Past, Present and Future of Product Safety Assessments

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

4th Seattle Symposium in Biostatistics:
Clinical Trials
Nov 23, 2010

Evolution of Risk Concern

Developers and reviewers of drugs, users of drugs, and patients have always been aware that drugs have risks/adverse effects, sometimes serious ones. Sometimes such serious risks are common and readily detected, "part of the deal," if you like, for serious risks that accompany use of cytotoxic chemotherapy, many anti-virals, and anti-coagulants. Common but less serious adverse effects are recognized for most drugs. In these cases, the known risks can be weighed against the known benefits.

But in most cases serious risks are not recognized early but turn up later, generally surprising us. This often provokes a concern that we didn't know about them sooner. The question increasingly being asked is whether there is something different we should be doing to catch them earlier so that we can factor them into our risk-benefit assessment or at least rule out risks of excessive size.

What Kinds of Risk Are There?

Risk is, of course, always balanced against benefit and even severe risk can be accepted where benefit is large and there is no safer alternative or when risks are very uncommon:

- Life-threatening effects of many oncologics, antivirals, and inflammatory disease modifiers
- Clozapine agranulocytosis, and bepridil TdP for angina therapy failures
- Fatal GI bleeding with NSAIDs (attempts to lower that risk with COX-2 selective agents perhaps led to a new concern).
- Coumadin bleeding; antibiotic superinfections and diverse other toxicities
- And on and on (read labels)

How Do We Find Serious Risks?

1. Severe, not so rare, obviously drug-related

   Serious adverse events that generally do not occur spontaneously can be detected in trials if common enough, and are acceptable, as noted, for drugs that have important benefits or no alternatives that don't have the effects

   - Oncologic
   - Anti-inflammatory TNF inhibitors
   - Clozapine
   - Thrombolytic bleeding; anticoagulant bleeding

These are generally seen during development, especially in pooled ISS analyses

How Do We Find Serious Risks?

2. Rare and severe, obviously drug-related, because little or no spontaneous occurrence

   Less common serious risks can sometimes be detected in clinical trials if they are single case interpretable (liver injury, ARF, Stevens-Johnson, rhabdomyolysis, hematologic, TDP) or anticipated if there is a good surrogate (Hy's Law, marked QT prolongation), but often they have not been detected pre-marketing, at least with current databases, which are, however, growing in size. Historically, these have been the main reason for drug withdrawals and major limitations.

   These are, generally, readily detected and attributed post-marketing, because they are far more common than the background rate, with risk ratios probably > 100, or even > 1000, as the many drug withdrawals for liver injury, TdP, and SJS indicate.

How Do We Find Serious Risks?

3. Events that occur spontaneously fairly often, that are not obviously drug-related, but that are greatly increased in rate by drug

   There have been serious events that are not easily interpretable as single or small numbers of events (DVT, PE on OC's; valvulopathy with fenfluramine or pergolide; endometrial cancer from unopposed estrogens) because the events are relatively common in the untreated population. But if the HR is large, HR > 4-5, they can be reliably detected in an Epi study, as all of the above were.

How Do We Find Serious Risks?

4. Events that occur spontaneously but that are only modestly (less than 2 fold) increased by the drug.

   Typically these are CV events, but they could be tumors, adverse outcomes of the disease (increased asthma deaths in people treated with long-acting beta agonists).

   Epi data are uncertain in these cases, especially when risk is 1.5 fold or less (I acknowledge different views on what the cut-point is), but even a modest increase (say 10-20%) would be of great concern given the wide use of the drugs, the existing rate of occurrence of the event, and the severity of the event.

   But large enough clinical trials can detect such effects and in some cases such trials are conducted pre-marketing or early post-marketing. There is new and growing interest in this kind of effects, a real change in focus.

How Do We Find Serious Risks?

The next few slides show 30 drugs withdrawn for safety reasons (some returned, usually with a RiskMAP), mostly since 1962 and through 2007. This can be considered one measure of the risks we worry most about, although a listing of drugs never approved because of similar concerns would also be of interest and would include hepatotoxicity (dilevalol, tasosartan, lumiracoxib, and ximelagatran) and QT prolongation (sertindole and others).

