



# Communication of CER Findings

Robert Temple, MD  
Deputy Center Director for Clinical Science  
Center for Drug Evaluation and Research  
Food and Drug Administration

Asymmetry in the Ability to Communicate CER Findings:  
Ethics and Issues for Informed Decision Making  
Feb 9, 2012

# Overview

1. Regulatory/guidance Limitations on responses (CER or other).
2. Particular problems with CER.

Caveat: some of these issues will need further in house discussion. Specific questions to me, Office of Medical Policy, or Tom Abrams at Office of Prescription Drug Promotion will help.

## Limitations on Promotion

As everyone here knows, drug manufacturers may not promote drugs in ways inconsistent with labeling e.g., by claiming effectiveness for unapproved uses, by minimizing adverse effects, or by claiming unsupported advantages over other treatments. [21 U.S.C. 352; 21 CFR 202.1]

In a variety of ways, however, FDA has tried to write guidance that allows appropriate (not promotional) transmission of additional information:

- Supporting independent educational activities
- Distributing reprints of controlled studies related to unapproved uses
- Responding to unsolicited requests for off-label information

## Support for Scientific and Educational Activities

FDA's 1997 guidance makes a clear distinction between company-directed activities (these must not promote unapproved uses, must not lack fair balance, and may not be false or misleading) and support of independent educational activities, which are not subject to those limitations. The guidance is clear in seeking open discussion in scientific and educational environments (so long as they are actually independent).

## Good Reprint Practices

FDA's 2009 Good Reprint Practices Guidance again reflected the view that distribution of independent (not influenced or controlled by the company) scientific information is a benefit to the medical community. It specifically allows distribution of high quality peer-reviewed scientific or medical journal articles about off-label uses that present results of adequate and well-controlled studies (even including historically controlled studies, PK and PD studies, and well-planned meta-analyses) that are not false or misleading (e.g., at odds with the rest of the evidence) or dangerous.

So, good science can be disseminated (NB, no accompanying promotion).

## Unsolicited Requests for Off-label Information

FDA's 2011 draft guidance on responding to unsolicited requests for off-label intervention recognizes 1) the potential need for information on off-label uses (which can be therapeutically important) and 2) that manufacturers may have important information about these uses. Requests for such information travel many routes, from direct requests via letters, e-mail, or telephone calls to requests put forth in a public forum. FDA's draft guidance describes various approaches that industry can follow so that responses to such requests will not be considered evidence of an intent to promote off-label use.

I.e., its OK, if...

## Unsolicited Requests (cont)

### OK, if

1. Not solicited (guidance illustrates many kinds of solicitation)
2. Responses are truthful, balanced, non-misleading, and non-promotional and respond to the specific request
3. For non-public requests,
  - Respond only to the person asking
  - Answer only the question asked
  - Rely mostly on published journal articles & texts
  - Avoid promotional tone
  - Scientific, not sales personnel
  - Include labeling and important safety information
  - A statement that the drug is not approved for the off-label use

## Unsolicited Requests (cont)

4. For public requests, where responses would go to a large audience, we suggest

- Respond only when request is about the firm's product, not solely a competitor's product
- Limit public response to contact information, i.e., not a broad dissemination of off-label information (it then becomes a non-public request)



## FDAMA 114

Finally, FDAMA 114 allows manufacturers to provide healthcare economic information not in labeling to formulary committees and similar entities. This is a limited, but potentially important exception that has some potential relation to CER.

# FDAMA 114

- FDAMA 114 amends FFDCA section 502(a)
  - Defines health care economic information (HCEI) as an analysis that identifies, measures, or compares the economic consequences of the use of the drug to the use of another drug, another healthcare intervention, or no intervention
  - Restricts the audience of HCEI to formulary committees or similar entities
  - States that HCEI will not be considered false or misleading if it is based on competent and reliable scientific evidence (note, different from adequate and well-controlled studies)
  - States that HCEI will not be considered false or misleading if it directly relates to an approved indication

## FDAMA 114

- This provision is not intended to provide a path for promoting new off-label indications or claiming clinical advantages of one drug over another when these claims do not satisfy FDA's evidentiary standards for the claims being made.
- E.g., HCEI based on prolonging patient survival would not be considered appropriate under this provision for agents approved for the symptomatic treatment of heart failure, but not approved for prolonging survival in heart failure patients.

