

Enrichment Strategies

Robert Temple, MD

Deputy Center Director for Clinical Science

Center for Drug Evaluation and Research

Food and Drug Administration

FDA/DIA Statistics Forum 2012
April 25, 2012

Outline

As all know, PDUFA 4 promised a guidance on Enrichment Strategies for Clinical Trials to Support Approval. We are well along.

1. Definition

2. Three principal enrichment strategies

- Practical enrichment – Decreasing heterogeneity and “noise”
- Identifying high-risk patients – prognostic enrichment
- Choosing patients likely to respond to treatment – predictive enrichment

3. Design considerations and cautions

4. Regulatory issues

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).

Enrichment - Definition

Enrichment is prospective use of any patient characteristic – demographic, pathophysiologic, historical, genetic, and others – to select patients for study to obtain a study population in which detection of a drug effect is more likely than it would be in an unselected population.

Enrichment could also refer to a subset in a study that is to be used in the primary analysis, even if a broader population is studied.

The increased study power facilitates “proof of principle” (there is a clinical effect in some population) but it leaves open 1) the question of generalizability of the result and how the drug will work in other populations and 2) how much data are needed before or after approval in the “non-selected” group. (Do these patients benefit at all? Are they harmed?)

As will be noted, the main reason for enrichment is study efficiency – increasing the chance of success, often with a smaller sample size – but it also provides major benefits of individualization, directing treatment where it will do the most good and sparing people who cannot respond potential harm.

Enrichment

Enrichment is usually focused on effectiveness, but it is pertinent to safety.

- 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 in ruling out risk.
- 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 difference if there is one.

Kinds of Enrichment

There are 3 broad categories of enrichment

1. Practical – virtually universal – decrease heterogeneity
 - Define entry criteria carefully
 - Find (prospectively) likely compliers
 - Choose people who will not drop out (VA BP Studies)
 - 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 variance, enrichment strategies fall into two distinct types:

2. Choosing high risk patients, i.e., those likely to have the event (study endpoint) of interest, or likely to have a large change in the endpoint being measured during the study. This is “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 toxicity.

3. Choosing people more likely to respond to treatment. This is “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).

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 likelihood of cardiovascular outcomes
 - Severity of heart failure
 - Cholesterol, blood pressure levels; angiographic appearance
 - Diabetes
 - Recent events (AMI, stroke)
 - Elevated CRP (JUPITER Study of rosuvastatin)
 - Family history
 - Gender, race, age
- Risk factors in cancer
 - Previous breast cancer to predict contralateral tumor
 - Tumor histology

Prognostic Enrichment

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 larger effect of tamoxifen (% reduction) in the people selected, but more events would be prevented.