Plainly, most WDs have occurred because of uncommon (most well < 1/1000 and many 1/10,000 or less). Usually, this was unacceptable because of the severity of the ADR and availability of alternatives without such risk.

Of about 30 withdrawn drugs, only 5 had a problem identified by epidemiologic studies (phenformin, PPA, fenfluramine, cerivastatin, pergolide), identified in the tables as type 2 data, sometimes with prior individual case "hints" and only 4 had the problem identified by good-sized controlled trials (encainide, flosequinan, rofecoxib), identified as type 3 data, or a meta-analysis (tegaserod) also type 3.

History of Drug Withdrawals

Drug	Date App'd	Date WD	Data: 1, 2, 3	Adverse Effect
iproniazied (Marsilid)	1950 (?)	1956	1	DILI
azarabine (Triazure)	1977	1977	1	Arterial thrombosis
phenformin		1978	2	Lactic acidosis
ticrynafen (selacryn)	1979	1980	1	DILI
Benoxaprofen (Oraflex)	1982	1982	1	DILI
zomepirac (Zomax)	1980	1983	1	Anaphylaxis
methaquolone (Qualude)	1960's	1984	1	OD very hard to treat
nomifensine (Merital)	1984	1986	1	hemolytic anemia
suprofen (Suprol)	1985	1987	1	ARF
** encainide (Enkaid)	1986	1991	3a	CAST; mortality (HR=2)
temafloxacin (Omniflox)	1992	1992	1	Hemolysis, renal failure
flosequinan (Manoplax)	1993	1992	3a	Mortality (HR – 1.5)

* 1= individual cases
2 = epidemiologic data
3 = RCT's: 3a large trials; 3b MetaA
** NOT withdrawn, but limited
*** Returned to market

History of Drug Withdrawals

Drug	Date App'd	Date WD	Data: 1, 2, 3	Adverse Effect
fenfluramine (Pondimin)	1973	1997	2	Valvulopathy
terfenadine (Seldane)	1985	1997	1	TdP
mibefradil (Posicor)	1998	1998	1	Interactions; TdP
bromfenac (Duract)	1997	1998	1	DILI
** trovafloxacin (Trovan)	1997	1998	1	DILI
astemizole (Hismanil)	1988	1999	1	TdP
grepafloxacin (Raxar)	1997	1999	1	TdP
troglitazone (Rezulin)	1997	2000	1	DILI
cisapride (Propulsid)	1993	2000	1	TdP
*** alosetron (Lotronex)	2000	2000	1	ischcolitis; constip'n
PPA	<1962	2000	2	hemorrhagic stroke
rapacuronium (Raplon)	1999	2001	1	bronchospasm
cerivastatin (Baycol)	1997	2001	1, 2	rhabdomyolysis

* 1= individual cases
2 = epidemiologic data
3 = RCT's: 3a large trials; 3b MetaA
** NOT withdrawn, but limited
*** Returned to market

History of Drug Withdrawals

Drug	Date Approval	Date WD	Kind of Data: 1, 2, 3	Adverse Effect
levacetyl methadol (orlaan)	1993	2003	1	TdP
rofecoxib (VIOXX)	1999	2004	3a	AMI
*** natalizumab (Tysabri)	2000	2005	1	PML
pemoline (Cylert)	1975	2005	1	DILI
valdecoxib (Bextra)	2001	2005	1	Stevens-Johnson
pergolide (Permax)	1998	2007	2	Valvulopathy
tegaserod (Zelnorm)	2002	2007	3b	CV events

Serious, Rare Risks

It is possible that larger databases, and perhaps, but not necessarily, large controlled trials, could have detected some of these toxicities earlier (especially hepatotoxicity, now that we know the clues to look for), but if the adverse reactions are in the 1/> 3500 or so rate, that is probably not realistic.

What Worries Us Now?

As noted, in the past, our main concern was primarily the deadly rare events, and spontaneous reporting systems, especially with recent improvements, have become very good at detecting them. Bromfenac and troglitazone hepatotoxicity and mibefradil interactions with simvastatin leading to rhabdomyolysis were detected very swiftly (a few months).

But our interests have broadened to include the less self-evidently drug-related events because they are numerically far more important.

