Clinical Trial Efficiency

There is broad recognition that the costs of clinical trials are growing and concern that this will limit our ability to get the information we need about the effectiveness and safety of treatments, including both the effectiveness and safety of novel drugs and the comparative data that is very much on people’s minds.

The clinical community is therefore thinking of a variety of ways to make trials more efficient:

- Adaptive designs
- Collecting only critical information
- Better targeted monitoring
- Carrying out trials in healthcare environments, making use of already-collected data

Today, I will talk about a major contributor to efficiency, the use of a variety of methods that improve study power, specifically the likelihood of showing a drug effect if there is one, by choosing the right patients for the trials.

Enrichment

We don’t do clinical trials in a random sample of the population. We try to make sure people have the disease we’re studying (entry criteria), have stable disease with stable measurements (lead in periods), do not respond too well to placebo (placebo lead in periods), have disease of some defined severity, and do not have conditions that would obscure benefit. These efforts are all kinds of ENRICHMENT, and almost every clinical trial uses them. There are, in addition, other steps, not as regularly used, that can be taken to increase the likelihood that a drug effect can be detected (if, of course, there is one).

In December 2012, FDA published a draft guidance: Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products.
Enrichment

Enrichment is the prospective use of any patient characteristic – demographic, pathophysiologic, historical, genetic, and others – to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population.

This occurs to a degree in virtually every trial, although enrichment may not be explicit, and is intended to increase study power in 3 principal ways, by:

- Decreasing heterogeneity (noise); choosing an appropriate population, i.e. patients who definitely have the disease
- Finding a population with many outcome events, i.e., high risk patients, or patients with relatively severe disease – prognostic enrichment
- Identifying a population capable (or more capable) of responding to the treatment – predictive enrichment

The increased study power facilitates “proof of principle” (there is a clinical effect in some population) but, depending on the specific enrichment mechanism used, it can leave open 1) the question of generalizability of the result and how the drug will work in other populations, as well as 2) the question of how much data are needed before or after approval in the “non-selected” group.

Enrichment Designs

Enrichment designs sometimes make people nervous and cause them to wonder about generalizability. With empiric designs, e.g., doing studies in people who respond to an open screen, there really is no way to identify the responder population; you just know that there is one. In some cases, the remedy is to:

- Use these designs early, to show unequivocal drug effect
- Don’t make the enrichment study the only study, at least not usually
- Be aware of what you’ve done and don’t hide or overstate results

But it is more and more recognized that the selected population is in fact the one where treatment makes the most sense. After all, results in an unselected population may be driven by a subset of the population; you just never know about it.
The guidance is focused on studies intended to demonstrate effectiveness but it is also pertinent to safety studies.

- In the studies of oral hypoglycemics to rule out CV risk, we recognize the need to include high risk patients to have any chance at success (prognostic enrichment).
- One could show a drug lacks a class adverse effect by studying people who had the effect on another member of the class; enriching the population for likelihood of having the AE on the control and facilitating a showing of a difference if there is one (predictive enrichment). Note that this is an enrichment that assesses comparative safety.

Kinds of Enrichment

1. Decreasing heterogeneity – virtually universal: A variety of practical steps to decrease heterogeneity (noise) are often used and include:
   - Define entry criteria carefully to be sure patients have the disease being studied
   - Find (prospectively) likely compliers (VA hypertension studies; Physicians’ Health Study)
   - Choose people who will not drop out
   - Eliminate placebo-responders in a lead-in period
   - Eliminate people who give inconsistent treadmill results in heart failure or angina trials, or whose BP is unstable
   - Eliminate people with diseases likely to lead to early death
   - Eliminate people on drugs with the same effect as test drug

   In general, these enrichments do not raise questions of generalizability

Kinds of Enrichment (cont)

Apart from efforts to decrease heterogeneity, enrichment strategies fall into two distinct types:

2. Choosing high risk patients, i.e., those likely to have the event (study endpoint) of interest – prognostic enrichment.

   This has study size implications, of course, but also therapeutic implications. A 50% change in event rate means more in high risk patients (10% to 5%) than in low risk patients (1% to 0.5%) and could lead to a different view of a drug’s toxicity.

