Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

March 2019
Clinical/Medical

Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products

Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353; Email: druginfo@fda.hhs.gov
https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

and/or

Office of Communication, Outreach, and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, rm. 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010; Email: ocod@fda.hhs.gov
https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

March 2019
Clinical/Medical

TABLE OF CONTENTS

Contains Nonbinding Recommendations

Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products Guidance for Industry[1]

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA office responsible for this guidance as listed on the title page.

I. INTRODUCTION

The purpose of this guidance is to assist industry in developing enrichment strategies that can be used in clinical investigations intended to demonstrate the effectiveness of drug and biological products. Enrichment is the prospective use of any patient characteristic 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. Although this guidance focuses on enrichment directed at improving the ability of a study to detect a drug's effectiveness, similar strategies can be used in safety assessments.

The enrichment strategies described in this guidance are intended to increase the efficiency of drug development and support precision medicine, i.e., tailoring treatments to those patients who will benefit based on clinical laboratory, genomic, and proteomic factors. This guidance also discusses design options for enrichment strategies and discusses the interpretation of the results of studies that use enrichment strategies.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

[1] This guidance was developed by the Center for Drug Evaluation and Research in coordination with the Center for Biologics Evaluation and Research at the Food and Drug Administration. Although the principles and examples discussed in this guidance relate primarily to the safety and effectiveness of drugs and biologic products many of these principles also apply to studies for other medical products, including devices. FDA encourages a sponsor that is considering applying these principles to the study of a device to discuss such a proposal with the Center for Devices and Radiological Health, specifically with the organizational unit responsible for that product area.

II. BACKGROUND

Sponsors of investigational drug products use a variety of strategies to enrich the study population by selecting a subset of patients in which the potential effect of a drug can more readily be demonstrated.

Three broad categories of enrichment strategies as listed below are addressed in this guidance:

(1) Strategies to decrease variability — These include choosing patients with baseline measurements of a disease or a biomarker characterizing the disease in a narrow range (decreased interpatient variability) and excluding patients whose disease or symptoms improve spontaneously or whose measurements are highly variable (decreased intrapatient variability). The decreased variability provided by these strategies would increase study power (see section III., Decreasing Variability).

(2) Prognostic enrichment strategies — These include choosing patients with a greater likelihood of having a disease-related endpoint event (for event-driven studies) or a substantial worsening in condition (for continuous measurement endpoints) (see section IV., Prognostic Enrichment Strategies — Identifying High-Risk Patients). These strategies would increase the absolute effect difference between groups but would not be expected to alter relative effect.

(3) Predictive enrichment strategies — These include choosing patients who are more likely to respond to the drug treatment than other patients with the condition being treated. Such selection can lead to a larger effect size (both absolute and relative) and can permit use of a smaller study population. Selection of patients could be based on a specific aspect of a patient's physiology, a biomarker, or a disease characteristic that is related in some manner to the study drug's mechanism. Patient selection could also be empiric (e.g., the patient has previously appeared to respond to a drug in the same class) (see section V., Predictive Enrichment — Identifying More-Responsive Patients).

Enrichment characteristics can be dichotomous (e.g., sex, presence of genetic marker(s), a concomitant illness) or continuous (e.g., age, blood pressure (BP)). Studies using the latter would ordinarily dichotomize the continuous variable (e.g., BP over 160 systolic) or examine several different cut-offs (e.g., BP over 140, over 160).

The enrichment strategies described in this guidance are discussed primarily in the context of randomized controlled trials (but could also be relevant to other designs, such as single-arm studies or historically (externally) controlled trials). In almost all cases, the strategies for patient selection are prospectively planned and fixed prior to study initiation (with a few exceptions for adaptive strategies to be noted later). These strategies, therefore, generally do not compromise the statistical validity of the trials or the meaningfulness of the conclusions reached for the population actually studied. The illustrative examples and design options described have been used in the past, but they should not be regarded as an exclusive list or a limitation. FDA encourages the development of additional approaches as experience with these strategies grows.

The principal concerns with the use of enrichment strategies relate to the generalizability and applicability of the study results. When considering use of an enrichment design, sponsors should consider whether the enrichment strategy could be used in practice to identify the patients to whom the drug should be given and whether the drug might be useful in a broader population (see section VII.A., Summary — The Decision to Use an Enrichment Strategy). The extent to which patients should be studied who do not meet the selection criteria for enrichment (see section VII.B., Study of Marker-Negative Patients) is therefore a critical consideration. In addition, in the setting of predictive enrichment strategies, the accuracy of the measurements used to identify the enrichment population and the sensitivity and specificity of the enrichment criteria used to distinguish treatment responders and nonresponders are also critical issues.

III. DECREASING VARIABILITY

Approaches to increasing study power (the ability of a clinical trial to demonstrate a treatment effect if one is present) by decreasing heterogeneity (nondrug-related variability) are widely practiced. The following strategies are useful and generally accepted ways to decrease variability:

- Defining entry criteria carefully to ensure that enrolled patients actually have the disease that is being studied.

- Training investigators to adhere to protocol-specified entry definitions and criteria.

- Identifying and selecting patients likely to adhere to treatment to decrease variability in drug exposure. *Note*: Excluding poor compliers identified after randomization in the analysis is not acceptable because such patients are not likely to be a random sample of the study population and because adherence itself has been linked to outcome, even adherence to a placebo treatment (Coronary Drug Project Research Group 1980).

- Using placebo lead-in periods before randomization to eliminate patients who improve spontaneously or have large placebo responses.

- Decreasing intrapatient variability by enrolling only patients who give consistent baseline values (e.g., for BP measurements, treadmill exercise tests, pulmonary function tests, or patient-reported outcome measures).

- Excluding patients taking drugs that are pharmacologically similar to, or that could interact with, the study drug.

- Excluding patients unlikely to tolerate the drug.

- Excluding patients likely to drop out for nonmedical reasons (e.g., because they have difficulty getting to the study site).

- Excluding patients with comorbid illness that would make completing the treatment period unlikely.

Two of the strategies described above — encouraging adherence and reducing the number of spontaneous improvers or placebo responders — are further discussed in sections III.A., Encouraging Adherence, and III.B., Decreasing Placebo Responses and Spontaneous Improvement.

Some strategies to decrease variability can result in studies that provide too little information about the full range of patients who will receive a drug in clinical practice, such as the elderly, patients with multiple illnesses, and patients taking multiple drug therapies. It is not clear that concomitant illnesses that do not affect survival or other endpoint measurements or concomitant drugs unrelated to a test drug really do interfere with assessment of a treatment effect. Therefore, the implications of using these strategies should be carefully considered before they are used and should be balanced against the need for information in critical patient subgroups.

A. Encouraging Adherence

Practices that have become standard for ensuring adherence include: encouraging adherence by making patients aware of the conditions and demands of the trial, avoiding overly rapid titration of drugs that could cause intolerable early adverse reactions, using adherence prompts and alert systems, and counting pills (or using smart bottles to monitor drug use) so that nonadherent patients can be encouraged to perform better. On occasion, more protocol-specific efforts have been used to identify and enroll good adherers into clinical trials. For example:

- In the Veterans Administration Cooperative hypertension studies of the late 1960s and early 1970s, prospective patients were given placebo tablets containing riboflavin during a single-blind prerandomization period. Patients' urine was then examined for fluorescence; and randomized only patients whose urine fluoresced (evidence that the patients had been taking the riboflavin tablets) on two consecutive visits during a 2- to 4-month observation period were randomized (Veterans Administration Cooperative Study Group on Antihypertensive Agents 1967, 1970).

- The Physicians' Health Study used an 18-week, prerandomization placebo run-in during which patients (all physicians) self-reported whether they were taking the drug as specified in the protocol (Steering Committee of the Physicians' Health Study Research Group 1989). About one-third of the screened patients were not randomized because of self-reported poor adherence. Adherence during the randomized study was reported as a very satisfactory 90% over the 5 years of the study, greatly increasing its power (Lang et al. 1991).

B. Decreasing Placebo Responses and Spontaneous Improvement

In placebo-controlled trials of drugs for symptomatic conditions (e.g., depression, anxiety, angina) or laboratory/vital sign abnormalities (e.g., dyslipidemia, hypertension), use of a single-blind, placebo lead-in period and exclusion of placebo responders (i.e., randomize only patients

whose signs or symptoms remain above some threshold value) are relatively common. This is done to identify and not randomize patients who would have had an improvement for a reason other than a response to the test treatment (e.g., spontaneous improvement, a placebo response, an effect of expectations on observations) that resolved or reduced a patient's symptoms or signs, making the patient less likely to show a response to treatment. Also, many signs and symptoms vary spontaneously; therefore, initial screening values that would support enrollment may represent *random highs* of the disease course that would be followed by regression to the mean, leaving the patient with mild to no symptoms and little opportunity to no opportunity to benefit from a drug.

IV. PROGNOSTIC ENRICHMENT STRATEGIES — IDENTIFYING HIGH-RISK PATIENTS

Prognostic enrichment strategies are designed to increase the proportion of patients likely to have a particular disease-related endpoint event or substantial worsening in condition.

Event-based studies: For any given desired power in an event-based study, a study to lower the rate of cardiovascular (CV) events or tumor recurrence, the appropriate sample size will depend on effect size and the event rate in the control group. If the patients enrolled have a high event rate in the course of the study, the power of a study to detect any given level of relative risk reduction will increase. A wide variety of prognostic indicators has been used to identify patients with a greater likelihood of having the outcome event of interest (or a large change in a continuous measure of interest, e.g., worsening of symptoms). These prognostic indicators include clinical and laboratory measures, medical history, genomic, and proteomic measures, among others. In many cases patients at high risk for events are chosen for the initial outcome study of a drug, and if successful, larger studies in lower risk patients are conducted later.

Progression-based studies: Prognostic enrichment strategies are also potentially applicable to the study of drugs intended to delay progression of a variety of diseases, such as Alzheimer's disease, Parkinson's disease, rheumatoid arthritis, and multiple sclerosis, and other conditions for which patients likely to have more rapid progression could be selected. A prognostic marker may also be predictive (see section V., Predictive Enrichment — Identifying More-Responsive Patients). That is, the more rapidly progressing patients could be less responsive to treatment (i.e., rapid progression would be a negative predictor of response) or more responsive to treatment, as illustrated in the angiotensin converting enzyme (ACE) inhibitor case described in section IV.A.1., CV Studies.

For any given desired power in an event-based study, the appropriate sample size will depend on effect size and the event rate in the control group. Prognostic enrichment does not affect the relative risk reduction but will increase the number of events in a shorter time period, generally allowing for a smaller sample size. For example, reduction of mortality from 10% to 5% in a high-risk population is the same relative effect as a reduction from 1% to 0.5% in a lower risk population, but the reductions in absolute risk are 5% and 0.5%, respectively. The sample size needed would be smaller in the high-risk population. The increased absolute effect could also enhance the benefit-risk relationship in the studied population.

A. Experience With Prognostic Enrichment Strategies

1. CV Studies

In CV disease, the severity of the illness being studied and other factors that can indicate increased risk have been used to identify patients at greater risk for CV events, considerably reducing the sample sizes needed to show an effect in outcome studies. Examples include a history of recent myocardial infarction (MI) or stroke; the presence of concomitant illness such as diabetes or hypertension; and certain blood markers, such as very high low-density lipoprotein (LDL) cholesterol, low high-density lipoprotein (HDL) cholesterol and high C-reactive protein (CRP). Outcome studies using ACE inhibitors in heart failure (HF) and 5-hydroxy-3-methylglutaryl-coenzyme A (HMG CoA) reductase inhibitors in hyperlipidemia (the enalapril and statin trials, respectively) illustrate this approach.

In the enalapril trials, mortality reduction and decreases in morbid events (such as hospitalization) were first assessed in a very ill HF population of New York Heart Association Class IV patients (Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS)), then in less ill patients (Studies of Left Ventricular Dysfunction (SOLVD) treatment trial), and eventually in asymptomatic patients (SOLVD prevention trial) (CONSENSUS Study Group 1987; SOLVD Investigators 1991, 1992). In the later studies, composite endpoints were needed because the number of early deaths was too low to allow a mortality effect to be demonstrated; event rates were about 15% at 1 year on placebo in the SOLVD treatment trial and about 5% in the SOLVD prevention trial, far lower than the 44% 6-month mortality in CONSENSUS in the placebo group. The very high early mortality in CONSENSUS, together with the large effect size (40% reduction in mortality), allowed demonstration of a survival effect in just 253 very ill Class IV patients, while the studies in less ill patients required sample sizes of 2,000 to 4,000 patients. The higher risk patients enrolled as a result of prognostic enrichment showed, as expected, a larger absolute effect size, but relative effect size was also greater in the more ill patients, suggesting that severity was also a predictive marker (see section V., Predictive Enrichment — Identifying More-Responsive Patients).

More recently, patients with a ventricular ejection fraction less than 35% to 40% (normal 55–65%) and levels of plasma B-type natriuretic peptide (BNP) or N-terminal pro-BNP above specific levels were enrolled in a study of the neprilysin inhibitor sacubitril in HF because these markers were thought to reflect HF severity. The study showed a significant effect on the combined endpoint of CV death plus HF hospitalization, which occurred in the control group at 26.5% over about 2 years (McMurray et al. 2014).