## FDAMA 114

It is apparent that economic considerations are unavoidably related to clinical effectiveness. Thus we interpret FDAMA 114 as follows:

- **Economic consequences directly related to labeled effect of the drug**
  - Must be supported by the approved product labeling
  - Any comparative claims must be supported by substantial evidence that directly compares the treatments in question (i.e., head-to-head clinical trials that provide a valid comparison of two drugs)

# FDAMA 114

- **Economic consequences that follow from the labeled effectiveness**
  - Can be supported by competent and reliable scientific evidence related to the consequences of an effect once the primary clinical outcomes have been established
  - E.g., Modeled economic consequences of prevention of fractures due to osteoporosis would be permitted for drugs approved for prevention of fractures due to osteoporosis because the modeled data may be considered competent and reliable for making an assumption about the secondary consequences of an osteoporotic fracture. Similarly, a treatment that provides tight control of blood sugar (insulin) in type 1 diabetes could have cost savings attributed to prevention of retinopathy if well-controlled studies show such an effect of tight control.

## FDAMA 114

- **Therefore, supporting evidence for HCEI depends on which component of the analysis is involved:**
  - Establishing the economic costs and consequences used to construct HCEI would be based on standards widely accepted by economic experts
  - Establishing the clinical outcome assumptions used to construct HCEI would be based on the evidence of effectiveness used to support effectiveness

## Bottom Line

In a number of areas, the ability of manufacturers to provide information they have is supported and recognized, so long as promotional behavior is avoided. We do not, that is, want an open door for promoting unsupported claims.

Apart from the guidance described, companies also support clinical trials whose results are published and regularly announce, in news releases, results that excite them. Even here, the report needs to avoid promotional elements. They also regularly rebut publicly findings, e.g., adverse epidemiologic results, they do not agree with. Again, they cannot contradict labeling or promote off-label uses.

# Rebuttals and Claims

Let me now turn to the subject at hand; the “asymmetry” between what researchers, payers, or the government can say about CER findings and what a sponsor, whose “communications are highly regulated by the FDA,” can say. It is suggested that the asymmetry can leave companies “unable to respond to erroneous communications made by others about their products.” There seems to be particular concern about the following possibilities:

1. A government conclusion (AHRQ or other, VA perhaps) that a drug is inferior or has a major disadvantage or is “not worth it.” Specific illustrations would help.
2. A published meta-analysis or observational study reporting a risk or other disadvantage of a drug. A recent example also might be newspaper assertions that antidepressants have no long-term benefits and really don’t work.

I will also consider what support is needed for a comparative claim, not quite the topic, but an interest of mine.

## Rebuttals

As the illustrations I've cited show, there is no FDA view that drug companies are condemned to silence about their products outside of formal promotion or perhaps published articles. Moreover, anyone reading newspapers can see that companies provide a wide range of public "disclosures" about how a study came out, how an FDA meeting went, what FDA's letter on an NDA said (without ever publishing the letter, by the way).

It seems clear that a sponsor could "correct," or dispute, a CER statement by a payer, researcher, or even government agency, so long as the correction was not itself promotional.

## Rebuttals (cont)

I don't know if there are specific government or payer CER statements people have in mind, and details matter, but in recent months we have seen companies disagree with meta-analytic or epidemiologic conclusions they considered unsupported on methodologic or other grounds and we certainly would not deny them that right. (note, again, making their own comparative claim would be quite another matter).