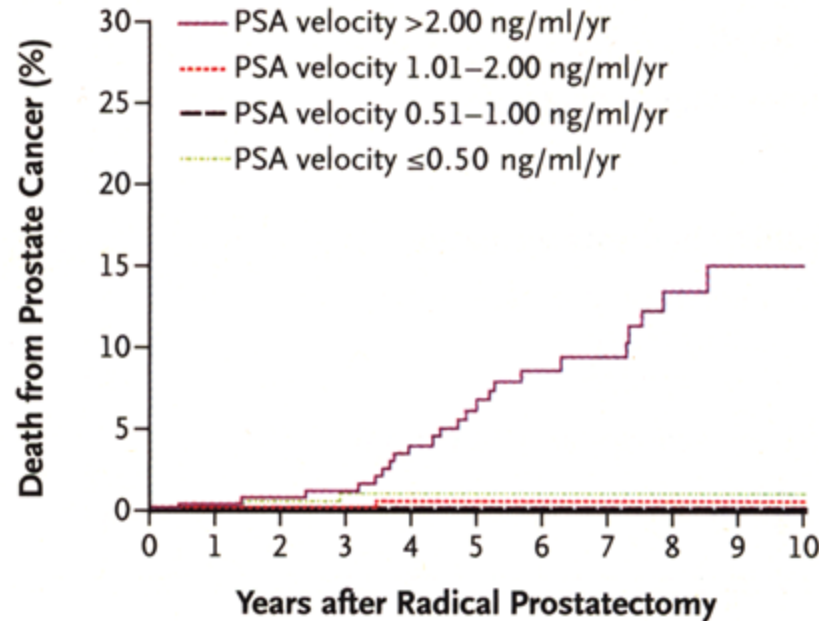
Prognostic Enrichment

1. Oncology (cont.)

Potential (not used or maybe not fully accepted, but a good illustration) selection method for more 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. Given concerns about effects of treatment on survival, an adjuvant prostate Ca study would surely want to include patients at risk of death.

C



No. at Risk

PSA velocity >2.00 ng/ml/yr	262	257	248	226	187	157	123	92	60	36	22
PSA velocity 1.01-2.00 ng/ml/yr	288	275	248	229	194	158	131	91	58	36	20
PSA velocity 0.51-1.00 ng/ml/yr	289	281	260	227	176	131	94	55	36	18	11
PSA velocity ≤0.50 ng/ml/yr	256	236	200	163	139	108	81	61	34	20	9