A similar strategy of identifying people at very high risk was used in the statin trials. The early CV outcome trials with statins were able to evaluate the effects of the drugs on mortality because the patients enrolled in the trials had a history of heart disease and very elevated cholesterol levels, an indicator of patients whose mortality risk was substantial (Pedersen et al. 1994). As the benefit of statins became established in high-risk patient populations, patients with less marked LDL cholesterol elevations and without known coronary artery disease were enrolled in subsequent CV outcome trials. These populations had not yet been shown to benefit from LDL

cholesterol lowering and, therefore, could still be ethically studied; patients in those trials were identified as high risk because of some other illness (e.g., type 2 diabetes mellitus) or risk factor (e.g., low HDL cholesterol, elevated high-sensitivity CRP). As the population's risk became lower, sample sizes increased considerably, but prognostic factors made the studies possible. For example, in the JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trials Evaluating Rosuvastatin) study (n=17,802), a statin was shown to have an effect on outcome in patients with LDL cholesterol levels that were considered "normal," but who were at higher CV risk based on factors other than LDL cholesterol, including age, one additional CV risk factor, and a high-sensitivity CRP greater than or equal to 2 mg/L (Ridker et al. 2008). As the magnitude of risk declined in these study populations, relying on composite endpoints often became necessary because the mortality rate was too low to allow a mortality trial of reasonable size.

Choosing patients at relatively high risk of CV events can also be critical for safety studies to be able to rule out a given level of CV risk with a reasonably feasible study size. This approach is now recommended for the development of new antidiabetic treatments[2] and has been a consideration in the design of studies to evaluate the CV risk of nonsteroidal anti-inflammatory drugs.

2. Oncology Studies

Studies intended to show reduced or delayed occurrence or recurrence of cancer are clearly more likely to be successful in patients identified as high risk for such events (e.g., by genetic or other characteristics). For example, adjuvant therapy studies of tamoxifen showed that the drug not only delayed development of metastases in patients with breast cancer but also reduced the risk of contralateral tumors (new primary tumors) in this high-risk group (high risk because the patients previously had breast cancer). Tamoxifen was then studied in 13,000 high-risk women (calculated using the Gail model) without a prior diagnosis of breast cancer who were followed for 4 years (National Surgical Adjuvant Breast and Bowel Project (NSABP) P-1) (Fisher et al. 1998). The study showed a 44% relative reduction in risk of invasive breast cancer, and FDA approved tamoxifen for reducing the risk of breast cancer in high-risk individuals identified using the Gail model calculator. A study in patients at lower risk would have required a substantially larger sample size. For example, a study of patients with a risk that was 25% of the risk of the NSABP P-1 study population would have needed about 20,000 patients to detect an effect of the size observed in the NSABP P-1 study with 90% power.

3. Pulmonary Studies

History of a recent exacerbation in patients with chronic obstructive pulmonary disease (COPD) is generally thought to predict the likelihood of exacerbations in the subsequent year, and for this reason, most COPD clinical trials use history of exacerbation as an inclusion criterion (e.g., only

[2] See the guidance for industry *Diabetes Mellitus — Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes* (December 2008). We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

patients with at least 1 exacerbation in the past 12 months are enrolled). FDA has also qualified plasma fibrinogen as a prognostic biomarker for all-cause mortality and COPD exacerbations, improving the power of COPD clinical trials, even when history of exacerbation is also used for enrichment.[3]

4. Neurology Studies

Studies of drugs intended to decrease exacerbations of multiple sclerosis have generally required patients to have had a recent exacerbation (one in the last year or two in the last 2 years) or specified magnetic resonance imaging findings before entry.

B. Potential Strategies for Prognostic Enrichment

Examples of additional strategies that may prove useful for prognostic enrichment are described below. Whether these strategies are useful as enrichment tools is not yet established.

1. CV Studies

There may be additional approaches to identifying high-risk CV patients. A report in 2005 noted that a higher resting heart rate, a small increase in exercise heart rate, and delayed recovery of heart rate in a population of asymptomatic working men between the ages of 42 and 53 years were all strong predictors of sudden death, suggesting potential enrichment strategies in studies of drugs to prevent sudden death (Jouven et al. 2005). The potential for risk prediction based on genetic factors has been examined, as has the predictive value of coronary artery calcium score (Ripatti et al. 2010; Polonsky et al. 2010; Ioannidis and Tzoulaki 2010).

2. Oncology Studies

a. Prostate cancer

In men with localized prostate cancer following radical prostatectomy, high prostate-specific antigen (PSA) velocity (PSA increase greater than 2 ng/mL during prior year) has been reported to strongly predict prostate cancer recurrence and mortality over a 10-year period (D'Amico et al. 2004). Prognostic markers such as PSA velocity, if validated in future studies, could be used to identify high-risk patients. Studies of adjuvant treatment for prostate cancer would likely be better able to detect an effect on survival if patients with a high risk of death were enrolled.

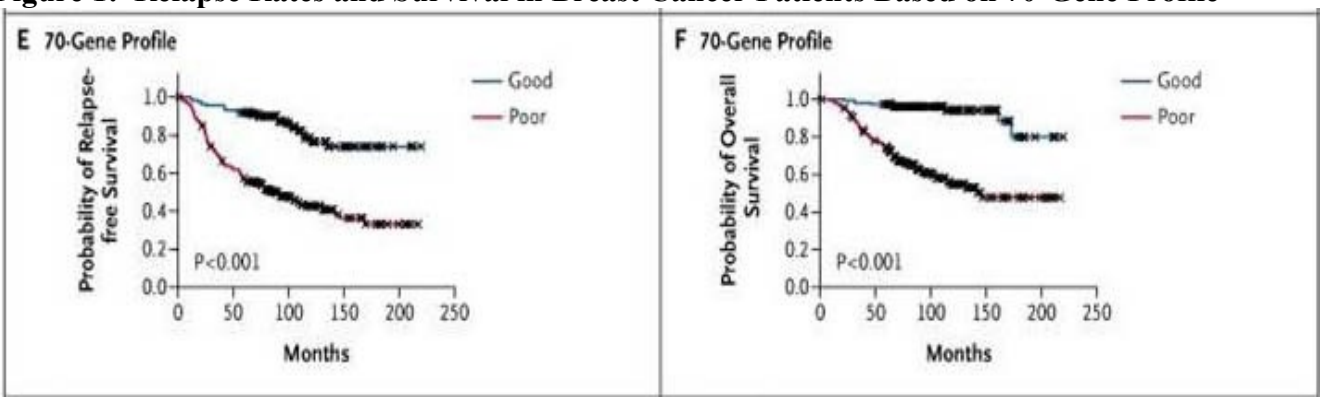
b. Breast cancer

Many investigators have reported gene expression profiles that appear to predict the likelihood of breast cancer recurrence after surgery. For an adjuvant therapy trial to be successful in showing a reduction in tumor recurrence and increased survival, selection of a population with a high rate of recurrence and poor survival is critical. In a report on the use of five different gene expression

[3] See the guidance for industry *Qualification of Biomarker — Plasma Fibrinogen in Studies Examining Exacerbations and/or All-Cause Mortality in Patients With Chronic Obstructive Pulmonary Disease* (September 2016).

profiling approaches in a nonrandomized 285-patient sample treated with local therapy, tamoxifen, tamoxifen plus chemotherapy, or chemotherapy alone, Fan et al. (2006) found that four of the five approaches for classifying patients had a striking ability to predict large differences in both recurrence and mortality (as illustrated in Figure 1) showing the difference between patients with *good* and *poor* 70-gene profiles. It is apparent that showing improvement in the low relapse population would require studies of enormous size. Studies in high-risk patients can be much smaller.

Figure 1: Relapse Rates and Survival in Breast Cancer Patients Based on 70-Gene Profile*



* Fan C et al., 2006, Concordance Among Gene-Expression-Based Predictors for Breast Cancer, N Engl J Med; 355(6): doi: 10.1056/NEJMoa052933.

FDA has cleared MammaPrint (an in vitro diagnostic test using the gene expression profile of fresh breast cancer tissue samples to assess a patient's risk for distant metastasis) as a prognostic test for certain breast cancer patients. As noted, use of such a diagnostic test represents a potential enrichment strategy for adjuvant trials to identify a population at higher risk for recurrence.

Women with a deleterious BRCA 1 or 2 mutation have a lifetime incidence of breast cancer and ovarian cancer of 60% and 15% to 40%, respectively, compared to a risk of 12% and 1.4%, respectively, in women without a BRCA mutation.[4] Selecting women with such markers for a primary prevention trial in breast cancer or ovarian cancer would increase the likelihood of cancer events during the trial, thereby permitting a smaller sample size and a shorter study.

V. PREDICTIVE ENRICHMENT — IDENTIFYING MORE-RESPONSIVE PATIENTS

There are many possible ways to identify patients more likely to respond to a particular intervention, and these strategies have long been used in clinical trials when selection of patients has been based on a specific aspect of pathophysiology, past history of response, or a disease characteristic that is related in some manner to the study drug's mechanism of action (e.g., genomic or proteomic factor). For example:

[4] See the National Cancer Institute web page BRCA1 and BRCA2: Cancer Risk and Genetic Testing at https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet.

- Congestive heart failure (CHF) can result from either systolic or diastolic dysfunction. Presumably, a population with systolic dysfunction would be more likely to respond in a study of an inotrope that would increase ventricular contraction.

- High renin status predicts a greater antihypertensive response to beta-blockers, ACE inhibitors, and angiotensin II receptor blockers (ARBs), all of which work by inhibiting aspects of the renin angiotensin system. A population with high renin hypertension would be more responsive than a general hypertension population in studies of drugs in these classes.

- Antibacterial drug effects are best evaluated in patients who are infected with an organism that is potentially responsive to the antibacterial drug. Most commonly, patients are randomized before the specific infecting organism is known, but only those patients with the type of organism targeted by the test antibacterial are evaluated for effectiveness in the primary analysis. The category of organism is a baseline characteristic, even though the assessment occurs post-randomization.

- An initial screening for response — a biomarker measurement (e.g., radiographic response, reduction of ventricular premature beats (VPBs)), early clinical response, or full-fledged clinical response — in an open-label prerandomization period can be used to identify a responder population that would then be randomized in the controlled study. This approach is of particular value when responders constitute only a small fraction of the overall population to be treated.

- A population of nonresponders to a different drug can be randomized to the new drug or to the drug they did not respond to. The comparison is enriched with respect to the active control comparison because the population is expected to have a poor response to the original drug compared to the test drug. These designs are not appropriate when effectiveness is critical to survival or another irreversible outcome or the intolerance is serious or life-threatening.

- Protein or genetic markers related to a drug's mechanism of action can be used to identify potential responders. Examples include use of human epidermal growth factor receptor 2 (HER2) overexpression in breast cancer to indicate responsiveness to trastuzumab (a monoclonal antibody that targets HER2) for breast cancer; tumor surface epidermal growth factor receptor (EGFR) measurements or mutations indicating responsiveness to EGFR tyrosine kinase inhibitors for lung cancer, and cystic fibrosis transmembrane conductance regulator mutation type indicating responsiveness to ivacaftor for cystic fibrosis.

- A protein or genetic marker shown to predict response, even without a documented mechanism of action.

Identifying a responder population (i.e., a subset of the overall population with a larger than average response to treatment) and studying this population in a clinical trial can provide two

major advantages: (1) increased study efficiency or feasibility and (2) an enhanced benefit-risk relationship for the patients in the subset compared to the overall population.

A. Increased Efficiency or Feasibility

Identification of a population with a high rate of response or a larger response greatly increases the chance that a study of an effective drug will be able to detect a treatment effect, if one exists, and allows a study to demonstrate this effect with a smaller sample size than would be needed for a study in an unselected population. The strategy can be particularly useful for early effectiveness studies because it can provide clinical *proof of concept* for later studies. When the treatment responder population constitutes only a small fraction of all patients, for example, 20% (a common situation in oncology settings), enrichment can permit a showing of effectiveness when a study in an unselected population may have difficulty showing any effect.

The prevalence of the enrichment marker and the relative effectiveness of the drug in the marker-positive and marker-negative populations will determine how much the enrichment strategy can reduce the sample size needed to adequately power a study. Table 1 illustrates how sample size ratios — the ratio of the number of subjects needed in an unselected population versus the number needed if only the marker-positive population is studied — change with varying prevalence of marker-positive patients and different magnitudes of treatment effect in marker-negative patients (treatment effect in marker-negative patients of either 0% or 50% of the effect in marker-positive patients). Table 1 assumes the classification of patients into positive versus negative is 100% accurate.

Table 1: Sample Size Ratios as a Function of the Prevalence of Marker-Positive Patients

Prevalence of Marker-Positive Patients	**Treatment Effect 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

In general, the lower the prevalence of marker-positive patients in the unselected population and the smaller the treatment effect in the marker-negative population, the more the sample size can be reduced in a study of marker-positive patients compared to a study in an unselected population. For example, when the prevalence of marker-positive patients in a population is only 25% and no treatment effect is expected in the 75% of patients who are marker-negative, the required sample size in a study of an unselected population would be 16 times the sample size needed for a study that included only marker-positive patients. Simon and Maitournam (2004) have presented a more detailed description of these results and the conditions under which the results were obtained.

B. Enhanced Benefit-Risk Relationship

Identification of a responder population can enhance the benefit-risk relationship of a drug by avoiding exposure and potential toxicity in patients who would not benefit from the drug. For drugs with significant toxicity and a low overall response rate in the overall population with the medical condition to be treated — factors that could deter further development — identifying a responder population could make a risk more acceptable and facilitate continued development and approval. For example, the significant survival advantage (approximately 5 months) seen with trastuzumab in the treatment of metastatic breast cancer in patients with HER2 overexpressing tumors (about 25% of breast cancers) ultimately supported use of the drug in the marker-selected population despite the significant cardiotoxicity that emerged (see further discussion in section V.C.2.c., Proteomic and genetic markers with known pathophysiologic effect). The much smaller effect (less than 2 months) that would have been observed in an unselected population (i.e., everyone with metastatic breast cancer), and the fact that only about one-fourth of patients would have benefited, might have made approval difficult to support in the face of the observed cardiotoxicity of the drug.