3. Choosing people more likely to respond to treatment – predictive enrichment.

   Choices could be based on patient characteristics, (pathophysiology, proteomic/genomic) or be empiric, based on patient history of response to similar drugs, early response of a surrogate endpoint (e.g., tumor response on some radiographic measure), or past response to the test drug (randomized withdrawal study), discussed further later.
Past Selection of High Risk Patients (Prognostic Enrichment)

Although the information distinguishing individuals with respect to risk is growing exponentially, we’ve had such information before:
- Epidemiologic risk factors for cardiovascular outcomes
  - Recent events (AMI, stroke)
  - History of angina, TIA, PAD
  - Cholesterol, blood pressure levels
  - Diabetes and other concomitant illness
  - Elevated CRP (JUPITER Study of rosuvastatin)
  - Family history
  - Gender, race, age
- Individual measurement/history in CV, cancer, and other outcomes
  - Vascular injury on angiography, ECHO findings
  - Tumor histology

Enrichment – High Risk Patients

In one way or another, it is routine to try to find people at high risk so that an intervention will have events to prevent. This is common in both oncology and CV medicine and there are growing possibilities:
- Breast or ovarian cancer prevention in people at high risk
- Outcome studies of lipid-lowering agents (hx of AMI, very high LDL, cholesterol, low HDL, elevated CRP)
- Studies of anti-platelet therapies in angioplasty patients

There is great potential for pharmacogenomically or proteonomically identifying high risk patients, e.g., in Alzheimer’s Disease, various cancers. Not so clear yet in CV disease.

When these methods are used, there is always a question about the effects and benefit/risk relationship in lower risk patients, usually resolvable only by more study, but at least you’ve been able to show an effect in some population.

Enrichment – High Risk Patients

1. Oncology

Tamoxifen prevented contralateral breast tumors in adjuvant setting (very high risk); it was then studied in people with more general high risk. This was needed a) to have enough endpoints to detect a possible effect and b) because of concern about toxicity. It was labeled for the group studied, with access to Gail Model calculator to assess risk. There was no reason in this case to expect a larger % effect in the people selected, but more events would be prevented.
Enrichment – High Risk Patients

1. Oncology

Potential selection method for frequent endpoints:
D’Amico reported [NEJM 2004; 351:125-135] that in men with localized prostate Ca, following radical prostatectomy, PSA “velocity” (PSA increase > 2 ng/ml during prior year) predicted prostate Ca mortality almost 100% over a 10 year period. There were essentially no deaths from prostate Ca (many from other causes), even though recurrence rates were not so different (NB; not used yet).

Kaplan-Meier Estimates of Disease Recurrence (Panel A) after Radical Prostatectomy, According to the Quartile of PSA Velocity during the Year before Diagnosis

Kaplan-Meier Estimates of the Cumulative Incidence of Death from Prostate Cancer (Panel C) after Radical Prostatectomy, According to the Quartile of PSA Velocity during the Year before Diagnosis
Enrichment – High Risk Patients

1. Oncology (cont)


Four of the 5 methods had high concordance and a striking ability to predict outcome and the differences were very large. One of them, a 70 gene profile, is shown on the next slide. The implications for patient selection are obvious, whether the endpoint is recurrence or survival. Studies should select poorer prognosis patients to have a better chance of showing a drug effect.

Recent approval of MammaPrint, an in vitro test based on gene expression profile will facilitate such selection.

Enrichment-High Risk Patients

2. Cardiovascular

Long routine to choose, in outcome studies, patients at high risk (secondary prevention, post-AMI or stroke, very high cholesterol, very severe CHF, undergoing angioplasty) so there will be events to prevent. For example:

- CONSENSUS (enalapril) in NYHA class III-IV patients studied only 253 patients, showing dramatic survival effect in only 6 months study. Mortality untreated was 40% in just 2 months, and treatment showed a 40% reduction. Later studies needed many 1000's of patients
- First lipid outcome trial (4S - Simvastatin) in a post-ML, very high cholesterol population: 9% 5 year CV mortality, needed only 4444 patients for a mortality effect. Later trials larger, used composite endpoints.
Selection of High Risk Patients

2. Cardiovascular (cont)

Recent JUPITER study by Ridker, et al [Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2009; 359: 2195-207] randomized a relatively low risk, not very high LDL population:

- 17,802 healthy (no hx CVD) people (M>50, F>60)
- LDL < 130 mg/dL
- CRP ≥ 2 mg/L
- No prior lipid Rx, current HRT, uncontrolled HT (190, 100), diabetes,

randomized to rosvastatin 20 mg or placebo.