## CER Claims

After effectiveness, there is nothing of greater interest to physicians and patients than how one treatment compares with alternatives in effectiveness and safety. That said, there are surprisingly few areas, especially considering symptomatic conditions, where one drug has been shown more effective than others, e.g.

- Not for anti-depressants
- Not for NSAIDs
- Not for AEDs (although some probably are more effective)

But it can be done, and it is fairly common where clinical studies of a new drug must compare the new with the old (e.g., in oncology, anti-coagulants, anti-platelet drugs).

It is also unusual to see a showing that one member of a drug class has less of a class side effect; e.g., do some antidepressants cause less sexual dysfunction than others? Yes for bupropion; no other data.

## CER (cont)

There are a few within-class claims (some ARB's have candesartan and larger effects than some other ARBs), but it took well over 1000 patients to show such differences because differences are usually small.

But might one member of a class, or a drug of a different class, work in people who fail on another drug? That is a critical question for physicians.

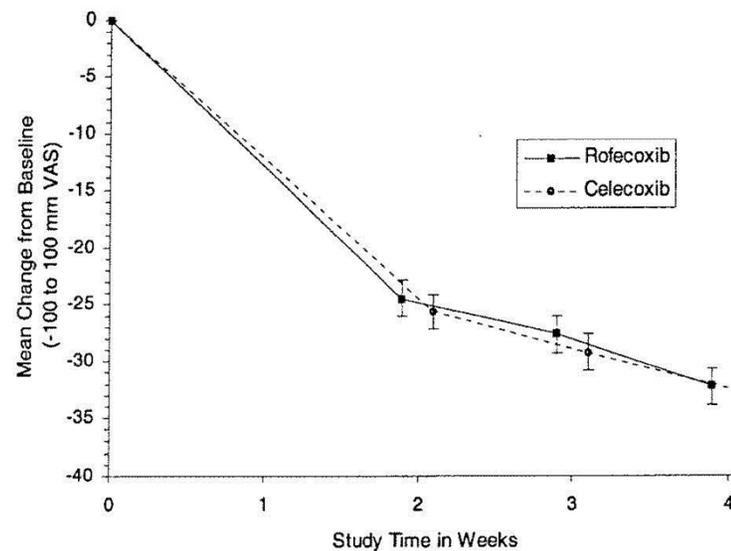
## Superiority in a Subset (Non-responders)

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.

Restricted  
Confidential  
Limited access

**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.

## Non-Inferiority Studies

We are seeing far more studies that compare one drug with another, not to show one drug is better but to show “non-inferiority” i.e., to show that a new drug is “close enough to the established drug to allow a conclusion that the new drug works.

# Non-Inferiority Studies

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 of the control 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 the control effect is of perceived great value, we usually ask that some fraction of the control effect be preserved (typically 50%). But showing you've kept at least 50% of the control would certainly not support a claim of "equivalence."

This IS a kind of relative effectiveness requirement; the new drug cannot be TOO inferior.

Those NI studies rarely allow a CE claim, unless the new drug WINS; then the claim goes into labeling.

# Superiority Claims

Superiority claims have been sought and we would allow them when our approval standard as been met: 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; or the comparator was used inappropriately.
- 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).

An unfair comparison might well deserve a rebuttal.

## 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.

What this suggests to me is that manufacturers may well have a basis for rebutting CER claims in many cases.

## 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).

Failure to detect differences does not mean they do not exist.

# 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 one:

### Hormone replacement therapy

The incorrect results of epidemiologic studies in this case, 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 8506	Placebo 8102	HR	95% CI
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.

## Comparative Effectiveness – Difficulties

Sample size needs to be 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.

## CER and Comment

Again, without specifics, it's hard to be precise, but the current hunger for CER will surely lead to use of data that are often not reliable at detecting small differences between drugs

- Epidemiologic studies
- Meta-analyses
- Cross-study comparisons

I think it is likely that manufacturers will have reason to comment on such comparisons and, on its face, that does not seem promotional and would not be barred in any way.