Identifying a more responsive population does not necessarily indicate that no benefit exists for the remaining population. It is therefore generally desirable to have some data in the nonselected (nonenrichment) population to determine whether that population responds less well or indeed does not respond at all. These data also can provide an assessment of safety in the nonselected population in the event that such patients are exposed after approval. The data in the nonselected population need not be obtained in the controlled trials supporting effectiveness but could be obtained in earlier studies showing absence of an effect (clinical or biomarker) in a clinical trial, absence of a critical pharmacologic effect, or even lack of an effect in pertinent nonclinical studies. A strong mechanistic rationale can make study of the nonenriched population unnecessary (e.g., study of effects in an infection caused by an organism clearly known to be resistant; study in a genetically determined disease such as cystic fibrosis, where some patients do not have the specific genetic variant affected by the treatment).

A trial intended to provide evidence of effectiveness to support approval could include a broad range of patients but be prospectively designed to evaluate in its primary analysis the effect in the enriched population subset. This is a standard (and unavoidable) approach when the baseline characteristic can only be determined after randomization (e.g., the infectious organism, tumor characteristic), but the approach (preferably with stratified randomization) is also valuable in other settings to gather some information on the marker-negative population.

C. Approaches to Predictive Enrichment

The discussion below considers five predictive enrichment strategies: (1) empiric strategies; (2) pathophysiologic strategies; (3) empiric genomic strategies; (4) randomized withdrawal studies; and (5) studies in nonresponders or patients intolerant to other therapy.

1. Empiric Strategies

With an empiric strategy, the selection of likely responders for a study is not based on any understanding of the basis for differences in response between patients with or without a particular characteristic, but on observations of a response during screening periods or prior experience with the drug or related drugs.

a. Open-label single-arm trial followed by randomization

A straightforward enrichment strategy, in cases when a treatment response can be identified shortly after treatment initiation, is to give the investigational drug to all patients; identify apparent responders, either on the planned study endpoint or on a biomarker or other short-term response thought to be predictive of clinical response; withdraw the treatment; and then randomize only responders into a placebo-controlled trial. This strategy is particularly useful when the rate is low. This strategy has been discussed in the past and was used throughout the 1970s to develop new antiarrhythmic drugs (Temple 1994; Roden et al. 1980). Patients were titrated on the investigational drug until they had an acceptable reduction of VPBs. Only the responding patients were then randomized into placebo-controlled trials, often fixed-dose, dose-response studies, and sometimes with active controls as well. This approach is useful only when the response to the drug is of relatively short duration after drug withdrawal. The approach would not be acceptable if withdrawal would be dangerous to the patient.

The Cardiac Arrhythmia Suppression Trial (CAST) — a study of the mortality effect of suppressing VPBs in patients with recent acute myocardial infarctions (AMIs) and at least six VPBs per hour — is one of the best-known studies conducted in apparent responders (Echt et al. 1991). Patients with greater than 10 VPBs per hour after an AMI were known to have a fourfold increase in the rate of sudden death. Previous failed attempts to show survival benefits with antiarrhythmics had been criticized because of the low rate of VPB suppression achieved. The CAST used an open-label screening period to identify responders to two drugs, encainide and flecainide, which were shown to be very effective in suppressing VPBs in a previous study (Cardiac Arrhythmia Pilot Study Investigators 1988). Only patients who had at least a 70% VPB reduction were randomized. Unfortunately, despite the enrichment effort, these antiarrhythmic drugs did not decrease mortality but instead more than doubled it. This result reflects either the inadequacy of the surrogate endpoint of VPB reduction as a predictor of an effect on mortality or, more likely given the adverse effect on survival, an *off-target* pro-arrhythmic effect of the test drugs. The result did not, however, reflect a problem with the study design. The enrichment design yielded a study capable of showing an effect of VPB suppression and allowed clear interpretation of the study, which showed, contrary to expectations, that even in VPB responders the drugs were not helpful and, indeed, were harmful.

Use of an initial open-label phase without a control group does raise some concerns that need to be considered. For example, the deaths that occurred during the screening period for CAST (which were not unexpected given the recent infarction) were difficult to interpret in an open, uncontrolled setting where all patients received the active drug. In CAST II (ethmosin versus placebo), the initial screen for VPB suppression used a randomized comparison of drug to placebo, with the responders then randomized into the placebo-controlled trial (CAST-II

Investigators 1992). This strategy showed that the drug used in the screen was itself lethal (19 deaths on ethmosin versus 1 on placebo), and the study was stopped early.

A similar problem was described in outcome trials of carvedilol in the treatment of chronic HF (Pablos-Méndez et al. 1998). These studies, unlike CAST, clearly showed a benefit of treatment. In two large studies, some patients were excluded during a run-in period because they could not tolerate carvedilol. Some of those patients died. The dropout rates during the subsequent controlled trials were undoubtedly decreased by the screening procedure that excluded patients intolerant to beta-blockers, and the results made carvedilol seem better tolerated than it would actually have been in patients starting therapy. The randomized comparisons and the benefits demonstrated are fully valid in these trials for the populations studied, but the benefits and risks facing unselected patients, who would be treated in clinical use of the drug, may be different from those benefits and risks observed in the clinical trial, requiring close attention to the screening period results. These results would need to be described in the drug's labeling.

There are many other outcome study settings in which it would be possible to select patients more likely to benefit from treatment. Patients with a lipid abnormality might be given the planned treatment in a screening period to evaluate their biochemical responses. For the randomized trial, only patients with a response of a certain size might be randomized, giving a greater mean effect on the lipid level and, presumably, a larger effect on outcome. That approach could be useful in an early outcome trial, but it would also be possible to randomize a broader population stratified by such an initial response with the intent of making the primary study endpoint the result in the high-response subgroup while also gaining some information about the less responsive group. Again, the response in such a selected group would not describe the response in an unselected population.

Active, open screening for empiric responders is particularly advantageous when a population includes subsets with potentially different responses to interventions that are not identifiable before treatment based on genetic or other pathophysiologic assessments. Although it is hard to know in advance when this is true, certain difficult-to-study conditions, such as irritable bowel syndrome or fibromyalgia, might be candidates for this approach (Temple 1994).

The overall strategy (open trial followed by randomization of responders) is an efficient way to document effectiveness, but it cannot be used prospectively to identify the responder population when the drug is used in clinical practice. In some cases, however, an early response could be used to determine who should stay on the drug, which is usually how symptomatic treatments are used in practice.

b. An individual's history of response to a treatment class

Information about prior responses to a drug in a pharmacologic class, if available, can be used to identify potential responders for a study of a new member of that class. As is the case with an open-label trial followed by randomization, use of patient history of response to a drug class can greatly increase the efficiency of a trial in demonstrating effectiveness. In most cases, however, it will not help identify the population to be treated in clinical practice.

A study enriched with prior responders to a pharmacologic class can be useful in demonstrating effectiveness at the proof-of-concept stage. This design may be particularly advantageous for randomized, fixed-dose, dose-response studies (the preferred dose-response study design as described in the ICH guidance for industry *E4 Dose-Response Information to Support Drug Registration* (November 1994)). A responder population provides a larger overall treatment effect and, therefore, a steeper dose-response curve, which generally allows for easier interpretation of the curve (identifying the steep area and plateau of the curve) and more precise characterization of dose-response, especially for doses providing near-maximum effects. For example, a dose-response study of indapamide in known responders to diuretics demonstrated mean decreases of 29/12 mmHg (systolic/diastolic) for the 2.5 mg dose and 37/15 mmHg for the 5 mg dose, an increase in effect with the 5 mg dose that was considerably larger than that seen in studies of unselected patients, where 2.5 mg and 5 mg gave similar results.

c. Factors identified in results from previous studies

Analyses of results of previous studies can sometimes point to a substantially greater effect in a specific subset of the overall population and provide a basis for studying that subset in a subsequent study, either as the sole population studied or as the identified primary endpoint subset in a study of a broader population. For example, isosorbide dinitrate/hydralazine hydrochloride, a treatment for severe HF, was approved on the basis of a placebo-controlled study in 1050 patients carried out entirely in self-identified blacks (the African-American Heart Failure Trial) (Temple and Stockbridge 2007; Taylor et al. 2004). The selection of a black population was based on two previous studies (the Vasodilator-Heart Failure Trials (V-HeFT) I and II) of an isosorbide dinitrate/hydralazine hydrochloride combination versus placebo in a racially mixed population that strongly suggested effectiveness in blacks (Cohn et al. 1986; Cohn et al. 1991). In those studies, the combination had not shown an overall benefit, but post hoc analyses revealed a nominally significant effect in black patients in V-HeFT I and apparent equivalence to enalapril in V-HeFT II. In contrast, there was little or no effect of the combination in whites in V-HeFT I and nominally significant inferiority of the combination to enalapril in whites in V-HeFT II. The replication of the observed effect in blacks was strong, with only a suggestion of a modest effect in whites, perhaps a third of the effect in blacks. A trial to establish this small effect in a white population would have required 16,000 patients. The product was approved for "self-identified blacks" only.

2. Pathophysiological Strategies

Pathophysiological strategies involve selection of likely responders to a drug based on a patient's individual physiology or on the assessment of disease pathophysiology that suggests that only certain patient subgroups will respond to the mechanism of the drug or that certain subgroups will respond better than others.

Indicators of an individual's pathophysiology could include biomarkers (e.g., a specific mutation that affects tumor proliferation), imaging findings, and possibly even demographic or clinical characteristics that correlate with certain disease phenotypes (e.g., age and race may associate with renin-angiotensin-aldosterone system activity in hypertension).

a. Metabolism of the test drug

For a drug that acts through an active metabolite, as is the case for the antiplatelet drug clopidogrel, patients may differ in their ability to metabolize the prodrug to its active metabolite. Some patients may not form the active metabolite at all, and others may not make enough to respond to the dose selected. Including these patients in a trial will dilute the overall drug effect and can also lead to inefficient or inappropriate use of the drug in practice if the two subsets of patients are not identified and treated differently. In some cases, adjusting (increasing) the dose in the poor metabolizers would be possible, but patients who cannot make the active metabolite at all should probably not be included in the trial or in the planned primary analysis. A closely related approach is the assessment of uptake of the test drug by a tumor (Stroobants et al. 2003). Historically, before treatment of thyroid tumors with I-131, a low dose was given to determine whether the tumor did, in fact, take up iodine and the extent of uptake so that the needed dose could be estimated.

b. Effect on tumor metabolism

Patients for a cancer trial can be selected by screening for an effect on a tumor metabolic response, as assessed by a positron emission tomography (PET) scan. For example, response to the tyrosine kinase inhibitors imatinib and sunitinib in patients with gastrointestinal stromal tumors (GISTs) has been shown to correlate well with metabolic responses (decreased tumor glucose utilization) assessed by 15F-fluorodeoxyglucose PET imaging (Prior et al. 2009). A trial could enroll only the identified metabolic responders or could enroll all patients, stratified by metabolic response, with the primary hypothesis to be tested being the treatment effect in the metabolic responder stratum.

c. Proteomic and genetic markers with known pathophysiologic effect

Increasingly, cancer treatments are directed at enzymatic, hormonal, or other functions that are tied to tumor cell surface or intracellular receptors and enzymes. The following examples illustrate use of proteomic markers, or genetic markers that are linked to a proteomic marker, that are known to be essential for the activity of the drug.

- Trastuzumab was developed to bind to the HER2 receptor, which is present on normal and malignant cells but is overexpressed in about 25% of breast cancers. Binding of trastuzumab to the HER2 receptor blocks receptor-mediated growth-stimulating intracellular signaling, decreasing cellular repair after chemotherapy and radiation therapy and also increasing apoptosis. In activity-estimating trials, antitumor activity in patients with lower levels of HER2 receptor expression (1+ by immunohistochemical staining) was minimal, so that efficacy trials in patients with metastatic disease were limited to patients with HER2 2+ or 3+ overexpression.

  In the treatment of metastatic disease, when added to either of two background regimens, trastuzumab was estimated to have increased median survival by about 5 months, about three to four times the effect that would have been expected in an unselected population, assuming no response (which a modest amount of testing showed was the case) in the

HER2-negative patients. Enrichment thus allowed a modest-size study to show a striking effect and directed treatment to the population that could benefit. In addition, because the drug was shown to be moderately cardiotoxic in the metastatic breast cancer trials, the focus on potential responders (i.e., patients with HER2 receptor overexpression) was considered critical when designing adjuvant studies (Smith et al. 2007).

- Evidence from the metastatic breast cancer setting has demonstrated that the likelihood of response to endocrine therapy is related to the hormone receptor status of the tumor. For example, when treated with tamoxifen, a selective estrogen receptor modulator, patients whose tumors express both estrogen receptors (ERs) and progesterone receptors (PRs) have a response rate of approximately 70%; patients whose tumors express either ER or PR, but not both, have a 20% response rate; and patients whose tumors are ER and PR negative have a response rate less than 5%. As a consequence, testing of all breast cancer specimens to direct decisions regarding endocrine therapy, in both the early stage and the advanced setting, has become the standard of care and would be expected in any trial of endocrine therapy (Early Breast Cancer Trialists' Collaborative Group 2005).

- A more recent illustration is the use of somatic mutations in the gene encoding the serine-threonine protein kinase BRAF to identify potential responders to vemurafenib in melanoma; 40% to 60% of all melanomas carry this activating mutation. In a study of 49 patients with melanoma, 11 of 16 patients with BRAF V600E mutation who received vemurafenib had a tumor response, compared to 0 of 5 without the mutation (the remaining 28 patients did not undergo BRAF mutation testing). The phase 3 trial compared vemurafenib to dacarbazine in 675 patients with metastatic or unresectable melanoma who had the BRAF V600E mutation. The trial was stopped after an interim analysis showed a 63% reduction in the risk of death with vemurafenib. The confirmed response rate was 48% for vemurafenib versus 5% for dacarbazine (Chapman et al. 2011).

