Enrichment Strategies for Clinical Trials

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

CDER Enrichment Webinar
March 25, 2013

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In December 2012, FDA published a draft guidance: Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products.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

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 it 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.

Enrichment

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). Not 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.

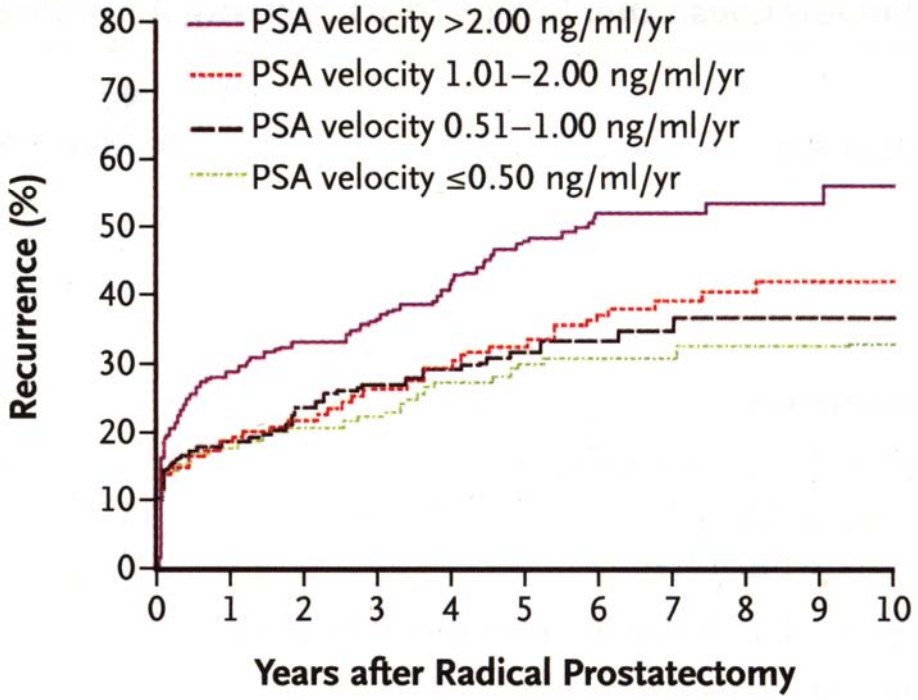
Enrichment – High Risk Patients

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).