How Do We Find Serious Risks? Epidemiologic Studies

There are, as noted, many safety problems detected by epidemiologic studies, useful where risks are relatively large

- DVT/PE with OC's, including lower rates with lower estrogens.
- Trasylol outcome results, still under active discussion, of increased CV risk.
- Endometrial Ca with unopposed post-menopausal estrogen, now not used in women with intact uterus.
- Valvulopathy with fenfluramine.

The difficulty in detecting many of these events is that they need relatively long exposure to occur or to occur in sufficient numbers, longer than would be typical of controlled trials in most development programs. But given the large increased risk, trials of familiar size could probably detect them, if we were to have more long-term controlled trial exposure, as opposed to current open-label extensions.

Serious Risks – Small Increase

Although trials of familiar size, perhaps longer, might detect markedly increased risk, in recent years we've become concerned about a far more challenging set of concerns: the possibility that drugs, most of them intended for long-term use, often to "prevent" or reduce risks, usually CV risk, might instead increase that risk or some other significant risk.

And the level of increase of concern is not the 2-fold, 3-fold risk of the past, risks that might be detected in post-approval epidemiologic studies, but far more modest increases (e.g., 10-20%) increases detectable only in randomized trials, or possibly pooled analyses of RCTs, of far greater size and duration than those of the past. Indeed, what these are is really "outcome trials," the same kinds of trials needed to detect small benefits.

How Do We Find Serious Risks? Randomized Trials

We have, of course, detected such adverse effects in the past, but several recent situations, both pre and post-marketing, have greatly increased interest:

- COX-2 selective NSAID and NSAIDs generally, CV risk
- Rosiglitazone and anti-diabetic treatments, more generally, CV risk
- Erythropoietin problems, both CV risk and tumor promotion
- Results with long-acting beta agonists, exacerbation of asthma decompensation
- The WHI, estrogen/progesterone replacement, CV and carcinogenicity
- The torceptrapib experience, CV risk (pre-marketing)
- Meta-analyses of antidepressants and AEDs showing increased suicidality

Before reflecting on where this new interest might lead, I want to briefly review the past history of randomized trials that showed adverse effects and some of their difficulties, considering some specific cases.

Outcome Trials in CV Disease

Most of the RCTs that have shown or looked for adverse CV effects have been studies of CV diseases, or less commonly, an examination of CV effects in non-CV disease

1. In most cases there was an attempt to show an outcome benefit, generally because there was a known favorable effect on a surrogate endpoint. That is, there was no adverse hypothesis.

2. In some cases there was an existing safety concern about a drug that seemed valuable (e.g., with clear short-term benefit, strong suspicion of longer term benefit, but short of proof); in some of the cases the CV concern that arose was a complete surprise.

First consider attempts to show a benefit. In these cases, of course, an adverse outcome is a more or less complete surprise, but conducting the study raises no ethical dilemma.

Outcome Trials in CV Disease/DM

A. Attempts to show benefit

1. UGDP

The UGDP, an NIH trial conducted in the late 60's, sought to show the value of improved glucose control with insulin given as fixed or variable doses and the sulfonylurea oral hypoglycemic tolbutamide, compared to placebo. It showed an adverse effect of tolbutamide on cardiovascular deaths compared to the other treatments, a point made in labeling for all sulfonylureas since then, although the results have been challenged and are controversial.

2. Subsequent trials in DM

Numerous studies in type 2 Diabetes (recently stopped ACCORD – NIH trial, many other commercial, European and U.S. trials) have mostly shown no benefit on CV outcomes. But ACCORD reported that very tight control gave a roughly 25% increase in mortality.

Outcome Trials in CV Disease

A. Attempts to show benefit (cont)

3. CAST

Cardiac Arrhythmia Suppression Trial (CAST), an NIH trial, was intended to show improved survival compared to placebo in post-infarction patients with ≥ 6 VPB's/hour in patients who responded to encainide or flecainide with substantial reduction of VPB's (median response 100%).

Result was > 2X mortality risk in the treated patients.

Going in conviction of benefit was so strong that some questioned the ethics of the trial.