The examples of pathophysiologic selection described above reflected, at least initially, variables related to cellular receptors that could be described as proteomic variables, but that were in many cases later identified as tumor genetic markers (e.g., EGFR and BRAF genetics). In such cases, the genetic marker may sufficiently define the pathophysiologic state.

When proteomic and genetic markers are used in an enrichment strategy, adequate characterization of the test for the marker is critical. An inaccurate assay will undermine an enrichment effort if the study aims to demonstrate superiority or noninferiority of the test treatment (Wang et al. 2011). It is also important to gain as much information as possible about the marker-response relationship.

3. Empirical Genomic Strategies

Studies directed at subsets of patients with specific genomic patterns that appear to be associated with outcomes (e.g., RNA expression profiles, single nucleotide polymorphism arrays) are becoming increasingly common. Although most genomic markers (e.g., for a tumor surface property) have been linked to a pathophysiologic property, this linkage is not essential. Use of a

genomic marker could instead be an empiric strategy, identifying subsets of responders without providing a pathophysiologic basis for the difference in response (i.e., before such a basis is recognized).

Simon and coauthors, for example, in Freidlin and Simon (2005) and Freidlin, Jiang, et al. (2010), have suggested that a trial population could be divided into two portions, with an unblinded exploratory analysis of many different genetic markers to identify a predictive classifier in the first portion. A confirmatory analysis would then be carried out in the biomarker-defined subgroup in the remaining portion of the trial. Treatment effects would then be evaluated in the overall population and the biomarker-defined subset from the remaining portion, with appropriate control of the type I error rate ensured. Any such approach would need scrupulous attention to maintaining the blind, perhaps by using an independent group to do the biomarker analysis and should be thoroughly discussed with FDA in advance.

4. Randomized Withdrawal Studies

In a randomized withdrawal study, patients who have an apparent response to treatment in an open-label period or in the treatment arm of a randomized trial are randomized to continued drug treatment or to placebo treatment. Because such trials generally involve only patients who appear to have responded, this is a study enriched with apparent responders, an empiric strategy. The study evaluation can be based on signs or symptoms during a specified interval (e.g., BP, angina rate), on recurrence of a condition that had been controlled by the drug (e.g., depression), or on the fraction of patients developing a rate or severity of symptoms that exceeds some specified limit (i.e., a failure criterion).

The randomized withdrawal design was initially proposed as a way to establish long-term effectiveness of drugs in settings in which long-term use of a placebo would not be acceptable on either ethical or practical grounds. Angina was the initial example, but this would also be true for most psychiatric conditions, pain treatments, and antihypertensive drug treatments (Amery and Dony 1975). A randomized withdrawal design in which the study population is on treatment for an extended duration followed by blinded, randomized withdrawal of treatment for a short duration can provide evidence of prolonged effectiveness with only brief exposure to the placebo. The design can allow a patient to be removed from the study (for having reached an endpoint) when the condition returns at a specified severity, avoiding long-term exposure to an ineffective treatment.

The randomized withdrawal design can also be used as an initial trial to show effectiveness when there is an existing population of patients in an open-label treatment setting (e.g., under an IND or as an off-label use of an approved drug), as illustrated by the cases of nifedipine and gamma-hydroxybutyrate (GHB).

The approval of nifedipine for vasospastic angina (the first drug approved for this condition) illustrates the utility of this design. An open-label, historically controlled trial was considered inadequate to support approval because the natural history of vasospastic angina was not well established (Antman et al. 1980). A randomized withdrawal design (see Figure 2) was conducted in patients already receiving the drug, with a primary endpoint of recurrence of severe

vasospastic angina leading to study withdrawal. A total of 28 patients participated in the study. One-third of the patients randomized to placebo withdrew early, as compared to no withdrawals in patients randomized to nifedipine (see Table 2).

Figure 2: Nifedipine Randomized Withdrawal Trial in Vasospastic Angina

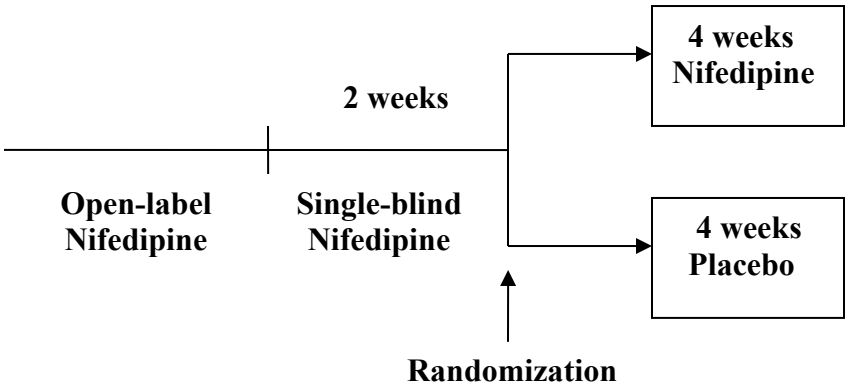


Table 2: Results of Nifedipine Randomized Withdrawal Study

	Nifedipine (n=13)	Placebo (n=15)
Early withdrawal	0	5*
Early withdrawal or AMI**	0	6*

* Statistically significant at $p \leq 0.05$
** Acute myocardial infarction

Another example in which patients already using a drug were studied involved GHB (sodium oxybate), which was approved for treatment of cataplexy on the basis of a single placebo-controlled study of conventional design and a second, small, randomized withdrawal study in 55 long-term (7 to 44 months) users randomized to 2 weeks of continued treatment with GHB or placebo. The second study produced a clinically and statistically impressive result ($p< 0.001$, as shown in Table 3) and needed little time for recruitment.

Table 3: Randomized Withdrawal Study of GHB* in Cataplexy

Median Attacks/2 Weeks		
Treatment Group	**Baseline**	**Change in Rate**
Placebo (n=29)	4.0	+21.0
GHB (n=26)	1.9	0

* GHB = gamma-hydroxybutyrate

By randomizing patients to different doses, the randomized withdrawal design can also be used to obtain long-term dose-response data. For example, this design was used to demonstrate effectiveness of a single weekly dose of fluoxetine in preventing recurrence of depression;

patients on 20 mg per day were randomized to a placebo, fluoxetine 70 mg per week as a single dose, or continued 20 mg per day. Both fluoxetine groups were superior to the placebo group in reducing the rate of recurrence. This design has become the standard for studies to demonstrate the ability of psychotropic drugs to prevent recurrent depression, psychosis, and anxiety.

5. Studies in Nonresponders or Patients Intolerant to Other Therapy

A comparative study can be enriched by selection of patients who failed to respond to an existing drug or who were intolerant of that drug. Although nonresponsive or treatment-intolerant patients are not more likely than an unselected population of patients to respond to or tolerate the new drug and would not enrich a placebo-controlled trial, they would generally be less likely to respond to or tolerate the existing comparator drug, giving the test drug a potential advantage if it was in fact more effective or better tolerated. Because patients in a trial sometimes respond to a drug to which they had previously failed to respond, in most cases studies in nonresponders are informative for the between-drug comparison of effectiveness only if patients are randomized to both the new and failed drug (i.e., not simply placed onto the new drug in a single-arm study or randomized to new drug versus placebo). This approach can provide important information to practitioners; it is critical to know whether another member of a pharmacologic class or a member of a different class can be useful in patients who fail on previous treatments. A drug's effectiveness in nonresponders can be a critical component of a risk-benefit assessment (i.e., it can allow approval of a drug with a substantial risk). It should be appreciated that for life-threatening diseases that are progressing, it may not be ethical to randomize to the failed treatment. The approach may be useful in two settings:

(1) Where the drug has a different mechanism of action from the previous treatments, it may be more effective in nonresponders to those treatments even if it is not more effective in an unselected population.

(2) Where the treatment effect of a new drug is moderately superior to the existing drug in an unselected population, but where a very large study would be needed to show superiority if the study included unselected patients, many of whom would respond to the less effective drug. For example, if the new drug response rate is 90% and the existing drug response rate is 80%, a study with 90% power to detect that 10% difference would require about 600 patients. In contrast, if only nonresponders to the existing drug were randomized (20% of the patients treated with the existing drug), few would respond to the existing drug; but if half of those patients responded to the new drug, the difference would be detectable with fewer than 40 of the nonresponders.

Note: In neither case would showing an advantage for the new drug in nonresponders to existing therapy establish superiority of the new drug in an unselected population.

a. Studies in nonresponders

Several past studies in nonresponders illustrate the value of this study design.

Captopril

To support a claim in severe hypertension unresponsive to other agents, a study was designed to evaluate patients who had not responded to standard triple therapy (propranolol 320 mg, hydrochlorothiazide 100 mg, and hydralazine 200 mg). Those patients, who had diastolic pressure that was severely elevated, were observed for 1 to 2 weeks on the same regimen (triple therapy lead-in), and if their diastolic pressure did not exceed a defined limit, the patients were randomized to the same standard triple therapy they had failed on, or to captopril, with a 2:1 captopril to triple therapy randomization ratio. The number of responders (diastolic pressure less than 90 mmHg or a decrease by at least 10 mmHg) clearly favored captopril in this difficult-to-treat population (see Table 4).

Table 4: Results of the Captopril Severe Hypertension Trial for the Group Randomized to Captopril or Triple Therapy

Time Period	DBP* Findings	Captopril (n=66)	Triple Therapy (n=30)
Week 4	Normalized DBP ≤ 90 mmHg	21 (32%)	5 (17%)
	Reduction in DBP ≥ 10 mmHg	8 (12%)	3 (10%)
Week 8	Normalized DBP ≤ 90 mmHg	22 (33%)	4 (13%)
	Reduction in DBP ≥ 10 mmHg	14 (21%)	3 (10%)

* DBP = diastolic blood pressure

Note: Approximately 25% of the triple therapy *nonresponders* actually responded (diastolic blood pressure less than 90 or 10 mmHg fall) to the previously failed therapy in the new trial. This finding reinforces the need for randomization to the new and reportedly failed therapy in a study of nonresponders. A study that had merely switched patients from the failed therapy to the new one and found 25% responders might have been interpreted as showing an effect of the new drug in the nonresponders to the failed therapy when, in fact, the response to the new drug was no better than the response to the failed therapy would have been in a new study. The finding of an effect in nonresponders helped overcome concerns about agranulocytosis seen with high doses of captopril.

Clozapine

Clozapine is an antipsychotic drug associated with serious toxicity, a greater than 1% rate of potentially fatal agranulocytosis. For clozapine to be approved, the sponsor had to show that the drug offered a clear advantage over safer alternatives. To show this, a study was conducted in hospitalized schizophrenic patients with a history of poor response to neuroleptics who, in addition, had failed to respond to 6 weeks of treatment with haloperidol. These patients were randomized to 4 weeks of treatment with clozapine or chlorpromazine plus benztropine. The results showed a striking advantage for clozapine on Clinical Global Impression and the Brief Psychiatric Rating Scale standard measures in antipsychotic drug trials (see Table 5). Despite its

serious risk, clozapine was approved for use in patients not responding to other antipsychotic drugs.

Table 5: Results of Clozapine Study in Nonresponders to Standard Psychotropic Drugs

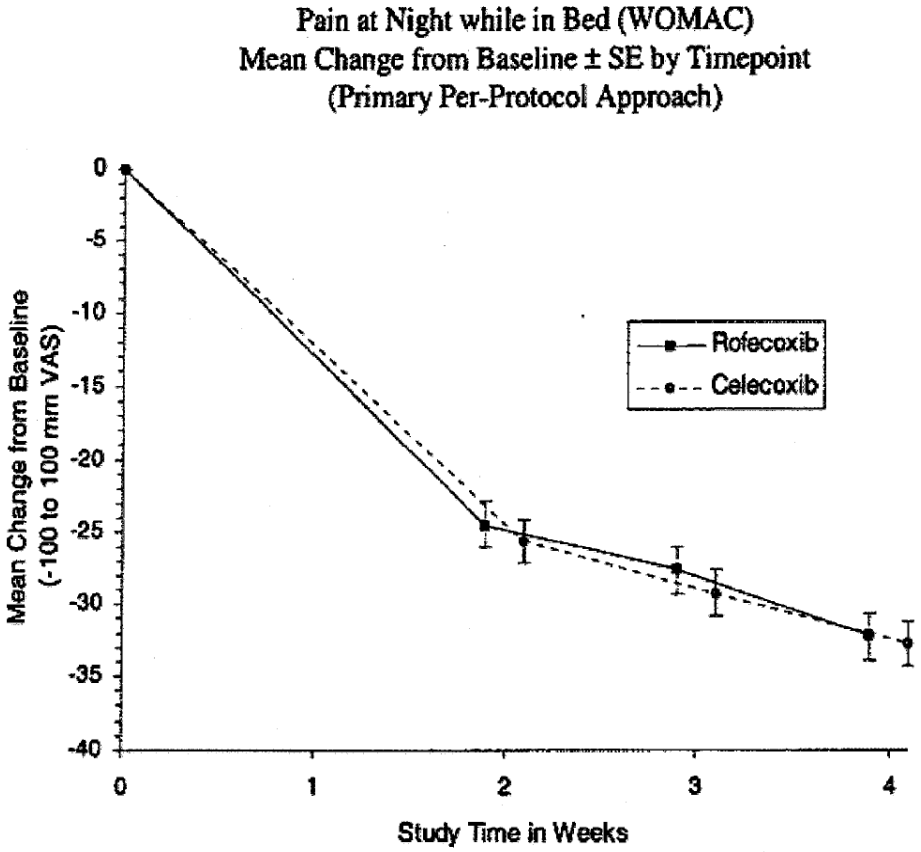
	Response (%)	
Measure	**Clozapine**	**Chlorpromazine**
CGI* (decrease > 1)	71	37
BPRS* items (decrease > 1)		
Conceptual disorganization	60	39**
Suspiciousness	64	42**
Hallucinations	59	51
Thought content	15	2**
CGI and BPRS	15	2**

* CGI = Clinical Global Impression; BPRS = Brief Psychiatric Rating Scale
** $p < 0.05$

Rofecoxib

Individual patients are often believed to respond differently to different nonsteroidal anti-inflammatory drugs. To examine this belief, a controlled trial was conducted in which osteoarthritis patients identified as nonresponders to celecoxib were randomized to celecoxib or rofecoxib (Temple 2012). In fact, no difference was observed between the treatments (see Figure 3).