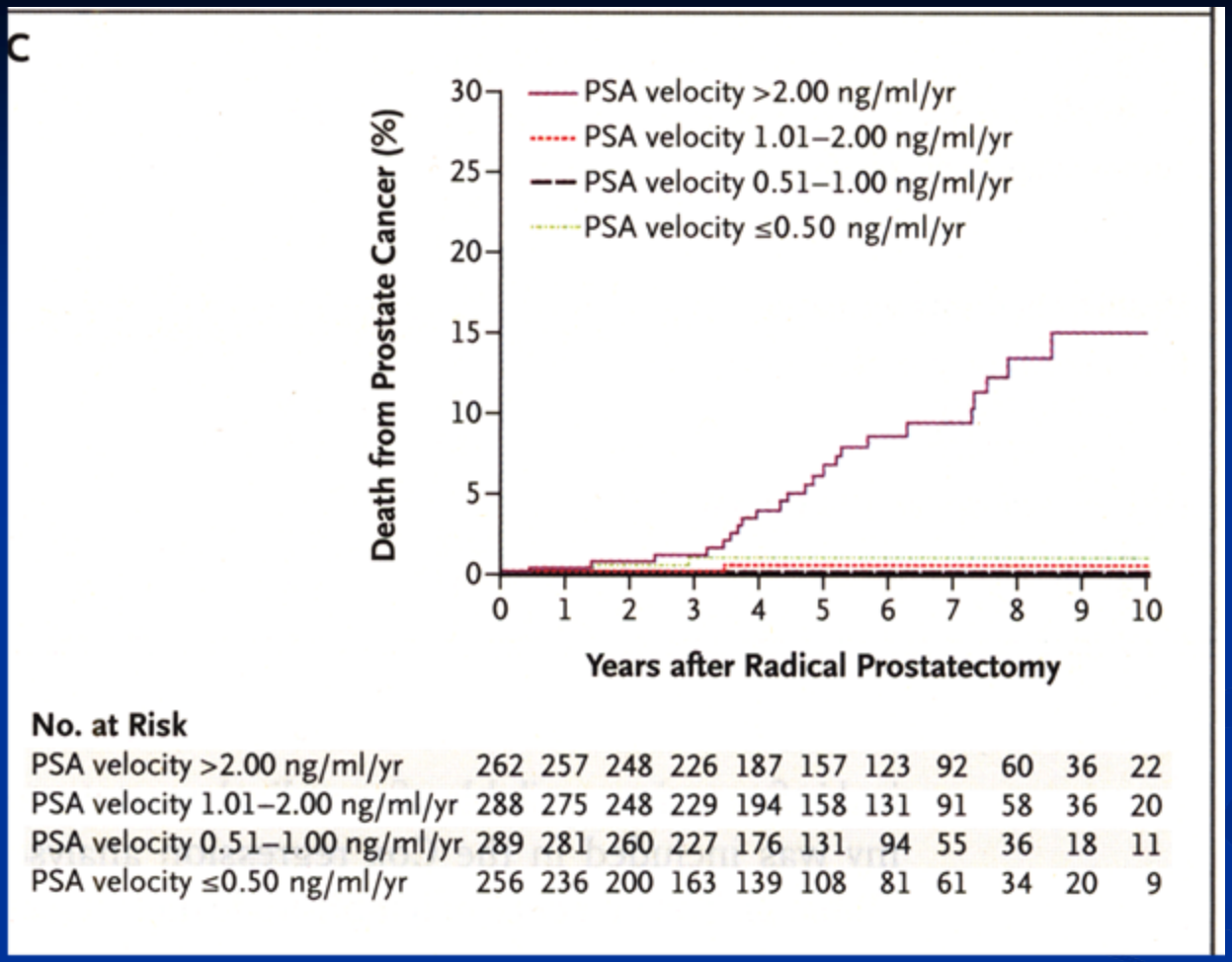
A



No. at Risk

PSA velocity >2.00 ng/ml/yr	247	173	155	132	104	81	60	45	31	19	13
PSA velocity 1.01–2.00 ng/ml/yr	280	218	191	167	133	101	84	56	36	19	15
PSA velocity 0.51–1.00 ng/ml/yr	287	226	193	158	120	92	64	36	23	14	9
PSA velocity ≤0.50 ng/ml/yr	249	190	156	128	103	84	58	43	24	13	5