Outcome Studies in CV Disease

A. Attempts to show benefit (cont)

4. NSAIDs

The studies of COX-2 selective NSAIDs that were most suggestive of increased CV risk were designed to show

- Reduced risk of major GI bleeds (VIGOR, TARGET, EDGE). This could be considered an attempt to validate the documented surrogate effect of decreased endoscopic ulcers.

- Reduced rate of colon polyp formation (APC, PreSAP trials of celecoxib; APPROVe trial of rofecoxib) or improvement in Alzheimer's Disease (ADAPT trial of celecoxib)

It is of note that it was the relatively older populations in all these trials that made CV outcome assessment possible.

Outcome Studies in CV Disease

A. Attempts to show benefit (cont)

5. Erythropoietin

Studies were intended to show what everyone believed – that titrating people to higher hemoglobin levels would be good for them, as epidemiologic data suggested. Instead we've seen unfavorable effects on tumor progression, stroke, and survival, thus far not well-explained.

Outcome Studies in CV Disease

A. Attempts to show benefit (cont)

6. Comparative benefit = comparative harm, doesn't it?

a. LIFE study

losartan gives less stroke than atenolol. Implications?

b. The ACCOMPLISH (NEJM, 2008) a trial compared amlodipine (5-10 mg) and HCTZ (13.5-25 mg), each added to benazepril. Amlodipine had fewer events (MACE & CV hospitalizations), 11.8% vs 9.6%, a 20% risk reduction for MACE alone, HR was 0.79 (p=0.002). BP control was very similar in both groups.

c. Dabigatran (RE-LY); RE-LY showed significant reductions in both ischemic and hemorrhagic stroke for dabigatran 150mg vs Coumadin.

Outcome Studies in CV Disease

B. Attempts to resolve concern because of the nature of the drug class where clinical benefit had been demonstrated (probably the case that is closest to the current discussion).

For two classes of CV drugs (inotropic drugs for CHF and antiarrhythmics generally), there was concern about possible adverse effects on survival. The Division of Cardiorenal Products has therefore required, for more than 2 decades, that these drugs be studied to assess CV risk. This was possible because the event rate in CHF and some antiarrhythmic settings was high and because an important benefit was expected (ethical consideration).

1. CHF

Early experience with beta-agonists raised concern about inotropes (or at least some inotropes) worsening outcome. Trials of various sympathomimetics, whether direct (beta agonists, dobutamine) or indirect (PDE inhibitors including milrinone, flosequinan, vesnarinone) showed adverse survival effects. Again, note that the very high risk patients allowed relatively small trials to be of adequate size.

Promise

	Placebo	Milrinone
n	527	561
Total mortality	127 (24%) P = 0.038, nominal P = 0.057, adjusted, early stop	168 (30%)
Total CV mortality	119 P = 0.016, nominal P = 0.037, adjusted	165

Packer, et al, N Eng J Med 325: 1468-1475 (1991)

PROFILE

Flosequinan vs Placebo in CHF

	Flo 75	Plbo	Flo 100	Plbo
Mortality	40/206	43/238	201/964	138/937
RR	1.05	1	1.48	
CI	.68/1.62		1.19-1.84	
P-value	0.8254		0.0004	

Flosequinan unequivocally improves exercise Capacity and CHF symptoms in NYHA Class II, III

Not tested in Class IV (can't exercise)

Outcome Studies in CV Disease

B. Concern over CV effects (cont)

2. Anti-arrhythmics

Post-CAST we have sought assurance, in at least some setting, of lack of harm.

DL sotalol for preventing AF recurrence had the post-AMI Julien trial (NS, but 18% reduction of mortality) while dofetilide had 2 "Diamond" studies showing no adverse outcome in CHF and post-infarction. Dronedarone showed increased mortality in patients with recent CHF exacerbation but the large effectiveness trial supporting approval (ATHENA), which excluded such patients, showed a favorable survival trend and a highly significant reduction in cardiovascular hospitalizations.

Outcome Studies in CV Disease (cont)

3. Other drugs/classes – interest has moved to other kinds of drugs

For antihypertensives, approved wholly on the basis of BP, there has been no requirement for long-term large outcome studies before or after approval. Depending on your view, ALLHAT either supports or challenges that position. Such trials, however, are extraordinarily difficult, as all involve multiple drugs. Placebo-controlled trials with a wide range of drugs have shown clear benefit.