Figure 3: Study Comparing Rofecoxib to Celecoxib in Celecoxib Nonresponders*



* Temple R, 2012, A Regulator's View of Comparative Effectiveness Research, Clin Trials, 9:56–65.

Considerable and prompt improvement in pain was reported in both groups. A baseline-controlled, single-arm trial of rofecoxib would have led to a clearly erroneous conclusion that rofecoxib was effective in celecoxib nonresponders and even a placebo-controlled trial of rofecoxib in this population might have shown an effect that would have been incorrectly interpreted as an effect in celecoxib nonresponders because whether they would have responded to celecoxib in the controlled trial setting would not be known.

b. Study in intolerants: ARBs in patients who cough on lisinopril

Studies of the tolerability of a new drug in patients who do not tolerate a previous treatment are also informative and efficient. Comparative studies in an unselected population could provide some information on relative tolerability, but a very large study would be needed to show small differences. For example, if the true rates of cough for an ACE inhibitor and an ARB were 5% and 1%, respectively, a study with 90% power to show a difference in an unselected population would need about 800 patients. In contrast, a study in patients known to cough on ACE inhibitors would need fewer than 20 patients, if, for example, the cough rate were greater than 90% in the ACE inhibitor arm and 20% in the ARB arm.

This approach was used in a study of 84 elderly hypertensive patients with history of coughing on an ACE inhibitor (Chan et al. 1997). Cough was confirmed by rechallenge with the ACE inhibitor lisinopril, administered for up to 8 weeks, which had to cause at least moderate coughing for patients to continue in the study. Lisinopril was then withdrawn for 4 weeks, and coughing had to disappear. The patients were then randomized to losartan 50 mg, lisinopril 10 mg, or metolazone (diuretic active control that does not induce coughing) for 10 weeks. The study achieved a very persuasive result with this small population (see Table 6), a result cited in losartan's labeling.

Table 6: Comparison of Coughing Rates With ACE Inhibitor, ARB, and an Additional Active (Non-Cough-Inducing) Control

	Lisinopril (n=28)	**Losartan (n=28)**	**Metolazone (n=28)**
Any cough	97%	18%*	21%*

* $p < 0.001$, lisinopril vs metolazone and losartan

As was the case in studying nonresponders to previous therapy, randomization to the poorly tolerated previous therapy and the new drug is critical to reach a conclusion that the new drug has superior safety because the adverse reactions do not always reappear when a treatment is repeated. For the same reason, FDA recommends that sponsors should include a placebo group (or, as in the above example, an active drug clearly lacking the adverse effect) to be certain that the adverse reactions were indeed reproduced in the previous treatment group. This study design is not feasible if the adverse reaction is dangerous to the patient.

VI. ENRICHMENT STUDY DESIGN AND OTHER CONSIDERATIONS

A. General Considerations

An enrichment design should be explicitly described in the protocol and final report and should fully detail the enrichment maneuvers and their effect on the interpretation of results (see section VI.E., Cautions in Interpretation).

Some enrichment strategies depend on a screening measurement for selecting the enriched patient population. For both prognostic and predictive markers, understanding the accuracy and performance characteristics of the test used to identify patients or marker-defined subgroups for enrichment is important. Note that the question of who to include in the study and what patients to analyze in the primary analysis are distinct issues and may differ for predictive and prognostic enrichment.

For prognostic enrichment, the ideal screening measurement to be used for patient selection will have good sensitivity and specificity. That is, it will reliably identify patients with a greater likelihood of having disease-related endpoint events or substantial worsening in their conditions (high-risk group) and reliably exclude patients in the lower risk group, both of which will serve to enhance the trial's efficiency. If the screening measurement's specificity is poor, many

patients who will not have clinical worsening will be selected, and the effect of enrichment will be diluted. If sensitivity is poor, patients who are at high risk will be missed in the screening process, delaying enrollment into the study. Note, however, that although exclusion of lower risk patients will increase absolute effect size, allowing a smaller trial, the relative effect is expected to be similar in higher and lower risk patients. This is in contrast to predictive enrichment, where patients without the enrichment factor would be expected to have a lower relative effect size.

Uncertainties about the performance characteristics of an enrichment strategy based on a prognostic marker have different implications for superiority and noninferiority studies. For a superiority study, a prognostic enrichment strategy with poor sensitivity and/or specificity can increase the sample size or duration required to observe the needed number of endpoint events. Although this will reduce the efficiency gained through enrichment, it will not lead to a false conclusion of effectiveness. For a study intended to demonstrate noninferiority, however, the use of a prognostic enrichment strategy with poor ability to select subjects for entry into the study could have the effect of increasing the chances of reaching a false noninferiority conclusion because the control effect size would be smaller than expected from historical experience (Wang et al. 2011). In general, patients in a noninferiority study should be selected according to criteria similar to the past studies of the active control.

For predictive enrichment, a screening measurement of a biomarker with both high sensitivity and specificity for identifying responders is also desirable. A screening measurement cut-off can be set to *cast a wide net,* i.e., with high sensitivity giving a high probability of identifying and treating patients who are more likely to respond to the treatment (higher sensitivity), at the risk of also treating more nonresponder patients (lower specificity). Alternatively, the measurement cut-off can be set to identify and exclude from treatment patients who might not respond (higher specificity), at the risk of also excluding more responder patients. These trade-offs can be explored using receiver operating characteristic analyses.

The performance characteristics of enrichment strategies based on predictive markers also have different implications for superiority and noninferiority studies. For a superiority study, predictive markers with poor sensitivity and specificity will lead to increased sample size but will not increase the chance of a false conclusion of efficacy. For a noninferiority study, the effect on the type I error rate is more complex and depends on whether the marker is pertinent to both treatments or only one treatment; again, however, patients should generally be selected according to criteria used in past studies of the active control.

B. Which Populations to Study

As will be described in more detail below, trials can be designed to: (1) include only patients with the enrichment factor; or (2) include patients with and without the enrichment factor, but with an intent to analyze only the patients with the enrichment factor as one of the primary study hypotheses. This is a potentially critical step in the predictive enrichment setting where patients without the factor are not expected to respond, thereby diluting the effect size if included. Studies including both populations need not include a wholly unselected population of patients with the disease to be treated but can designate separate sample sizes for patients with and

without the enrichment characteristic to allow collection of sufficient information to demonstrate effectiveness in the enriched subgroup and to allow a reasonable estimate of effect in the nonenriched group. Many design alternatives have been discussed in the literature (Mandrekar and Sargent 2009; Freidlin, McShane, et al. 2010).

A critical question in all settings in which enrichment is used is therefore to what extent the marker-negative population should be studied. In some cases, study of the general population (one including the marker-negative population) would not be expected. For example, if prognostic enrichment is used to ensure that there are sufficient events to make a trial feasible, even if the treatment effect is thought to also be present in the lower risk population without the marker (but at a lower absolute effect size), it may not be possible to design a trial that includes a significant fraction of the marker-negative population without greatly increasing the sample size — a strategy that may make the trial impractical and defeat a major purpose of prognostic enrichment. Whether to use the drug in the unstudied marker-negative population would depend on the particular circumstances. For example, the presence of significant toxicity could lead to doubts about the advisability of using the drug in the lower risk population. The variability-reducing factors discussed in section III., Decreasing Variability, would not ordinarily call for study of the population lacking the enrichment factor (e.g., patients with poor adherence).

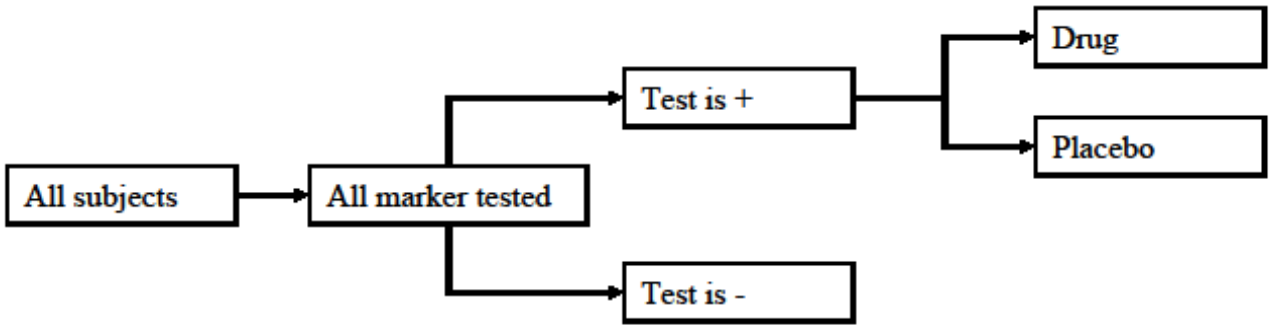
It is principally in the area of predictive enrichment, especially predictive enrichment using a pathophysiological measurement or biomarker, that the question of studying the population without the enrichment factor is most germane. Experience shows that the selected enrichment factors often do not precisely dichotomize patients into subpopulations that will and will not respond so that obtaining some information on the marker-negative population to assess performance of the factor is usually desirable. However, even an imperfectly characterized predictive marker can greatly increase the power and likelihood of success of a study. Moreover, in treating serious and life-threatening illnesses (especially when alternative treatments exist), using the test treatment in patients thought unlikely to respond raises critical ethical issues.

Efforts to use predictive enrichment thus offer a number of design choices. The study designs illustrated below are fixed sample size designs that can be used with predictive enrichment strategies (also see section VI.D., Adaptive Enrichment, concerning adaptive enrichment and nonfixed sample size). The examples describe trials intended to show superiority of the test treatment to a control (e.g., placebo, standard of care), but noninferiority studies would present similar issues.

1. Studying Marker-Positive Patients Only

A study randomizing only marker-positive patients is shown in Figure 4. Because a study that uses a marker-positive only population will provide no direct information about the marker-negative population, its use should generally be limited to situations in which information about the marker-negative population is not needed or is not feasible given the objectives of the study. For example, if it appears clear based on mechanistic, nonclinical, or early clinical data that the marker-negative patients will have no or minimal response or would be exposed to unreasonable risk, inclusion of the marker-negative patients would, in most cases, not be justified.

Figure 4: Study Design of Predictive Enrichment Trial With No Possible Effect in the Marker-Negative Group



The study shown in Figure 4 would support an effectiveness claim in the enriched population, but it would overstate the actual effectiveness for an unselected population, so that the fraction of marker-positive patients would be important information. The study would provide no new clinical evidence with respect to the marker-negative population and would not further characterize the predictiveness of the marker because of a lack of ability to compare effectiveness in marker positive and marker negative patients. Because no effectiveness or safety information on the enrichment-marker-negative patients would exist, the selection process should be fully described in labeling and in clinical practice patients would usually be assessed for the enrichment marker before exposure. Moreover, because assessment of marker status is critically important to determining whether the drug will be effective in patients, the test to assess the enrichment marker that would be used after the drug's approval would be an established, FDA-cleared or -approved, laboratory test explicitly labeled for this purpose as a companion diagnostic, although exceptions can be considered for a major advance in treatment.[5]

2. Studying Both Marker-Positive and -Negative Patients

FDA encourages inclusion of some predictive marker-negative patients in most trials intended to provide primary effectiveness support, unless earlier studies have established that the marker-negative patients do not respond or a strong mechanistic rationale makes it clear that they will not respond. Significant toxicity of the test drug could reduce the level of the evidence needed to conclude that the drug should not be studied in marker-negative patients. In general, the greater the uncertainty about the marker cutoff and responsiveness of marker-negative patients, the more important inclusion of a reasonable sample of marker-negative patients becomes. When substantial incentive exists to use the drug in the marker-negative population (e.g., for serious diseases with few alternative therapies), characterization of the response in the marker-negative population is more important, especially if the drug has important safety concerns.