Endpoint first major CV event (NFMI, NF stroke, hosp’t unstable angina, arterial revasc, or ‘confirmed’ CV death.

### JUPITER

<table>
<thead>
<tr>
<th>Event</th>
<th>Rosuvastatin</th>
<th>Placebo</th>
<th>HR (CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>142</td>
<td>251</td>
<td>0.56 (0.46-0.69)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>NFMI</td>
<td>22</td>
<td>62</td>
<td>0.33 (0.22-0.58)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>NF stroke</td>
<td>30</td>
<td>58</td>
<td>0.52 (0.34-0.82)</td>
<td>0.009</td>
</tr>
<tr>
<td>All death</td>
<td>198</td>
<td>247</td>
<td>0.80 (0.67-0.97)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

In this population, the rate of primary endpoints was pretty low (1.36/100 PY on placebo) and deaths were 1.25 per 100 PYs, so a good-sized study was needed to show even a good-sized effect.

I have little doubt the result was made possible by the enrichment.

Prognostic Enrichment

3. Other

Identifying people at high risk is especially important in "prevention" or risk reduction efforts. Apart from the CV setting, there is major interest in reducing the risk of Alzheimer’s Disease or particular cancers. To have any chance of doing so it will be critical to find a high risk population identified genetically or through early signs of disease (e.g., people with minimal brain dysfunction in studies of prophylaxis of Alzheimer’s Disease; women with BRCA 1 or 2 mutations) in studies to prevent breast or ovarian cancer).
Selection of Likely Responders (Predictive Enrichment)

Identifying the people who will respond to a treatment, then formally studying them, greatly enhances the power of a study and has clear implications for how a drug will be used.

It can be especially critical when responders are only a small fraction of all the people with a condition, e.g., because they have the “right” receptor. In such a case finding an effect in an unselected population may be practically impossible.

Selection can be based on understanding of the disease (pathophysiology, tumor receptors) or it can be empiric (e.g., based on history, early response, response of a biomarker).

Selection of Likely Responders

Pathophysiology

- Hypertension can be high-renin or low-renin. High renin population would show a much larger effect than a mixed population to ACEIs, AIIbs, or BBs.
- We study antibiotics in bacterial infections sensitive to the antibacterial or, if not identifiable initially, we examine the subset that had the relevant organism.
- A well-established genetically determined difference could be the basis for a pathophysiologically selected population. Many tumor genetic or surface markers are related to well-understood effects on enzymes or growth stimulus: Herceptin for Her2+ breast tumors; selection of ER+ breast tumors for anti-estrogen treatment, many other receptor markers.

Even if pathophysiology is unclear, likely responders could be identified by an initial short-term response, an empiric approach. There is a history of this:

- CAST was carried out in people who had a 70% reduction of VPB’s. Only “responders” were randomized.
- Trials of topical nitrates were carried out only in people with a BP or angina response to sublingual nitroglycerin.
- Anti-arrhythmics were developed by Oates, Woosley, and Roden by open screening for response, then randomizing the responders.
- Every randomized withdrawal study has this characteristic.
- History of response to a class.
Selection of Likely Responders

Selection could be based on response of a biomarker; that is, study the entire group and randomize only those with a good response. Possibilities

- Tumor that shows early metabolic effect on PET scan
- Tumor that shows early response on blood measure (PSA)
- Tumor that doesn’t grow over an n-week period (it would be hard to randomize tumor responders to Rx vs. no Rx)
- Only patients with LDL effect > n (or some other less studied lipid)
- Only patients with CRP response > x

Advantages of Predictive Enrichment

1. Efficiency/feasibility
   When responders are a small fraction of the population, predictive enrichment can be critical.