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

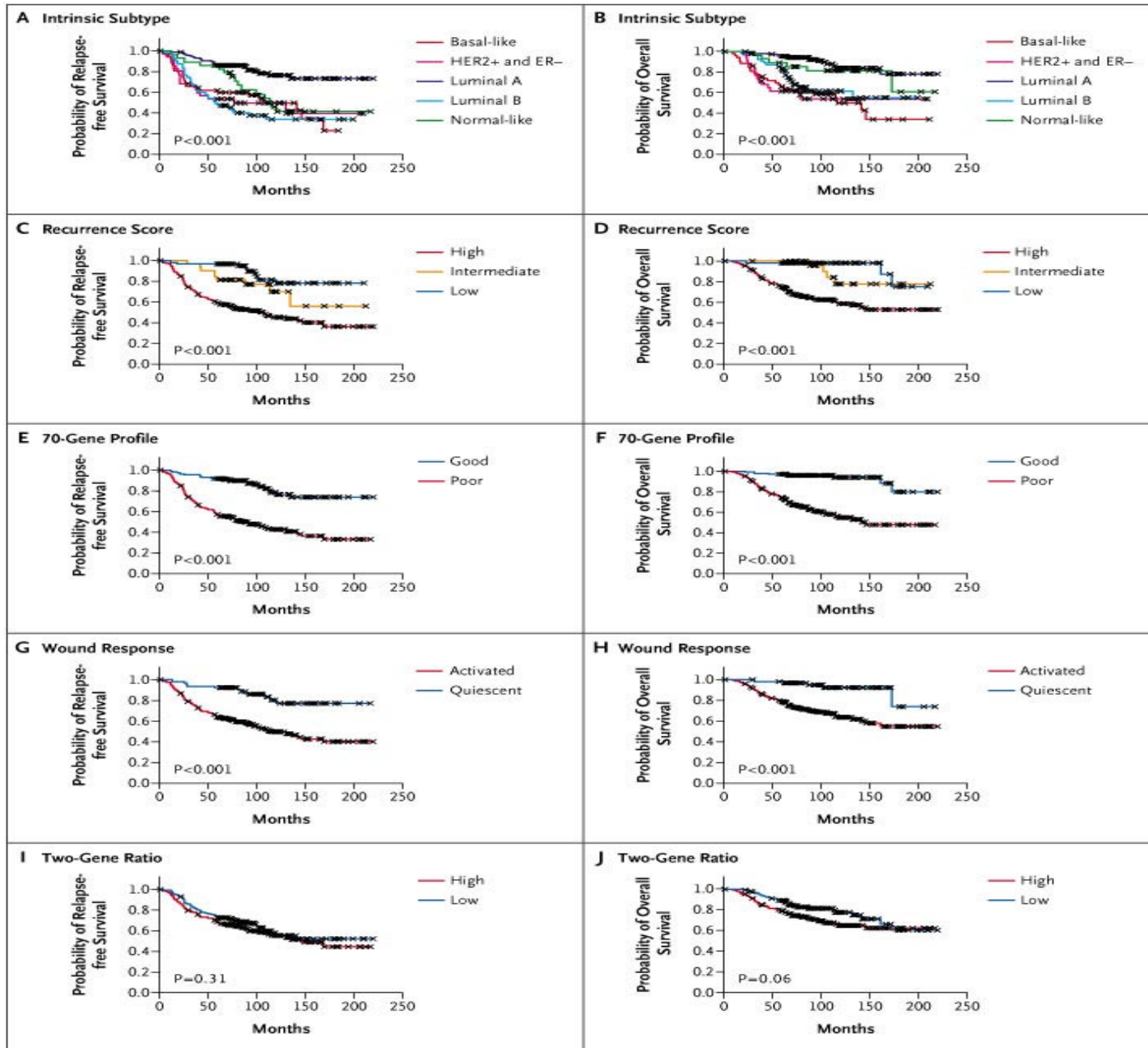
Prognostic Enrichment

1. Oncology (cont)

Fan, et al [NEJM 2006; 355: 560-69] recently applied 5 different gene-expression profiling approaches, intended to predict breast cancer recurrence rates, to a 285 patient sample treated with local therapy, tamoxifen, tamoxifen plus chemo, or chemo alone.

The results and methods used are shown on the next slide. Four of the 5 methods had high concordance and a striking ability to predict outcome and the differences were very large. 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.



Prognostic Enrichment

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) study was in NYHA class IV patients. It needed only 253 patients to show a dramatic survival effect in a 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) was in a post-MI, very high cholesterol population: 9% 5 year CV mortality, needed only 4444 patients for a mortality effect. Later trials larger, used composite endpoints (i.e., not survival).

Prognostic Enrichment

2. Cardiovascular (cont)

JUPITER study by Ridker, et al [Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. NEJM 2009; 359: 2195-207] compared rosuvastatin and placebo in people not usually considered candidates for statin treatment:

17,802 healthy (no hx CVD) people (M>50, F>60)

LDL < 130 mg/dL (below level where Rx is recommended)

Endpoint first major CV event (NFMI, NF stroke, hosp'n unstable angina, arterial revasc, or "confirmed" CV death). To obtain enough events on placebo, elevated CRP was used to select higher risk patients (about 0.7% over course of study).

	Rosuv	Plbo	HR(CI)	P-value
Primary	142	251	0.56 (0.46-0.69)	< 0.00001
NFMI	22	62	0.33 (0.22-0.58)	< 0.00001
NF Stroke	30	58	0.52 (0.33-0.80)	0.003
All death	198	247	0.80 (0.67-0.97)	0.02

Prognostic Enrichment

3. Other

Identifying people at high risk is especially important in “prevention” or risk reduction efforts. Apart from the CV risks we know about, there may be genetic predictors of risk (e.g., for Alzheimer’s Disease or particular cancers) or early signs of Alzheimer’s Disease (people with minimal brain dysfunction or other abnormalities). A population without such a predictor might have few or no cases over several years, making a demonstration of an effect impossible.

Predictive Enrichment

Probably the most exciting enrichment today is predictive enrichment, finding the patients with the greatest likelihood of responding to treatment. Studying people who will respond to a treatment greatly enhances the power of a study, facilitating approval, but it may also have critical implications for how a drug will be used.

It can be especially important 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 a survival 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).

Predictive Enrichment

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, rather, we analyze the patients who turn out, after randomization, to have a sensitive 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 tumor growth rates; Herceptin for Her2+ breast tumors; selection of ER⁺ breast tumors for anti-estrogen treatment, and use of many other receptor markers illustrate this.

Predictive Enrichment

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

- CAST was carried out in people who had to have a 70% reduction of VPB's during a screening period. 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.