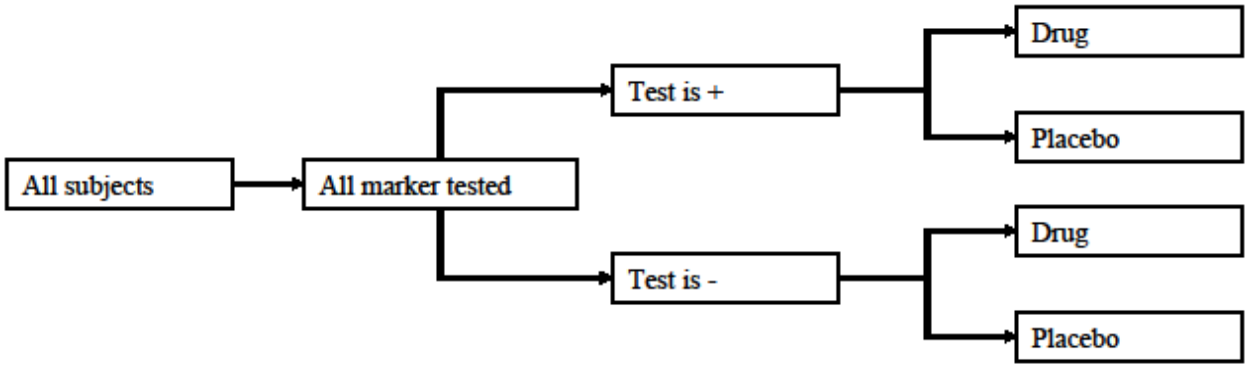
There are two cases to consider in studies that include both marker-positive and marker-negative patients: (1) when the marker can be assessed before randomization; and (2) when the marker can be assessed only after randomization. Figures 5 and 6 provide sample study designs for these two cases (Freidlin, McShane, et al. 2010).

[5] See the guidance for industry and FDA staff *In Vitro Companion Diagnostic Devices*.

In the first case (see Figure 5), marker status is determined for all patients, and randomization is stratified by marker status. The primary study objective would usually be a statistically rigorous demonstration of the treatment effect in the marker-positive patients, and the study would be powered for the effect in that group. The size of the marker-negative group would be determined separately, and randomization of all marker-negative patients would not be necessary. Because the treatment effect would be expected to be much smaller (if there were any effect) in the marker-negative population, the size of the marker-negative population would usually be too small to provide a definitive answer on the effect; however, the marker-negative patients would provide at least some estimate of the effect in that population.

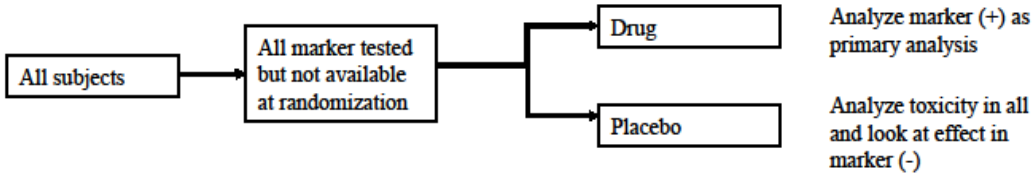
This design could also enable a risk-benefit assessment for the drug when used in the overall population with the medical condition to be treated, which would be advantageous if some exposure in marker-negative patients is anticipated in clinical practice (e.g., because the test is not widely available). When substantial uncertainty exists about whether a marker is predictive (i.e., whether it can select a population in which treatment is effective), the primary endpoint could be the effect in the overall population, or study alpha could be divided between the two endpoints (overall population and marker-positive population). In each of these scenarios, the hypotheses should be clearly specified and control of the type I error rate should be addressed (see section VI.C., Type I Error Rate Control for Enriched Study Subpopulations). In addition, tests for heterogeneity of treatment effects (i.e., treatment by biomarker interactions) should be prespecified.

Figure 5: Study Design of Predictive Enrichment Trial With Possible Effect in the Marker-Negative Group



The second case (see Figure 6) is one in which a drug is expected to be effective only in the marker-positive subset (e.g., only in patients with a targeted type of organism), but the drug must nonetheless be given to all patients because the marker result is not available at randomization. Having the primary study outcome be the effect in the marker-positive subset would still be appropriate, but the risk-benefit assessment would reflect results in the entire population (i.e., the population that would be exposed to treatment). In these cases, when the marker test results will not be known before drug administration and when no patient management decisions will be made on the basis of the test result (e.g., decisions to discontinue treatment in marker-negative patients), FDA clearance or approval of the test contemporaneously with approval of the drug is not needed.

Figure 6: Study Design of Predictive Enrichment Trial Where All Patients Are Randomized Because Marker Cannot Be Assessed Before Randomization of the Marker-Negative Group



C. Type I Error Rate Control for Enriched Study Subpopulations

Generally, even if patients both with and without a predictive enrichment characteristic are studied, the primary endpoint is expected to be driven by the result in the enriched subgroup. For enrichment designs that enroll patients both with and without the enrichment characteristic, the type I error rate for the study can be shared between a test conducted using only the enriched subpopulation and a test conducted using the entire population. While allowing the assessment in the enriched subgroup, this alpha allocation scheme also allows for assessment of the treatment effect in the entire population when there may be some effect in patients without the enrichment characteristic. Determining the required sample size needed to provide reasonable power to test the different hypotheses while controlling the type I error rate (usually including a prespecified order of testing or a multiple testing procedure allowing testing of both hypotheses) is challenging. The interpretation of the study finding would have to take into consideration the magnitude of the effect in the nonenriched group.

D. Adaptive Enrichment

Enriching a clinical trial using prognostic or predictive markers can pose a challenge if there is uncertainty at the planning stage about the performance characteristics of the enrichment strategy, such as incomplete information on the prevalence of the marker and/or the type and strength of the marker-outcome relationship. This can be a particular problem for predictive enrichment. A study design that incorporates planned adaptations to the enrichment strategy, taking advantage of information gained on marker performance during the course of a clinical trial, may be useful in meeting these challenges. For example, the prospective determination of the sample size required to detect a particular treatment effect with sufficient power, either in the overall population or in a marker-defined subpopulation, can be difficult when there are uncertainties about the characteristics of the enrichment strategy, such as:

- The cutoff value of the biomarker used to identify patients in the marker-defined subgroup
- The proportion of patients who are in the marker-positive subgroup
- The magnitude of the treatment effect in patients with and without the marker

With a fixed sample enrichment design, such unknowns create uncertainty as to whether the study will successfully meet its objectives. An adaptive design in which the sample size and other design features can be modified in a prospectively planned manner to adapt to information obtained during the course of the trial (e.g., the frequency of marker positivity) may enhance the value of certain enrichment strategies. Similarly, determination that the enrichment factor has a greater or lesser effect on response than anticipated or that the patients without the enrichment factor have a higher or lower response than anticipated, could trigger planned changes in sample size or entry criteria in response to the accumulating information. Such changes should be prospectively planned and would often need appropriate type I error rate control to account for interim, unblinded analyses of the accumulating data as well as analyses of multiple subgroups. If the only change was increased sample size based on blinded, pooled results because the prevalence of the marker-defined subgroup was lower than expected, there would be no need for a type I error rate adjustment.

A few examples of adaptive enrichment trials have been reported in the literature. The Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and Molecular Analysis Phase 2 (I-SPY 2) is an example of a phase 2 screening trial using an adaptive enrichment strategy. This is a clinical trial for women with newly diagnosed, locally advanced breast cancer. The trial is designed to evaluate the effect of investigational drugs added to standard chemotherapy compared to standard chemotherapy alone on pathological complete response rates. The treatments being studied are neoadjuvant therapies, that is, drugs administered before surgery. The treatment phase of this trial involves testing multiple investigational drugs that are thought to target identified biomarkers of each patient's tumor type. The trial is an adaptive enrichment design in that biomarkers are used at baseline to identify patients likely to respond to treatment, and the accumulating responses are then used to inform treatment assignments for subsequent participating patients as the trial progresses. The results of I-SPY 2 are intended to inform the design of future trials for establishing efficacy and safety of those treatments that advance to phase 3 (Barker et al. 2009).

Although there are challenges with enriched study designs that involve preplanned changes to the enrichment strategy based on information accrued during the trial (e.g., change sample size after the start of the study), the following three examples illustrate potential adaptation strategies:

(1) In a study that includes both marker-positive and -negative patients, and with a primary endpoint that is the effect in the marker-positive group, a planned interim look could reveal, either on an early endpoint (e.g., imaging, pharmacodynamic biomarker, tumor response rate) or later endpoint (e.g., progression-free survival), that the marker-negative population has a much lower response than the marker-positive group. Additional enrollment of marker-negative patients could be reduced or stopped entirely, provided an adaptive enrichment strategy was planned and was appropriately accounted for in statistical analyses.

(2) Planned interim analyses could suggest changing entry criteria to increase enrollment in what may be a better responding subgroup. If such an adaptive enrichment strategy is undertaken, appropriate control of the type I error rate may need to be demonstrated, depending on the analyses used. As noted above, although type I error rate adjustments

would be needed for interim unblinded assessments of efficacy, no additional adjustment would be needed for changing the entry criteria per se, as long as all randomized patients based on the originally planned sample size were included in the final analysis (Mehta et al. 2009).

(3) A study could be designed to obtain more precise information on the performance characteristics of the marker or other characteristics used for enrichment, for example, by examining an early endpoint using several different biomarker cutoff values to determine the optimal cutoff value. If a cutoff value proved too low or too high, it could be changed. Such plans would need to be specified in advance, and appropriate type I error rate control demonstrated.[6]

E. Cautions in Interpretation

Any use of an enrichment design should be explicit in the protocol and study report and should fully detail the rationale for the design, the specific enrichment maneuvers, and their effects on the interpretation of results. For example, if only half of the patients screened passed the entry test, that should be noted, and the effect of this selection in terms of the expected response rate in the overall population and on the generalizability of the results should be evaluated. The importance of such descriptions is obvious for trials in which high-risk patients (prognostic enrichment) and probable responders (predictive enrichment) have been selected, where the description is critical to knowing to which patients the results apply, but such descriptions are important for all types of enrichment studies. Given the potentially complex interpretation of studies using enrichment designs, plans to use them should be discussed with FDA early in development.

When enrichment depends on a proteomic or genetic biomarker, particularly if a test for the biomarker is intended for use in practice to identify patients to be treated, the analytical validity of the test is critical. In addition to assay validity, for any marker used to select patients, even a familiar one, the biomarker's clinical sensitivity, specificity, and positive and negative predictive values should be well characterized. To the extent an enrichment strategy successfully identifies patients with high event rates or high response rates and leads to a successful study, study results could be said to speak for themselves (i.e., the randomized trial *did* show an effect; the event rate *was* high enough) and certainly support the effectiveness of the drug in the population studied. Again, however, the enrichment strategies should be clearly described in the study report and labeling to indicate how the drug is to be used and to whom the results may apply (groups of patients that are known to benefit or not to benefit and groups in which effect is not known).

Selection of the optimal predictive enrichment study design (specifically, whether to include both marker-positive and -negative patients and whether to introduce adaptive elements) can be difficult to determine when there is uncertainty about the properties of the enrichment marker. Many publications have addressed these issues (Freidlin and Simon 2005; Mandrekar and Sargent 2009; Freidlin, McShane, et al. 2010). One conclusion is that the greater the uncertainty

[6] The draft guidance for industry *Adaptive Designs for Clinical Trials of Drugs and Biologics* (September 2018). When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

regarding the marker cutoff and responsiveness of marker-negative patients, the more sense it makes to include a reasonable sample of marker-negative patients, perhaps using an adaptive design to exclude such patients if they are seen not to respond. In general, stratification by marker status would be sensible, especially when marker prevalence in the population studied is relatively low.

VII. ENRICHMENT — REGULATORY ISSUES

A. Summary — The Decision to Use an Enrichment Strategy

The decision to use an enrichment design is largely left to the sponsor of the investigation, but like the overall research and clinical communities, FDA is very interested in targeting treatments to the patients who can most benefit from them (i.e., individualization or "precision medicine"). FDA's interests also include the adequacy of the study (Will it successfully assess effectiveness in a defined population and, in so doing, support marketing approval?) as well as how study findings can be described in drug labeling. As discussed above, a critical question is almost always how much data will be needed in the off-target (nonenriched) population, particularly for predictive enrichment strategies. These issues are critical aspects of a development program that should be discussed with FDA early in development.

There are many reasons to use enrichment designs, including an enhanced benefit-risk relationship if a population with an increased likelihood of response can be identified, and efficiency in drug development, as smaller studies can often be used to demonstrate effectiveness. Sponsors, however, should consider the following two critical regulatory issues when contemplating the use of enrichment designs.

1. Does the Enrichment Strategy Identify the Patients to Whom the Drug Should Be Given?

When patients with an increased likelihood of response can be defined before treatment by a predictive marker (e.g., a pathophysiologic or genomic characteristic, a short-term screen such as response to a test dose), a straightforward method is available for selecting patients for treatment. In contrast, some empiric strategies that provide predictive enrichment (e.g., studying known responders in a conventional study or in a randomized withdrawal study) can efficiently establish the effectiveness of a drug in a subset of the population but provide no way for prescribers to prospectively identify patients with a greater likelihood of response or to predict the magnitude of response in an unselected patient. Although this type of untargeted treatment may seem troubling (treatment of many to attain a response in only some), the reality is that this is generally the case with treatments that are approved on the basis of conventional studies in a nonenriched population, where there is typically a wide range of responses, including no effect at all or even harm in some patients. However, it needs to be understood that the magnitude and/or likelihood of a treatment response for an unselected patient could be substantially less than the mean response observed in a clinical trial that employed an empiric enrichment strategy, and this should be clear in labeling. When the prescriber is able to quickly gauge the effectiveness of a drug in an individual patient (e.g., pain is relieved, cholesterol is reduced), the pretreatment

ability to predict the likelihood of a drug response with accuracy may not be as critical as it would be if response is delayed.

In some cases, enrichment cannot be used to select patients for study because the enrichment factor is not known until after the treatment is initiated but is used to identify the subset of the studied population to be analyzed — for example, the patients with a prerandomization sample of tissue, sputum, etc. that shows (later, after treatment is started) the presence of a targeted type of organism in studies of antimicrobial drugs. Again, the subset analysis documents effectiveness, but the population entered into the trial and randomized to treatment groups, at least initially, will be the unselected patients (i.e., a larger group than the population of potential responders). Such situations are unavoidable, however, if treatment is urgent and must be initiated before the enrichment test results are available.