<table>
<thead>
<tr>
<th>Prevalence of Marker-Positive Patients</th>
<th>Prevalence of Marker-Positive Patients</th>
<th>Response in Marker-negative Patients</th>
<th>Response in Marker-negative Patients</th>
<th>Sample Size Ratio</th>
<th>Sample Size Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>0%</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>75%</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>50%</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>25%</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Advantages of Predictive Enrichment (cont)

As the table shows, if 25% of patients have the marker that predicts effect and marker negative patients have no response, an unselected population would need 16 times as many patients [the gain is much less if marker negative patients have same response even if it is smaller].

2. Enhanced B/R if there is toxicity. Trastuzumab (Herceptin) is cardiotoxic. Studies in patients with metastatic cancer as well as adjuvant studies were conducted in patients with Her-2-neu positive tumors, enhancing B/R by removing patients who could not benefit. Her-2-neu negative patients have much less response, and the cardiotoxicity is unacceptable.

Data in the Marker-Negative (Off) Group

A trial done entirely in a marker-positive group is efficient but gives no information about the omitted patients (i.e., do they have some response?). Guidance urges (repeatedly) that, unless there is no real chance of an effect in marker-negative patients, some negative patients should be included because

- They may have some response
- They data can be used to refine the marker cut off

It would still be possible to make the primary endpoint the effect in the enriched stratum.

Selection of Likely Responders

We are at the very beginning of searching for genetic or other characteristics that will predict response. These could be pathophysiologic; that is, based on understanding of disease or drug mechanism (role of her 2 receptor in response to Herceptin; role of EGFR in response to erlotinib), generally with these factors identified prospectively, and with patients either selected by, or stratified by, that factor. But the selection could be simply empirical or descriptive: run a trial in unselected patients with depression, bipolar disease, lipid abnormalities, heart failure and link a genetic baseline finding with response. In fact, one could search widely for such a relationship. The usual course would then be to study the genetically described subset prospectively. Tarceva data illustrate the potential.
Selection of Likely Responders
Tarceva (erlotinib)

Randomized, DB, placebo-controlled trial of Tarceva 150 mg in 731 patients with locally advanced or metastatic NSCLC after failure of ≥1 prior regimen. Randomized 2:1 (488 Tarceva, 243 placebo). Study overall showed clear survival effect.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>HR</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarceva</td>
<td>0.73</td>
<td>0.61-0.86</td>
</tr>
<tr>
<td>Placebo</td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

1 year survival

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Survival (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarceva</td>
<td>6.7</td>
</tr>
<tr>
<td>Placebo</td>
<td>4.7</td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

EGFR+ (127)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Survival (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarceva</td>
<td>10.71</td>
</tr>
<tr>
<td>Placebo</td>
<td>3.84</td>
</tr>
<tr>
<td>p</td>
<td>0.033</td>
</tr>
</tbody>
</table>

EGFR- (111)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Survival (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarceva</td>
<td>5.35</td>
</tr>
<tr>
<td>Placebo</td>
<td>7.49</td>
</tr>
<tr>
<td>p</td>
<td>0.958</td>
</tr>
</tbody>
</table>

Kaplan-Meier Curve for Overall Survival of Patients by Treatment Group

Tarceva (erlotinib)

Tumors were examined for EGFR expression status in 238 (of 731) patients. EGFR+ was defined as ≥10% staining using DAKO EGFR pharmDx kit.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Survival (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR+</td>
<td>10.71</td>
</tr>
<tr>
<td>EGFR-</td>
<td>5.35</td>
</tr>
<tr>
<td>p</td>
<td>0.033</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Survival (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR+</td>
<td>0.65</td>
</tr>
<tr>
<td>EGFR-</td>
<td>1.01</td>
</tr>
<tr>
<td>CI</td>
<td>(0.43-0.97)</td>
</tr>
<tr>
<td>p</td>
<td>(0.65-1.57)</td>
</tr>
<tr>
<td>p</td>
<td>0.958</td>
</tr>
</tbody>
</table>
Recent Examples

1. Oncology
   - Use of imatinib in unresectable and metastatic GIST (gastrointestinal stromal tumor with c-kit, a tyrosine kinase activating mutation present in most GIST, 55/56 patients responded (durable). Also good results in an adjuvant trial.