NSAIDs currently under development are regularly the subject of large outcome studies, hardly surprising. Placebo-controlled trials are not possible in symptomatic patients, so all are active-control trials, a problem because the risks of the possible active control drugs are not well-defined.

Lipid-lowering drugs have been approved based on LDL cholesterol effects, but for other effects (HDL raising) outcome or plaque effect studies have been expected. The torceptrapib experience will almost surely strengthen that expectation (for outcome studies).

Outcome Studies in Other Classes

3. Other drugs/classes (cont)

Tegaserod (for IBS, constipation) was the subject of a meta-analysis (requested by the Swiss) of short-term RCT's, revealing large excess of CV events. The drug was withdrawn. What might the implications be for GI motility – modifying treatments, at least some of which involve long-term, or at least recurrent use?

The most famous meta-analysis (? ever) was of rosiglitazone. Various analyses included about 40-50 short-term trials and some (Nissen) added longer-term ADOPT and DREAM. The analysis suggested an adverse effect on heart attacks (non-fatal) and, less strongly, an effect on survival (The latter was not present in ADOPT, DREAM, and later, in RECORD).

What about Other Drugs/Classes

There are several recent suggestions from FDA and elsewhere (in addition to need for outcome data on drugs for heart failure and anti-arrhythmics)

1. Antidiabetics

   In commenting on the rosiglitazone meta-analysis and other studies (of pioglitazone, epidemiologic studies) that may, or may not, have offered a convincing case for increased AMI, Psaty and Furburg, Rosen, and Nathen, writing in the NEJM, said that the surrogate endpoint of HbA1c is inadequate and that outcome studies are needed. They are not clear on this but, as no drug, including insulin, has convincing outcome data in type 2 diabetes, insistence on a favorable CV outcome would eliminate all treatments for type 2 diabetes, including oral hypoglycemics and probably insulin as well.

What about Other Drugs/Classes

What they really meant, I think, was that FDA should require a large study, like ADOPT, RECORD, DREAM, and PROACTIVE (of pioglitazone), before, or as a condition of, approval, to rule out, or cap, the cardiovascular risk of the drug, a possibility raised by Joffe, et al of FDA in a NEJM letter on the rosiglitazone situation, and now established policy in FDA guidance.

Other Drugs/Classes

1. Antidiabetics (cont)

Following a July 2008 meeting of the Endocrinologic and Metabolic Drugs Advisory Committee, FDA in December 2008 issued a final guidance calling for

- Pre-marketing demonstration, generally based on pooled data, that there is not an unacceptable increase in CV risk, indicated by an upper bound of the 95% CI for risk of less than 1.8 (but watch the point estimate too).

- Post-approval demonstration, obviously in a larger database, that the 95% CI upper bound for CV risk is less than 1.3. This will be a post-marketing requirement under FDAAA.

Other Drugs/Classes

1. Antidiabetics (cont)

The studies are plainly demanding but not so daunting if patients at higher risk (with many events) are included in trials.

For a class of drugs directed at a population with high CV risk, this requirement seems hard to rebut. Of course:

1. The 1.8 and 1.3 are arbitrary (but reasonable, I think, and similar to the goal in the large PRECISION trial of celecoxib, ibuprofen, and naproxen). The 1.3 value, of course is the 95% upper bound of CI. The point estimate will matter too.

2. What about other chronic drugs?

Other Drugs/Classes

2. Brass, Lewis, Lipicky, Murphy, and Hiatt proposed some years ago in CP and T that for symptomatic CV treatments (where outcome is not part of development as it would be for, say, antiplatelet therapy) sufficient data should be obtained to rule out some upper boundary of risk, perhaps 50% (HR < 1.5). This conclusion would presumably also apply to approvals based on a surrogate endpoint, and seem applicable to a wide range of chronically used drugs.

   Brass, et al, recognized many difficulties.

   - Long-term placebo-controlled trials in symptomatic patients will not be possible, leaving only comparative trials available; even in outcome trials, placebo controls will often not be possible.
   - Trials of realistic size require high risk people to gain enough endpoints and will almost surely require a combined endpoint (death, AMI, stroke and perhaps more, like unstable angina, CHF, etc) but that may not be what one is worried about (CHF drugs, CAST, do not show increased AMI or stroke). So the size must be still greater if there is only one endpoint of interest or if the population is healthier.