Finally, knowing the precise utility of the enrichment strategy may not be critical to determining that a drug has an effect (that is if the selected study population shows an effect, the drug was in fact effective), but knowing the effect of enrichment and the difference in effect between populations are important for therapeutic use of the drug. It is plainly undesirable to treat patients who will not respond and to fail to treat potential responders because the enrichment marker has poor performance characteristics or because cutoff points for the marker were poorly selected.

These problems notwithstanding, if the enrichment strategy enables a drug of value to be developed and to be shown to be effective when disease and response variability would make nonenriched studies unable or unlikely to succeed, there is clearly an important gain from use of such strategies. Labeling will reflect limitations and concerns, but it seems clear that a drug shown to be effective and safe in an enriched study should be available even if the responder population is not identified as precisely as would be desirable.

2. Might the Drug Be Useful in a Broader Population Than Was Studied?

The data that should be obtained for the marker-negative patients are considered below, but it can be anticipated that less information will be available about those patients and there will be greater uncertainty as to their responses to the treatment. Studies in unselected patients (i.e., a nonenriched population), the typical basis for drug approval, simply ignore the question of identifying treatment responders and lead to treatment of many patients who will not benefit. There would thus seem to be a gain from a process that seeks to establish the characteristics that predict a drug response, rather than one that ignores the varied responses and overcomes the variability by simply increasing sample sizes.

In general, then, FDA's regulatory standards do not bar the Agency from approving drugs whose effectiveness has been demonstrated primarily or even solely in enriched populations, and FDA will seek to ensure truthful labeling that does not overstate the likelihood of a response, the magnitude of the response, or the predictiveness of the enrichment factor. But the extent of data that should be available on the nonenriched subgroup should always be considered. Postmarketing commitments or requirements may be requested to better define the full extent of a drug's effect (including efficacy and safety studies and trials in a broader population).

B. Study of Marker-Negative Patients

Well-controlled enrichment studies, if successful, provide evidence of effectiveness in the population studied. In many cases, however, questions will remain about whether some effect of the treatment occurs in marker-negative patients and the extent to which the treatment effect should be characterized in this population. A related issue is whether labeling should discourage use in marker-negative patients and, if so, how strongly such use should be discouraged. In general, heterogeneity-reducing efforts raise few issues of this kind, but prognostic enrichment strategies, and especially predictive enrichment strategies, do raise them. The solutions are not always clear and may be circumstance specific.

In general, studies using prognostic enrichment have been accepted as a basis of approval without a requirement to study broader populations. Nonetheless, when studies have shown that a drug reduced the rate of serious or irreversible endpoints, showing an effect in a high-risk population (e.g., high BP, high LDL with a history of MI, severe CHF), they have, historically, been followed by later assessments of effects in lower risk patients. Subsequent studies in lower risk populations have generally shown an effect, but the effect size has often been smaller, and the study endpoints have sometimes changed (mortality in the high-risk studies versus composite endpoints in the lower risk patients). Consequently, benefit-risk considerations may have changed. FDA has generally accepted the results from prognostically enriched studies, approved an indication based on the observed effect, and described the study, including the patient population, in the CLINICAL STUDIES section of labeling, with any enrichment selection criteria noted. In most cases, the specific patient population studied has been the indicated population in the INDICATIONS AND USAGE section of labeling, in addition to its description in the CLINICAL STUDIES section. Labeling should also note populations that were not studied.

When predictive enrichment is used in drug development and the treatment represents an important advance for the marker-positive group, delaying approval because of limited data in the marker-negative group would generally be unreasonable. Nonetheless, in such cases, potential effectiveness in the marker-negative group is of great interest, and questions will therefore arise about whether there is a treatment effect, even if a smaller effect, in the marker-negative patients and about the precision of the dividing line chosen to define marker positivity.

It must be appreciated that a study sized to show a treatment effect in the predictively enriched population, even if there is a modest representation of marker-negative patients, will have relatively little capacity to detect or rule out an effect in the marker-negative population. Nevertheless, the marker-negative population will generally have at least *some* data, as the study designs in Figures 5 and 6 indicate. In addition, information on possible differences in the treatment effect based on the enrichment factor can sometimes be obtained from pathophysiologic studies, nonclinical studies, in vitro clinical studies, or combinations of various kinds of information.

Determining the need to characterize the treatment effect in the marker-negative population will be based on potential benefit and risks in that population, including: (1) the nature of the efficacy shown in the marker-positive population (which could range from reduction in

symptoms to enhanced survival); (2) the risks of the drug, in particular risks that might be present in the marker-negative population; (3) whether the effect of treatment would be apparent to an individual patient; (4) the relative sizes of the marker-positive and -negative populations; and (5) the desire to use the drug in the marker-negative population (the impetus is greater when benefit is substantial and/or alternative treatments are limited).

When treatment benefit is clinically critical (e.g., leading to prolonged survival or prevention of significant disability), physicians may want to prescribe the drug in the marker-negative population, especially if treatment options are limited, and obtaining a reasonably reliable assessment of the effect in the marker-negative group is especially important.

When treatment benefit is not clinically critical, it is important to consider the expected toxicity in the marker-negative population, the relative size of the marker-positive and marker-negative populations, and how much evidence there is that no treatment effect occurs in the marker-negative population. The balance of these factors would determine both the need to discourage or limit use through labeling and other measures and the need to characterize the treatment effect in the marker-negative population.

When individual patients can determine whether they are deriving benefit from a drug (usually for drugs used to reduce symptoms in the short term), patients can decide to discontinue a drug if it fails to provide symptom benefit. In this situation, determining the treatment effect in the marker-negative population is generally less important. This is because if the drug is used in the marker negative population, the determination of whether or not to discontinue the drug is made in a straightforward manner based upon the patient's report. How reassuring this will be depends substantially on the seriousness of the drug's adverse effects.

For drugs used to reduce the risk of morbid outcomes, patients typically cannot determine whether or not they will ever derive benefit, and such drugs will usually be used for long periods of time. When the drug has the potential to cause significant harm in the marker-negative population, a greater need to assess the treatment effect in that population arises. Also, the larger the relative size of the marker-negative population, the more important characterization of the benefit in that population becomes. For example, if only 10% of the population is marker-positive, and marker-negative (90%) patients are not known to benefit from the drug, sponsors should assess the benefit in the marker-negative population and communicate the findings. Under these circumstances, the benefit of the treatment in the marker-positive population could be outweighed by the overall risk in the marker-negative population, if this population is treated extensively, and communicating both the risk and the lack of benefit in that population would be important.

A number of considerations may support collection of less (or sometimes no) information on the enrichment-factor-negative population:

- A clear pathophysiologic basis for concluding that the nonenriched population will not respond (e.g., because the patients lack the molecular target of the drug, or because they cannot convert a prodrug to its active metabolite). This type of support could be provided

from nonclinical or clinical pharmacology and biomarker studies. In some cases, such information could be obtained from the literature.

- Early clinical studies that show very marked differences in response between the enrichment and nonenrichment populations.

Sponsors should discuss how much information to collect on the enrichment-factor-negative population with FDA review staff.

C. Labeling

The use of enrichment designs will have implications for labeling, especially in the INDICATIONS AND USAGE and CLINICAL STUDIES sections. Labeling should accurately describe the enrichment strategies used, including any limitations or concerns they raise for clinical use of the drug. Sponsors should discuss the potential effect of enrichment strategies on labeling with FDA during drug development.

APPENDIX:
ADDITIONAL GUIDANCE RELATED TO ENRICHMENT

FDA has issued a number of guidances that provide additional and related information about clinical trial designs (including enrichment designs) and demonstrating effectiveness. See especially the following guidances.[1]

- The draft guidance for industry *Adaptive Designs for Clinical Trials of Drugs and Biologics* (September 2018)[2] considers the case of enrichment approaches introduced only after randomization and based on interim evaluations. Such a retrospective finding would have to be carefully implemented and highly compelling to be accepted without further study.

- The guidance for industry *Clinical Pharmacogenomics: Premarket Evaluation in Early-Phase Clinical Studies and Recommendations for Labeling* (January 2013) focuses particularly on use and evaluation of genomic strategies in early drug development and highlights identification of enrichment options for later trials.

- The guidance for industry and FDA staff *In Vitro Companion Diagnostic Devices* (August 2014) defines in vitro companion diagnostic devices that are essential for the safe and effective use of their corresponding therapeutic products. The guidance describes FDA's policies for approval and clearance and for labeling companion diagnostics contemporaneously with approval and labeling of the therapeutic product.

- The guidance for industry *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (May 1998) describes the amount and type of evidence needed to demonstrate effectiveness and is applicable to studies using enrichment designs.

[1] We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

[2] When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

REFERENCES

Antman E, Muller J, Goldberg S, MacAlpin R, Rubenfire M, Tabatznik B, Liang C, Hempler F, Achuff S, Reichele N, Geltman E, Kerin N, Nelf R, and Paramawalt E, 1980, Nifedipine Therapy for Coronary Artery Spasm, N Engl J Med, 302:1269–1273.

Amery W and Dony J, 1975, A Clinical Trial Design Avoiding Undue Placebo Treatment, J Clin Pharmacol, 15:674–679.

Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, and Esserman LJ, 2009, I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy, Clin Pharmacol Ther, 86(1):97–100, doi: 10.1038/clpt.2009.68.

Cardiac Arrhythmia Pilot Study (CAPS) Investigators, 1988, Effects of Encainide, Flecainide, Imipramine and Moricizine on Ventricular Arrhythmias During the Year After Acute Myocardial Infarction. The CAPS, Am J Cardiol, 61:501–509.

Cardiac Arrhythmia Suppression Trial-II Investigators, 1992, Effect of Antiarrhythmic Agent Moricizine on Survival After Myocardial Infarction, N Engl J Med, 327:227–233.

Chan P, Tomlinson B, Huang TY, Ko JT, Lin TS, and Lee YS, 1997, Double-Blind Comparison of Losartan, Lisinopril, and Metolazone in Elderly Hypertensive Patients With Previous Angiotensin-Converting Enzyme Inhibitor-Induced Cough, J Clin Pharmacol, 37:253–256.

Chapman P, Hauschild A, Robert C, et al., 2011, Improved Survival With Vemurafenib in Melanoma With BRAF V600E Mutation, N Engl J Med, 364:2507–2516.

Cohn JN, Archibald DG, Ziesche S, Franciosa JA, Harston WE, Tristani FE, et al., 1986, Effect of Vasodilator Therapy on Mortality in Chronic Congestive Heart Failure. Results of a Veterans Administration Cooperative Study, N Engl J Med, 314:1547–1552.

Cohn JN, Johnson G, Ziesche S, Cobb F, Frances G, Tristani F, et al., 1991, A Comparison of Enalapril With Hydralazine-Isosorbide Dinitrate in the Treatment of Chronic Congestive Heart Failure, N Engl J Med, 325:303–310.

CONSENSUS Study Group, 1987, Effects of Enalapril on Mortality in Severe Congestive Heart Failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS), N Engl J Med, 316:1429–1435.

Coronary Drug Project Research Group, 1980, Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project, N Engl J Med, 303:1038–1041.

D'Amico AV, Chen MH, Roehl KA, and Catalona WI, 2004, Preoperative PSA Velocity and the Risk of Death From Prostate Cancer After Radical Prostatectomy, N Engl J Med, 351:125–135.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005, Effects of Chemotherapy and Hormonal Therapy for Early Breast Cancer on Recurrence and 15-Year Survival: An Overview of the Randomised Trials, Lancet, 365:1687–1717.

Echt DS, Liebson PR, Mitchell LB, Pears RW, Obias-Manno D, Barker AH, Arensberg D, Baller A, Friedman L, Greene HL, Huther ML, Richardson DW, and the CAST Investigators, 1991, Mortality and Morbidity in Patients Receiving Encainide, Flecainide, or Placebo: The Cardiac Arrhythmia Suppression Trial, N Engl J Med, 324:781–788.

Fan C, Oh DS, Wessels L, Welgeh B, Nuyten DS, Nobel AB, van't Veer LJ, and Perou CM, 2006, Concordance Among Gene-Expression-Based Predictors for Breast Cancer, N Engl J Med, 355:560–569.

Fisher B, Constantino JP, Wickerham DL, et al., 1998, Tamoxifen for Prevention of Breast Cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study, J Natl Cancer Inst, 90:1371–1388.

Freidlin B, Jiang W, and Simon R, 2010, The Cross-Validated Adaptive Signature Design, Clin Cancer Res, 16:691–698.

Freidlin B, McShane LM, and Korn EL, 2010, Randomized Clinical Trials With Biomarkers: Design Issues, J Natl Cancer Inst, 102:152–160.

Freidlin B and Simon R, 2005, Adaptive Signature Design: An Adaptive Clinical Trial Design for Generating and Prospectively Testing a Gene Expression Signature for Sensitive Patients, Clin Cancer Res, 11:7872–7878.

Ioannidis JP and Tzoulaki I, 2010, What Makes a Good Predictor? The Evidence Applied to Coronary Artery Calcium Score, JAMA, 303:1646–1647.

Jouven X, Empana JF, Schwartz PJ, Desnos M, Cowben D, and Ducimetiere P, 2005, Heart-Rate Profile During Exercise as a Predictor of Sudden Death, N Engl J Med, 352:1957–1958.

Lang JM, Buring JE, Rosner B, Cook N, and Hennekens CH, 1991, Estimating the Effect of the Run-in on the Power of the Physicians' Health Study, Stat Med, 10:1585–1593.

Mandrekar SJ and Sargent DJ, 2009, Clinical Trial Designs for Predictive Biomarker Validation: Theoretical Considerations and Practical Challenges, J Clin Oncol, 27:4027–4034.