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



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

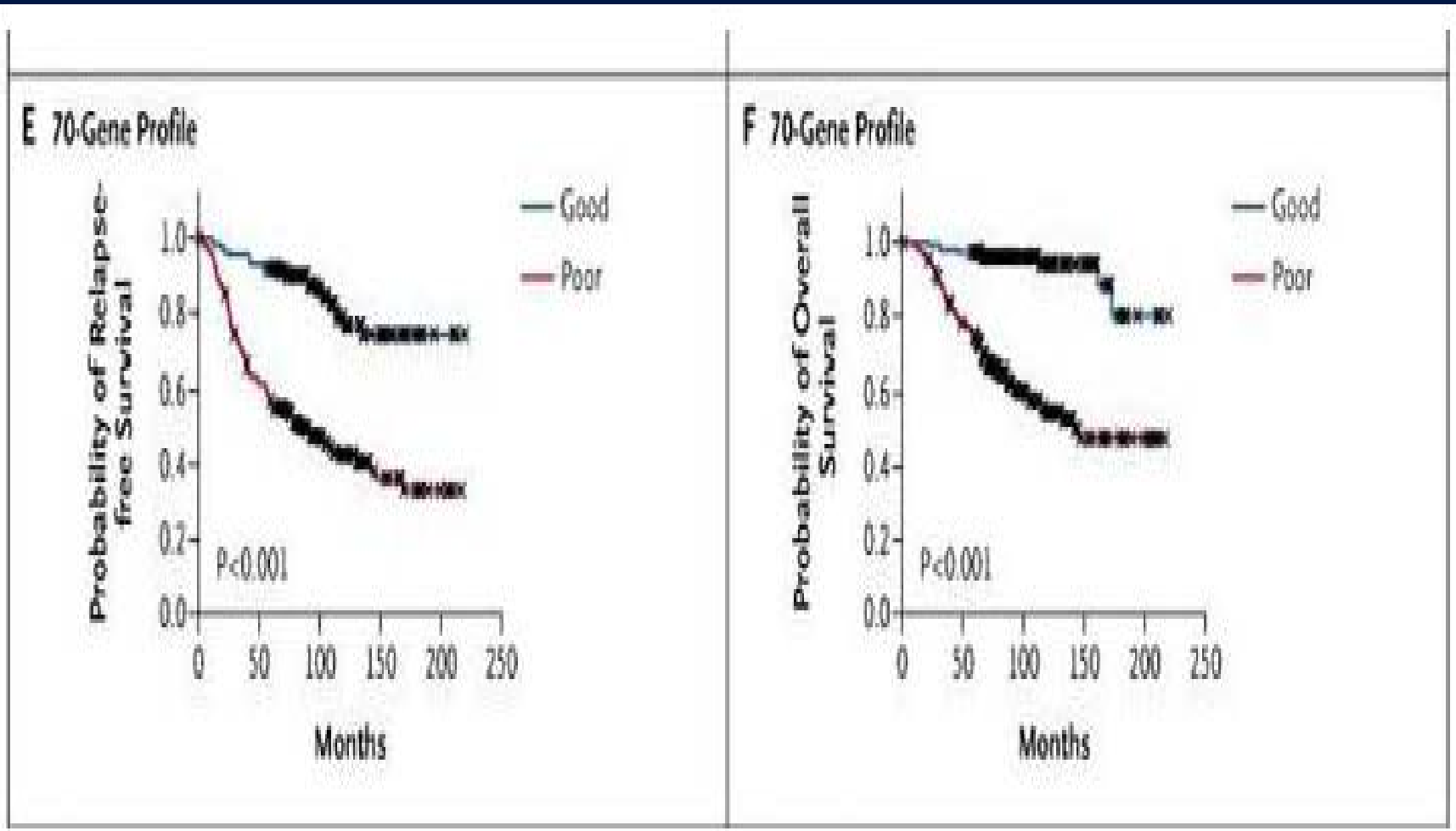
Enrichment – High Risk Patients

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.

   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.


E 70-Gene Profile
Probability of Relapse-free Survival
1.0
0.8
0.6
0.4
0.2
0.0
P<0.001
0
50
100
150
200
250
Months
Good
Poor
F 70-Gene Profile
Probability of Overall Survival
1.0
0.8
0.6
0.4
0.2
0.0
P<0.001
0
50
100
150
200
250
Months
Good
Poor

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-MI, 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. NEJM 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, uncont'd HT (190, 100), diabetes,

to rosuvastatin 20 mg or placebo.

Endpoint first major CV event (NFMI, NF stroke, hosp'n unstable angina, arterial revasc, or "confirmed" CV death.

JUPITER

	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

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.

Selection of High Risk Patients

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). This is especially critical if intervening early is important.

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.

Selection of Likely Responders

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.

Parallel Dose-Response Studies of Indapamide

	Study Dose (mg)	n	Baseline BP	Decrease from Baseline (S/D)	
				Standing	Supine
1. Micheal, et al	placebo	17	146/102	3/3	1/1
	1.0	14	143/103	7/5	6/6
	1.5	13	141/101	5/4	5/3
	2.0	15	150/102	21/9*	18/7
	2.5	14	151/104	20/9	17/7
2. Mroczek, et al	placebo	19	153/103	1/2	1/2
	1.0	21	155/104	12/5	10/5
	2.5	21	148/102	14/7	15/6
	5.0	20	153/102	14/6	13/6
3. Sanchez-Torres	placebo	8	163/103	+6/3	+0/6
	1.0	9	174/106	10/4	10/8
	2.5	9	164/104	29/12	22/6
	5.0	8	171/105	37/15	28/15
4. Multicenter vs HCTZ	2.0	30	141/101	12/8	11/7
	2.5	25	147/103	12/7	11/7
	HCTZ 100	28	150/101	12/8	11/6
5. Multicenter Long-term (40wk)	2.5	62	148/100	13/8	12/9
	5.0	71	145/101	14/10	13/9
	HCTZ 50	54	145/101	12/10	11/9

*Underlined values significantly different from placebo

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 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].

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.

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.

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 – recent cystic fibrosis drug
5. Virus genotype – hepatitis c drugs boceprivir and telaprivir treat genotype 1

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 $BRAF^{V600-E}$.