Predictive Enrichment

Selection could be based on response of a biomarker; that is, screen 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 2: Sample Size Ratios as a Function of the Prevalence of Marker-Positive Patients

Prevalence of Marker-Positive Patients	Response in Marker-negative Patients (% of marker positive response)	
	0%	50%
	Sample Size Ratio	Sample Size Ratio
100%	1.0	1.0
75%	1.8	1.3
50%	4	1.8
25%	16	2.6

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]. Recently, FDA approved ivacaftor for CF patients with a specific gene mutation that is present in just 4% of CF patients. A study in an unselected population would have had no chance of success.

2. Enhanced B/R if there is toxicity (Herceptin).

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. Her-2-neu negative patients have much less response, and the cardiotoxicity is unacceptable.

Data in the Marker-Negative (Off) Group

Two important questions arise when using such selection criteria. One is the quality of the test. The second is the sensitivity and specificity of the various predictive cut-off points (how positive must Her-2-neu be?) In general, unless there is no real chance of an effect in marker-negative patients, some negative patients should be included in studies (stratified) because

- They may have some response
- They may help refine the marker cut off

Early studies may solve this problem, but the larger numbers in later trials may give better answers. It would still be possible to make the primary endpoint the effect in the enriched stratum (routine in antibiotic trials where sensitivity of the organism is not known at randomization), while examining response in patients below the cut-off.

Predictive Enrichment – Empiric Approaches

There are many such possibilities. A few, with some illustrations, are:

1. Open trial 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 $\geq 70\%$ 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

3. Results in earlier studies (BiDil showed far greater response in blacks, allowing a definitive trial entirely in blacks)

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

Predictive Enrichment – Pathophysiology or genetic characteristics

1. Only people who make the active metabolite (clopidogrel)
2. Only people whose tumor takes up the drug (History, test for I 131 uptake in thyroid tumor to choose dose)
3. Effect on tumor metabolism, e.g., glucose uptake
4. Proteomic markers or genetic markers that predict response

Plainly, the wave of the future in oncology (Herceptin; imatinib inhibits c-KIT, a receptor for tyrosine kinase, that is mutated and activated in most GIST patients; vemurafenib in melanoma effective in patients with activating mutation BRAF^{V600-E}).

Usually the marker is pre-selected but Friedlin and Simon suggest a way to look for responsive subsets half-way and analyze both whole population and subset.