   - Vemurafenib had dramatic effects in melanoma patients with mutated genes encoding serine-threonine protein kinase BRAF. V is a BRAF inhibitor that blocks the mutated BRAF protein. A controlled trial vs dacarbazine in patients with unresectable melanoma and BRAF$^{v600E}$ mutation showed 74% risk reduction for mortality. Still a ways to go, as resistance often emerges, perhaps leading to combination therapies.

Recent Examples (cont)

   - Dramatic hepatitis C effects in patients with type 1 virus, the most resistant, with boceprevir and telaprevir, with better duration and many curves and shorter treatment than past (interferon plus ribavirin). Newer agents are having broader effects.

   - Ivacaftor is a new drug for CF directed at a specific mutation G551D, present in just 4% of patients, a mutation of the cystic fibrosis transmembrane conductance regulator (CFTR). The drug reverses the defect in sweat and mucus that leads to disease.
Predictive Enrichment – Empiric Approaches

The guidance describes these approaches in considerable detail:

1. Open observation followed by randomization
   - Oates, Woosley, Roden – anti-arrhythmic development
   - CAST: VPB suppression post-MI to prevent sudden death. Patients all screened for response; only randomized people with ≥ 70% VPB suppression.
     Drug “worked” but was lethal
   - Beta-blocker CHF studies – screened for tolerability. Then withdrawn and randomized. Not a prediction of favorable outcome but of ability to tolerate.

2. History of response to treatment class (indapamide).

3. Results in earlier studies: BiDil showed a large response in blacks in early study. Definitive study solely in blacks showed a 40% mortality reduction.

4. Adaptation: after interim look, include more of the responder population (e.g., men, disease severity), count everybody.

5. Randomized withdrawal study.

Randomized Withdrawal

Amery in 1975 proposed a “more ethical” design for angina trials, which then often ran 8 weeks to 6 months in patients with frequent attacks (before regular CABG and angioplasty).

Patients initially receive open treatment with the test drug, then apparent responders are randomized to test drug (at one or more doses) or placebo. Endpoint can be time to failure (early escape) or conventional measure (attacks per week).

These trials are all enriched with people doing well on treatment. Also, no new recruitment is needed. This is now a routine way to demonstrate long-term benefit of anti-depressants.

Early use in studying nifedipine in vasospastic angina (first approved use).

---

<table>
<thead>
<tr>
<th></th>
<th>Nifedipine</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Early withdrawal</td>
<td>0</td>
<td>5*</td>
</tr>
<tr>
<td>Early withdrawal or AMI</td>
<td>0</td>
<td>6*</td>
</tr>
</tbody>
</table>

* Statistically significant at p ≤ 0.05
Randomized Withdrawal (cont.)

Design has major advantages

- Efficient: “enriched” with responders giving a larger drug-placebo difference
- Efficient: patients already exist and known, e.g., a part of an open or access protocol
- Ethical: can stop as soon as failure criterion met, very attractive in pediatrics

We are seeing extensive use in showing persistent effects of pain medications and has been used to study needed duration of use of bisphosphonates and adjuvant breast cancer therapy.

Other Designs

1. Use More Sensitive Setting

It is well-recognized that field studies of anti-histamines require large patient samples (> 200 per group often) and fail regularly nonetheless. Chamber studies (antigen introduced, don’t depend on pollen, winds) are a kind of PD study (but with a standard clinical endpoint), require much smaller numbers.

Will also need field studies, but initial dose finding surely should be in chamber studies (also better for time of onset, duration of effect). Then try to confirm in field studies.

2. More Sensitive D/R

Consider conducting dose response studies for effectiveness in known responders to the drug or drug class to increase sensitivity, or identify responders pharmacologically, if possible, all enrichment designs. The only effect of including non-responders is to obscure (flatten) the dose response relationship. Note, though, that non-responders may have adverse effects that cannot be ignored.

It would usually be important to test non-responders separately to see if they merely have a shift in D/R, an important discovery, if true, and, if responders are not identifiable, studies in a non-selected population would be needed to assess overall B/R.
Other Designs

2. More Sensitive D/R

I don’t have other examples (because this just is not done) but it is obvious that a better picture of the effect of different doses can be found if drug effect is 20 mmHg than if it is 5 mmHg.