Genomic/proteomic selection

Increasingly, we are seeing predictive enrichment using genetic or proteomic characteristics that predict a response. These have been mainly in the oncology setting, but have more recently identified subsets of patients with cystic fibrosis who respond to ivacaftor (GSSI D mutations of CFTR gene), a small fraction (4%) of all CF patients. A study in an unselected population would have surely failed. As noted previously, we have approved two genomically directed drugs for hepatitis C.

So genomic prediction is spreading.

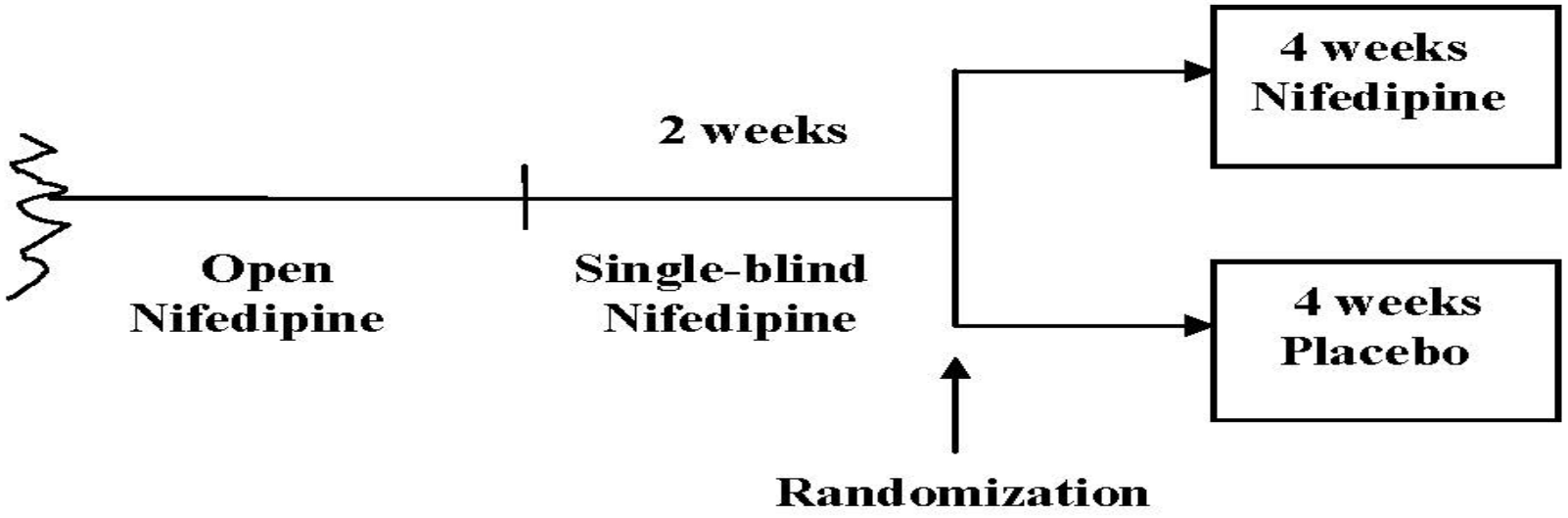
Randomized Withdrawal

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).



	Nifedipine	Placebo
N	13	15
Early withdrawal	0	5*
Early withdrawal or AMI	0	6*

* Statistically significant at $p \leq 0.05$

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 Predictive Enrichment

Studies in non-responders; randomize to new drug and failed drug. This is a particularly relevant comparative effectiveness study in a class of patients with a real need.

Studies in intolerants; randomized to new drug and poorly tolerated drug. Used to show losartan does not cause cough.

Both are enriched designs not by better response to drug but by poorer response (failure or intolerance) to the previous drug, giving larger drug-control difference.

Very valuable findings – rarely attempted, although basis for approval of several drugs with major toxicity: clozapine (agranulocytosis), captopril (agranulocytosis), and bepridel (TdP).

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 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)

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.