Randomized Withdrawal

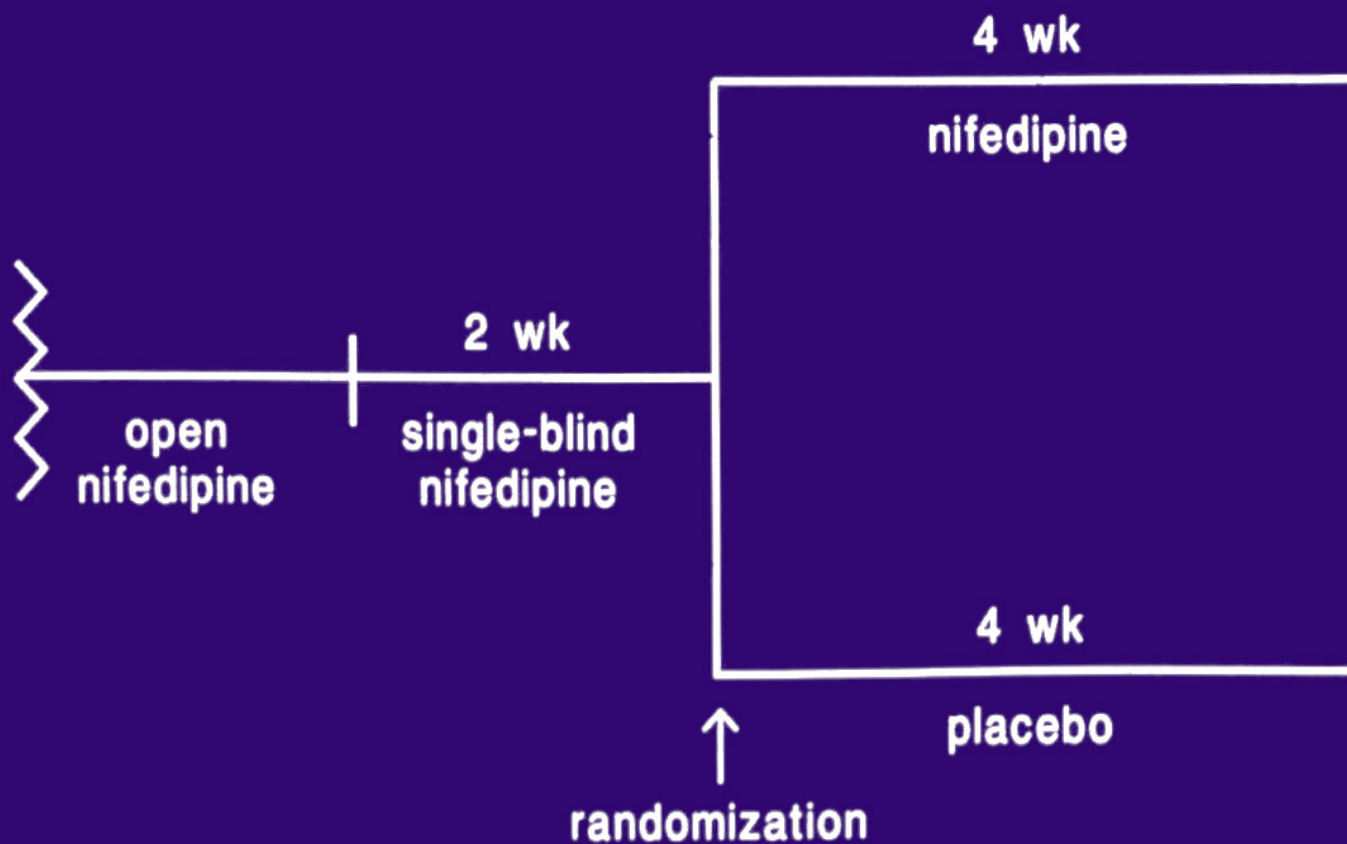
Amery in 1975 proposed a “more ethical” design for angina trials, which then often ran 8 weeks to 6 months in patients with frequent attacks (before regular CABG and angioplasty).

Patients initially receive open treatment with the test drug, then 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.

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

Nifedipine Randomized Withdrawal



	Nifedipine	Placebo
n	13	15
Early withdrawal	0	5*
Early withdrawal plus AMI	0	6*
Investigator's judgment of success	11	2*
Median angina/week	0	3.4*
Mean angina/week	0.7	18.4*
Change from baseline in attacks/week		
better (≤ 1)	0	0
same (± 1)	11	6
worse (≥ 1)	2	9

*p<0.05, one sided

Randomized Withdrawal (cont.)

Design has major advantages

- Efficient: “enriched” with responders, so will show 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

Other Predictive Enrichment

Studies in non-responders; randomize to new drug and failed drug. A comparison enriched with people who will not respond to the control drug, increasing drug-control difference.

Studies in intolerants; randomize to new drug and poorly tolerated drug, a comparison enriched with people who will do “badly” on the control drug.

Both should give a larger drug-control difference.

Very valuable findings – rarely attempted.

Design Considerations and Cautions

There are some things to watch out for considering predictive enrichment, notably the performance characteristics of the selection marker and how broad a range of patients (marker responses) should be selected.

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 cutoff 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 cut-off 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)

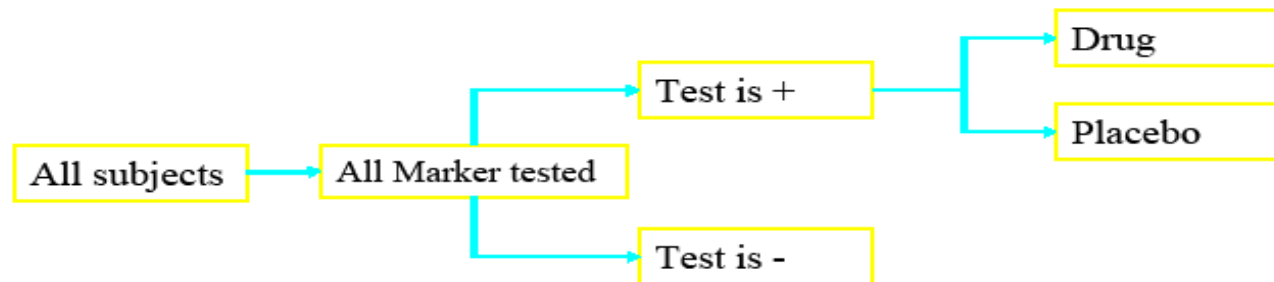
1. As noted, a classification method should be precise and its sensitivity and specificity should ideally be known.

