-effectiveness under current law.
- We have little/no basis for insisting on the kind of multi-drug comparisons, comparison with old generics, that are of such great interest, as has been done by
 - NHLBI (ALLHAT: chlorthalidone, lisinopril, doxazosin, amlodipine)
 - CATIE (risperidone, olanzapine, perhenazine, quetiapine, ziprasidone)
very interesting result, but not too surprising.
- But we are very interested in helping design them.

Difficulties / Limitations

- Cross-study comparisons almost never credible.
- Epidemiologic data not credible when looking at tiny differences (10-20%).
- One can hope for new ways, such as randomized studies in HMO's, but for most settings, persuasive and informative only if superiority shown.