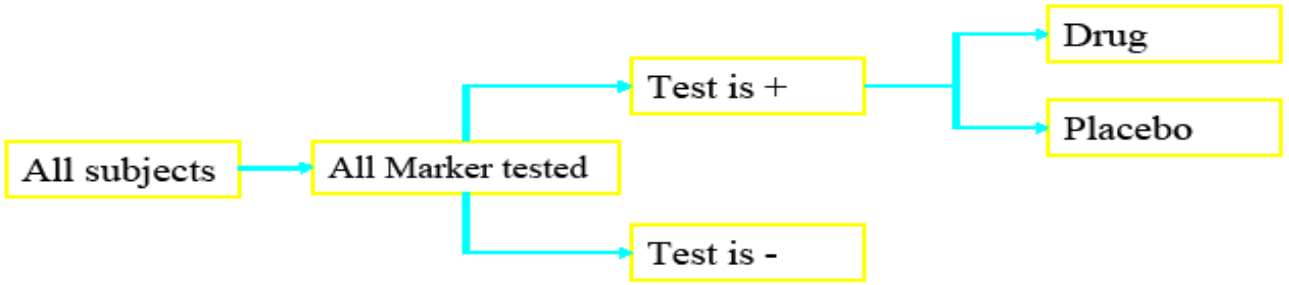
3. Who to include

   a. Only enrichment population patients

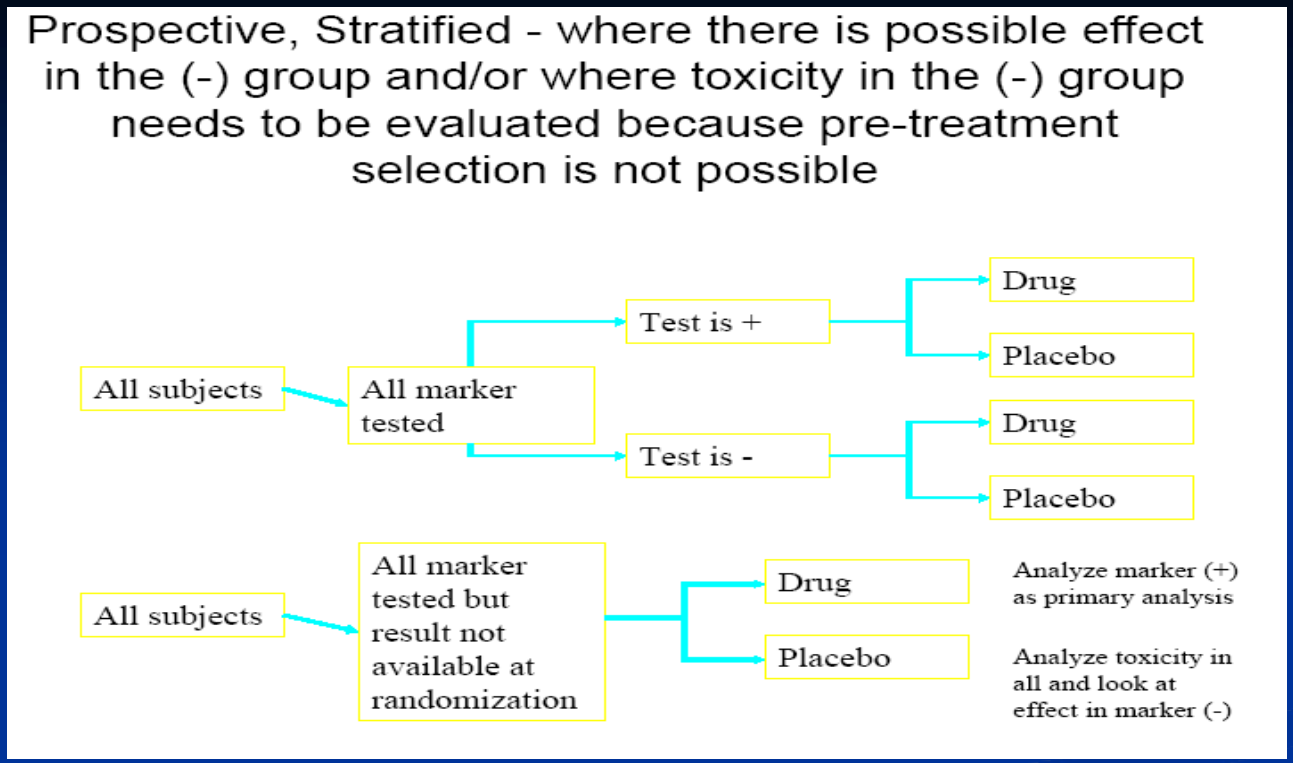
   b. All, but analyze only those with the marker as primary endpoint.

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.



• 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