2. When to develop the classifier is a critical question
Early studies can enter a broad range and evolving data can then help choose the cutoff. But early data may be limited in numbers and a phase 3 study with broad inclusion criteria could explore the impact of various thresholds. Even if a particular threshold is used to define the analytic population, analyses using various other thresholds can be examined descriptively.

3. For predictive enrichment using a marker, an important decision is who to include
 - a. Only enrichment population patients

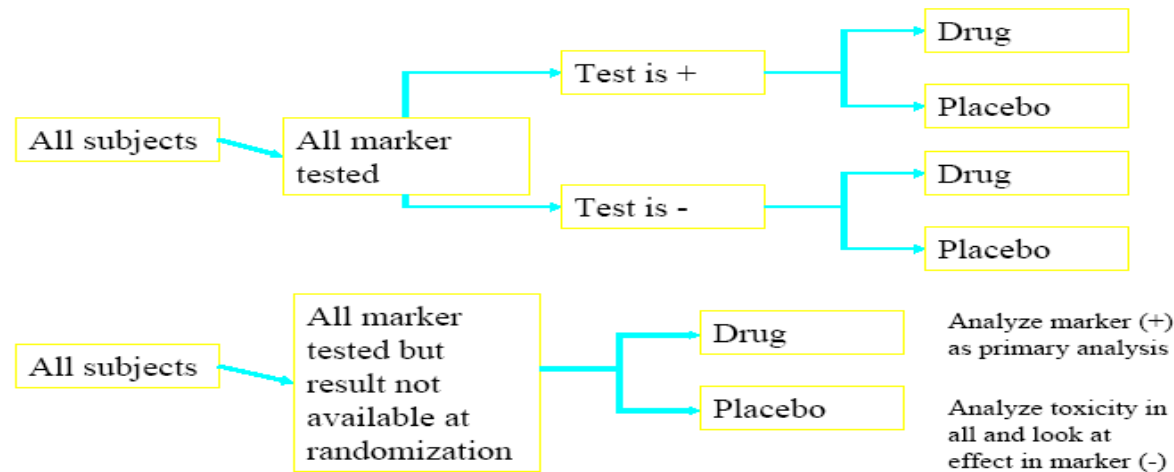
 - b. Both non-enrichment and enrichment patients, but with analysis of only those with the marker as primary endpoint.

Prospective, Screened - no possible effect in (-) group



- Supports effect in the enriched population
- Plainly overstates effect for an unselected population
- Gives information on people below the marker cutoff
- Suitable when there is clear evidence that marker negatives will not respond
- Labeling MUST identify only marker positive as suitable; usually will 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



- Do not need all patients; could include more marker positive; but you get some data on marker negative and sensitivity/specificity for various levels of marker (a receiver-operative curve)
- We would generally urge this design but probably not insist. Interest is partly to understand consequences as of off label use. Marker + group would usually be the primary endpoint and study size would be based on marker-positives patients.

Other Issues

1. Generally, we like individualization of treatment. But suppose you can't use the marker to choose (obtain a baseline value, but only after therapy begins). Could still analyze the marker-positive subset to show effectiveness (demonstrate effect) which might not be possible for whole population, if this were mostly non-responders).
2. Would prefer not choosing too narrow a population and miss some responders, so we encourage broad population. But would surely still approve a drug if it works in the narrow population studied, perhaps with post-marketing studies of a broader population.