Does leaving out the non-responders lead to a loss of information? I don’t think so; the non-responders can’t respond to any dose, so dose-finding in them is of no value (again, good to test larger doses in them).

Studies in Non-Responders

Studies in non-responders are a very attractive design. Showing an effect in such patients would give a new drug an edge and has allowed approval of drugs otherwise too toxic:

- Captopril (thought to cause agranulocytosis) was superior to diuretic, nesipram, 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. Non-responders on a previous exposure could respond to the same drug in a new study.
Studies in Non-Responders

standard drug

standard drug
non-responder

new drug

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

<table>
<thead>
<tr>
<th>Response (%)</th>
<th>Clozapine</th>
<th>CPZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGI (decrease ≥ 1)</td>
<td>71</td>
<td>37*</td>
</tr>
<tr>
<td>BPRS items (dec ≥ 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>concept disorganization</td>
<td>60</td>
<td>39*</td>
</tr>
<tr>
<td>suspiciousness</td>
<td>64</td>
<td>42*</td>
</tr>
<tr>
<td>hallucinations</td>
<td>59</td>
<td>51</td>
</tr>
<tr>
<td>thought content</td>
<td>65</td>
<td>40*</td>
</tr>
<tr>
<td>CGI and BPRS</td>
<td>15</td>
<td>2*</td>
</tr>
</tbody>
</table>

*p < 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 meeting about COX2 selective NSAIDs 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.

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

Other Designs

4. Studies in Intolerants

A study in people who cannot tolerate standard therapy is also enriched for the population of interest and is especially useful when adverse effects to the standard are uncommon.

For example, if you wanted to show AIIB causes less cough than ACEI’s (rate, say 5%) you need a study large enough to distinguish 5% from 1-2%, i.e., a big study. This has in fact been done, but there is an easier way.
Study in Intolerants

Lisinopril (ACEI) 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

<table>
<thead>
<tr>
<th></th>
<th>Lisinopril</th>
<th>Losartan</th>
<th>Metolazone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough rate, any</td>
<td>97%</td>
<td>18%</td>
<td>21%</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Enrichment Caveats

Evidence of effectiveness for marketing vs. proof of principle

A critical distinction when using prospectively enriched populations is between proof of principle (evidence that the drug is effective in some defined population) and evidence of efficacy suitable for approval. There is always an issue of generalizability when you enrich, probably manageable if the enrichment feature is easily used in practice, but there are special problems when you cannot enrich (or select patients) at the time of treatment because

- You can find the responders only by treating everybody. E.g., to do a study in people with a history of response or who respond in a screening period, you have to treat everybody, then watch. This is the case for randomized withdrawal study that provides primary evidence of effectiveness.

Enrichment Caveats (cont)

- You need to act too soon (e.g., infections, can’t wait for growth of organism or measuring its sensitivity); treatment of possible AMI with a thrombolytic before enzymes are available; treatment of stroke before it is known whether hemorrhagic or thrombotic. In these cases, all people receive the treatment but only some can benefit
- Relevant test is not available to practitioners

Cases: real and theoretical
- Antibiotics
- Sepsis drugs
- Stroke treatment

In these cases, the study’s role is to prove effectiveness but it may not provide guidance about how to use the drug.

Enrichment Caveats

Evidence of effectiveness vs. proof of principle

1. Antibiotics - Routine to randomize all patients with a condition and only analyze people with sensitive organism. Seems reasonable:
   - Logical - how could the drug work if infectious organism is not sensitive?
   - The organism is a baseline feature and should be randomly assigned (and removal of patients with insensitive organism is also therefore random)
   - A valid test of effectiveness, even if patients with relevant organisms are only a small part of the population

But suppose the drug has a significant adverse effect in some people. The benefit could outweigh risk in people with a sensitive organism but not others, who might be adversely affected.

So, analyzing only the people with sensitive organisms provides proof of principle (the drug clearly works on what it should treat) but you must look at all randomized for benefit/risk assessment.
Enrichment Caveats

Evidence of effectiveness vs. proof of principle

2. Sepsis - In some cases it seemed that there was an effect in people with right pattern of disease (e.g., gm negative sepsis), but no effect overall, perhaps because of an adverse effect on the other patients (tendency not to imagine an adverse effect on the non-responders, but in fact there can be one). CBER asked for evidence of effect in the overall population. An alternative might be reassurance of no adverse effect overall.