3. Other drug classes: asthma drugs (suggested by recent LABA experience); chronic GI drugs, weight loss drugs, anti-rheumatics, etc.

Outcome Trials for All Policy and Problems

1. Policy

It will be no surprise to you that I am not announcing an FDA policy on the universal need for outcome studies. What you can clearly see is that we are prepared to ask for such studies when:

- There is a suggestion of risk for the particular drug.
- There is concern about the class (NSAIDs, antidiabetics, anti-arrhythmics, drugs for CHF, or at least inotropes).

When else such studies would be reasonable is yet to be determined.

A minimum interim step that should be considered is to have the current level of expected long-term data for drugs (ICH E1: 300-600 for 6 months, 100 for a year) be from controlled trials, not open label extensions, as is now common. This, of course, would detect only major increases.

Policy and Problems

2. Issues and Problems

- Unless the population is high risk, CV studies will have little chance of showing anything unless they are enormous.
- In many cases, studies will need to be long, OK for outcome studies, very hard in symptomatic conditions, where, even in active control trials, dropouts are very common.
- In symptomatic settings, placebo control is unrealistic and risk of potential active controls is unevaluated (usually).
- If study is to follow up on even a weak suggestion of harm, ethical concerns have been raised, at least for drugs without unusual benefits. That has not stopped PRECISION, but TIDE is on hold.
- The large outcome trials are taking years, even decades to be conducted. Would a higher upper bound be an acceptable trade off for a quicker answer?

Policy and Problems

- We all know epidemiologic study results can vary, but recent experience shows even RCT results vary when effects are small. How sure of the results will we be able to be?
  - ASA secondary prevention: largest trial (AMIS) shows no effect and survival goes the wrong way; iib/iiia inhibitor trials in ACS vary from 0-50% reduction in death plus AMI at 30 days.
  - Of 3 placebo-controlled celecoxib trials of similar size, one showed a 2+ fold excess risk, one showed no risk at all, and the 3rd showed a small increased risk but numerically smaller than the risk of naproxen, thought by many to have least risk.
  - A fairly strong mortality effect in the short-term studies in the Nissen meta-analysis of rosiglitazone was not seen at all in ADOPT, DREAM, and RECORD. (RECORD is reported to have leaned favorably on mortality and, apparently, stroke, but further review is anticipated). For AMI's in that meta-analysis, in contrast, short-term studies did not show a significant effect and only by adding in the larger studies (ADOPT and DREAM) did a significant effect emerge. (As noted, RECORD is being analyzed further). None of this is to suggest what answer is correct or to suggest poor analytic process, but it does show the problem of selection in meta-analyses as well as the problem of variable results.

Other Problems

- Value, cost-benefit
  - Requiring outcome data, not surprisingly, can affect development. Antiarrhythmic policy is "associated with" with minimal anti-arrhythmic development (implanted defibrillators may have influenced also) and CHF treatments other than ACEIs, ARB's, eplerenone are hard to spot.
  - How likely, absent an animal or human signal, is a bad outcome? If these are very hard to detect, and detection is uncertain, how worthwhile is it compared to other important questions (how low to drive BP, LDL cholesterol) that also have life and death implications?
  - There is no doubt that expanded ability to conduct large trials (e.g. in HMO-type environments) would greatly enhance our ability to do such trials.

Conclusions

There is little doubt that there is a new interest in conducting studies to detect possible modest adverse (or beneficial, of course) effects of chronic-use drugs. Interest has spread from cardiovascular drugs (where it has long been present because of experience with anti-arrhythmics, various inotropes) to other chronic-use drugs, including anti-diabetics, NSAIDs, and chronic-use asthma drugs. Of course, some drugs have their effectiveness evaluated in long-term studies (anti-platelet drugs, bisphosphonates and other bone-preserving agents, adjuvant chemotherapy) that are of substantial size. These studies can already detect an adverse long-term effect, at least if the right population is studied.

There is no doubt that this issue will be the subject of much discussion.