Enrichment Caveats

Evidence of effectiveness vs. proof of principle

3. Stroke - Suppose a treatment needed to be used early, before clear Dx of stroke as thrombotic or hemorrhagic, and there was clear benefit shown after the study in former and clear harm in latter, just as might have been theoretically predicted for, say, a thrombolytic. The analysis of people with thrombotic stroke could give "proof of principle" (evidence of an effect in the only plausible population) but B/R would need to be based on all randomized, selected for thrombotic stroke as best you can (history, presenting characteristics, age, gender, etc.) [NB: TPA is approved for people shown to have had a thrombotic stroke; current labeling calls for demonstration of a thrombotic stroke before use.]

Enrichment Caveats

Evidence of effectiveness vs. proof of principle

If a genetic characteristic is used to define a treatment population, clearly predicts response, and is available for use to select patients, you have both proof of principle and relevant efficacy and B/R assessments. Presumably, in that case, we would usually ask that the diagnostic test be available. But perhaps we would not if Rx benign and effect easily assessed, e.g., on blood pressure or cholesterol. It could be easier and cheaper to do a therapeutic trial than to do the test. Probably would then need only modest data for people without the characteristic, perhaps none, if the drug obviously could not work in them.
Enrichment Caveats
Evidence of effectiveness vs. proof of principle

If the treatment cannot be directed at the subgroup with the genetic characteristic (timing, availability), the effect in the overall population might seem to be the relevant effectiveness measure and basis for benefit/risk assessment. But suppose the overall population does not show an effect, e.g., because the “responder subgroup” is small. Could the effect in the genetically appropriate population subset be the basis for efficacy determination (i.e., the primary endpoint) even if it can’t be used for selection once the drug is marketed?

It seems clear that it could, similar to the antibiotic case, but

- B/R would be based on the overall population, recognizing that only the subset benefits
- You’d want to be very sure the drug is not harmful to the non-responders, but how to do that is not so clear, beyond being sure there is no overall adverse effect.
- You’d expect an overall favorable “lean,” even if NS because the responder subset is too small.

Design Considerations and Cautions

A long section in guidance on what to watch out for in considering predictive enrichment designs and the properties (advantageous or not) of specific designs. Obviously, only highlights here.

1. Performance characteristics of the selection criteria

When a test (genomic, proteomic) is used to choose patients you need to know test precision and test performance (generally sensitivity/specificity/predictive value) and how any cutoffs used relate to S & S. E.g., for Herceptin, cut off at 2+ on Her-2-neu could find more responders than 3+ (increased sensitivity) but also more non-responders (poorer specificity). Ideally, would include a fairly broad range of marker values and assess performance, and define the best cutoff value. But clearly need a larger study to do that. May be able to modify by interim looks (e.g., no responses in her-2-neu 1+, so drop them).

Design (cont)

2. When to develop the classifier

Ideally, early studies enter a broad range and evolving data help choose cutoff. But a phase 3 study with broad inclusion criteria could explore the impact of various thresholds and plan analyses (correcting for multiplicity) using various thresholds.

3. Who to include
   a. Only enrichment population patients
   b. All, but analyze only those with the marker as primary endpoint.

Where there is an enrichment marker, a number of study designs can be considered.
Prospective, Screened - no possible effect in (-) group

- Supports effect for enriched population
- Plainly overstates effect for unselected population
- No information on people below the marker cutoff
- Suitable when there is little chance marker negatives will respond
- Labeling MUST identify only marker positive as suitable, usually need CDRH approval of test.

Prospective, Stratified - where there is possible effect in the (-) group and/or where toxicity in the (-) group needs to be evaluated because pre-treatment selection is not possible

- We would generally urge this (top), but probably not insist. Marker + subset is usually the primary endpoint. Study size based on marker-positives; the marker-negative group could be smaller.
- Get some data on marker negative (could randomize unequally).
- Bottom design is where you don’t have the marker when treatment starts