McMurray JVV, Packer M, Desai AS, Giong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K, and Zie MR, 2014, Angiotensin-Neprilysin Versus Enalapril in Heart Failure, N Engl J Med, 371:993–1004.

Mehta C, Gao P, Bhatt DL, Harrington RA, Skerjanec S, and Ware J, 2009, Optimizing Trial Design: Sequential, Adaptive and Enrichment Strategies, Circulation, 119:597–605.

Pablos-Méndez A, Barr RG, and Shea S, 1998, Run-in Periods in Randomized Trials: Implications for the Application of Results in Clinical Practice, JAMA, 279:222–225.

Pedersen TR, Olsson AG, Faergeman O, Kjekshus J, Wedel H, Berg K, Wilhelmsen L, Haghfelt T, Thorgeirsson G, Pyörälä K, and Miettinen T, 1994, Randomised Trial of Cholesterol Lowering in 4444 Patients With Coronary Heart Disease: The Scandinavian Simvastatin Survival Study (4S), Lancet, 344:1383–1389.

Perez EA, Romond EH, Sumam VJ, et al., 2007, Updated Results of the Combined Analysis of NCCTG N9831 and NSABP B-31 Adjuvant Chemotherapy With/Without Trastuzumab in Patients With HER2-Positive Breast Cancer, J Clin Oncol, 25:65 (suppl 185, abst 512).

Polonsky TS, McClelland RL, Jorgenson NW, et al., 2010, Coronary Artery Calcium Score and Risk Classification for Coronary Heart Disease Prediction, JAMA, 303:1610–1616.

Prior JO, Montemurro M, Orcurto MV, et al., 2009, Early Prediction of Response to Sunitinib After Imatinib Failure by 18F-Fluorodeoxyglucose Positron Emission Tomography in Patients With Gastrointestinal Stromal Tumor, J Clin Oncol, 27:439–445.

Ridker PM, Danielson E, Fonseca FA, et al., 2008, Rosuvastatin to Prevent Vascular Events in Men and Women With Elevated C-Reactive Protein, N Engl J Med, 359:2195–2207.

Ripatti S, Tikkamen E, Orho-Melander M, et al., 2010, A Multi-Locus Genetic Risk Score for Coronary Heart Disease: Case-Control and Prospective Cohort Analyses, Lancet, 376:1393–1400.

Roden DM, Reele SB, Higgis SB, Mayol RF, Gammans RE, Oates JA, and Woosley RL, 1980, Total Suppression of Ventricular Arrhythmias by Encainide, N Engl J Med, 302:877–882.

Shepherd J, Cobbe SM, Ford I, et al. for the West of Scotland Coronary Prevention Study Group, 1995, Prevention of Coronary Heart Disease With Pravastatin in Men With Hypercholesterolemia, N Engl J Med, 333:1301–1307.

Simon R and Maitournam A, 2004, Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials, Clin Cancer Res, 10:6759–6763.

Smith I, Procter M, Gelber RD, et al., 2007, 2-Year Follow-up of Trastuzumab After Adjuvant Chemotherapy in HER2-Positive Breast Cancer: A Randomized Controlled Trial, Lancet, 369:29–36.

SOLVD Investigators, 1991, Effect of Enalapril on Survival in Patients With Reduced Left Ventricular Ejection Fractions and Congestive Heart Failure, N Engl J Med, 325:293–302.

SOLVD Investigators, 1992, Effect of Enalapril on Mortality and the Development of Heart Failure in Asymptomatic Patients With Reduced Left Ventricular Ejection Fractions, N Engl J Med, 327:685–691.

Steering Committee of the Physicians' Health Study Research Group, 1989, Final Report on the Aspirin Component of the Ongoing Physicians' Health Study, N Engl J Med, 321:129–135.

Stroobants S, Goeminne J, Seegers M, et al., 2003, 18FDG-Positron Emission Tomography for the Early Prediction of Response in Advanced Soft Tissue Sarcoma Treated With Imatinib Mesylate (Glivec), Euro J Cancer, 39:2012–2020.

Taylor AL, Ziesche S, Yancy C, Carson P, D'Agostino R Jr, Ferdinand K, et al., 2004, Combination of Isosorbide Dinitrate and Hydralazine in Blacks With Heart Failure, N Engl J Med, 351:2049–2057.

Temple R, 1994, Special Study Designs: Early Escape, Enrichment Studies in Non-Responders, Commun Stat Theory Methods, 23:499–531.

Temple, R, 2012, A Regulator's View of Comparative Effectiveness Research, Clin Trials, 9: 56–65.

Temple R and Stockbridge N, 2007, BiDil for Heart Failure in Black Patients: The US Food and Drug Administration Perspective. Ann Int Med, 146:57–62.

Veterans Administration Cooperative Study Group on Antihypertensive Agents, 1967, Effects of Treatment on Morbidity in Hypertension: Results in Patients With Diastolic Blood Pressures Averaging 115 Through 129 mmHg, JAMA, 202:1028–1034.

Veterans Administration Cooperative Study Group on Antihypertensive Agents, 1970, Effects of Treatment of Morbidity in Hypertension II: Results in Patients With Diastolic Blood Pressure Averaging 90 Through 114 mmHg, JAMA, 213:1143–1152.

Wang SJ, Hung HMJ, and O'Neill RT, 2011, Genomic Classifier for Patient Enrichment: Misclassification and Type I Error Issues in Pharmacogenomics Non-Inferiority Trial, Stat Biopharm Res, 3:310–319.

Wang SJ, O'Neill RT, and Hung HMJ, 2010, Statistical Considerations in Evaluating Pharmacogenomics-Based Clinical Effect for Confirmatory Trials, Clin Trials, 7:525–536.

A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Ansalan Stewart, 240-402-6631, ansalan.stewart@fda.hhs.gov; (CBER) Outreach and Development, 800-835-4709 or 240-402-8010; (CDRH) Office of the Center Director, CDRHClinicalEvidence@fda.hhs.gov; Office of Good Clinical Practice, 301-796-8340; or Office of Regulatory Affairs (ORA) ORAHQBIMOInspectionPOC@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biological Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Office of Good Clinical Practice (OGCP)
Office of Regulatory Affairs (ORA)

March 2019
Procedural

A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm
and/or
Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm
and/or
Office of the Center Director
Center for Devices and Radiological Health
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 66, Room 5431
Silver Spring, MD 20993-0002
Email: CDRH-Guidance@fda.hhs.gov
https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/default.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biological Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Office of Good Clinical Practice (OGCP)
Office of Regulatory Affairs (ORA)

March 2019
Procedural

Contains Nonbinding Recommendations

Draft — Not for Implementation

TABLE OF CONTENTS

A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers Guidance for Industry[1]

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION

This document provides guidance on risk-based approaches to monitoring investigational studies of human drug and biological products, medical devices, and combinations thereof. This guidance contains recommendations on planning a monitoring approach, developing the content of a monitoring plan, and addressing and communicating monitoring results. This guidance expands on the guidance for industry *Oversight of Clinical Investigations – A Risk-Based Approach to Monitoring* (August 2013) (the RBM guidance)[2] by providing additional guidance to facilitate sponsors' implementation of risk-based monitoring.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND

Sponsors of clinical investigations involving human drugs, biological products, medical devices, and combinations thereof are required to provide oversight to ensure adequate protection of the rights, welfare, and safety of human subjects and the quality of the data submitted to FDA.[3]

[1] This guidance has been prepared by the Office of Medical Policy in the Center for Drug Evaluation and Research in cooperation with the Center for Biologics Evaluation and Research, the Center for Devices and Radiological Health, the Office of Good Clinical Practice, and the Office of Regulatory Affairs at the Food and Drug Administration.

[2] We update guidances periodically. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

[3] 21 CFR part 312, subpart D generally (Responsibilities of Sponsors and Investigators) and 21 CFR part 812, subpart C generally (Responsibilities of Sponsors).

FDA's regulations require sponsors to monitor the conduct and progress of their clinical investigations.[4,5,6] The regulations are not specific about how sponsors are to conduct such monitoring; therefore, a range of approaches to monitoring are compatible with the regulations.

The RBM guidance discusses the importance of identifying critical data and processes necessary for human subject protection and integrity of the investigation, conducting a risk assessment, and developing a monitoring plan specific to the investigation. The RBM guidance also encourages sponsors to tailor monitoring plans to the needs of the investigation, describes factors to consider in developing a monitoring plan, and provides examples of monitoring methods and techniques.

FDA believes risk-based monitoring is an important tool to allow sponsors to identify and address issues during the conduct of clinical investigations. FDA's experience since finalizing the RBM guidance in 2013 suggests that additional guidance would be beneficial regarding FDA's recommendations for planning a monitoring approach, developing the content of monitoring plans, and addressing and communicating monitoring results. The following questions and answers are intended to assist sponsors in planning and conducting risk-based approaches to monitoring.

III. QUESTIONS AND ANSWERS

A. Monitoring Approach

Q1. What is the purpose of the risk assessment and should sponsors document their methodologies and activities for assessing risk?

Consistent with the RBM guidance, sponsors should identify and perform a risk assessment on those critical data and processes that are necessary for human subject protection and integrity of the investigation.

The risk assessment serves to identify and understand the nature, sources, likelihood of detection, and potential causes of risks that could affect the collection of critical data or performance of critical processes. The risk assessment informs the development of a monitoring plan and may also support efforts to manage risks across a clinical investigation (for example, through modifying the protocol design or implementation) or across a product's development program. Therefore, sponsors should document their risk assessment, including methodologies used for the

[4] 21 CFR 312.50 requires a sponsor to, among other things, ensure "proper monitoring of the investigation(s)" and "that the investigation(s) is conducted in accordance with the general investigational plan and protocols contained in the IND." 21 CFR 812.40 states that sponsors are responsible for, among other things, "ensuring proper monitoring of the investigation, …"

[5] See also 21 CFR 312.53(d), 312.56(a), 812.43(d), and 812.46.

[6] For the purposes of this guidance, the terms investigation, trial, and study are used interchangeably to refer to a clinical investigation, consistent with how these terms are used in the 2013 RBM guidance, *Oversight of Clinical Investigations – A Risk-Based Approach to Monitoring.*

risk assessment, conclusions from the risk assessment, and how the assessment was used to make decisions on the management of the risks identified. Any such document should be available for review.

The monitoring plan should include information regarding the identified risks and how the monitoring methods will address those risks. (See Q6 for further details.) The inclusion of these components in the monitoring plan will enhance the utility of the plan by providing a clear explanation of the identified risks and how they will be monitored, managed, and mitigated.

Q2. Should sponsors monitor only risks that are important and likely to occur?

A risk-based approach to monitoring should focus sponsor oversight activities on preventing or mitigating important and likely risks to investigation quality, including risks to human subject protection and data integrity. Sponsors also should consider monitoring risks that are less likely to occur but could have a significant impact on the investigation quality. Sponsors should determine the types and intensity of monitoring activities best suited to address the identified risks. In addition, monitoring plans should permit monitoring activities to evolve based on additional issues and risks that may be identified during the conduct of an investigation.

Q3. What factors should sponsors consider when determining the timing, types, frequency, and extent of monitoring activities?

As described in detail in the RBM guidance, factors sponsors should consider include the following:

- Complexity of the study design
- Types of study endpoints
- Clinical complexity of the study population (for example, study populations that are seriously ill, have multiple co-morbidities, or are more vulnerable and may require more intensive monitoring and consideration of on-site monitoring visits to be sure appropriate protection is being provided)
- Geographic location of clinical investigator (CI) sites where there may be differences in standards of medical practice or less established clinical trial infrastructure
- Relative experience of the CI and of the sponsor with the CI
- Electronic data capture to be utilized[7]
- Relative safety of the investigational product

[7] See the guidance for industry *Electronic Source Data in Clinical Investigations* (September 2013).

- Stage of the study (progress of the study)

- Quantity of data

FDA also recommends that sponsors consider the following additional factors:

- Workload at the CI site

- Turnover of personnel at the CI site or among monitoring staff

  – Similar to workload, high personnel turnover may cause unintended disruptions to conduct of the investigation and sponsor oversight.

- Location where subjects will be seen and whether they will be seen at more than one location to complete investigation procedures (for example, data collection at the imaging center, at a local physician's office, or at the subject's home)

  – When designing the monitoring plan, sponsors should take into consideration where and how the data are going to be collected in the investigation relative to where the sponsor oversight activities will be conducted (for example, to confirm that appropriate controls, instructions, and training tools are in place).

- Benefit of an early monitoring visit or other early monitoring activities

  – By scheduling an early monitoring visit (for example, soon after the first trial subject(s) enrolls in the investigation) or by carrying out other early monitoring activities (for example, through remote processes), sponsors can help ensure early in the investigation that procedures are being performed correctly at CI sites. Alternatively, if early monitoring identifies issues, corrective action(s) can be implemented sooner.

- Experience and qualifications of the research coordinator

  – The research coordinator serves an important role in ensuring the quality of the execution of the investigation at the investigation site (for example, the research coordinator often recruits subjects, collects and evaluates study data, and maintains study records.)

- Safety profile of the investigational product

  – When developing a monitoring plan, sponsors should consider the known safety profile, including the available human and non-clinical safety information for the product and the class, and the mechanism of action.

- Characteristics of data to be collected

  – When developing a monitoring plan, sponsors should consider the amount and complexity of the data collected.

Q4. How can a risk-based approach to monitoring that includes centralized monitoring help minimize missing data, protocol violations, or protocol deviations?

There may be situations in which poor trial conduct or adherence to the investigational plan causes or contributes to incomplete data collection. Therefore, by reviewing important investigation activities, in real-time across CI sites, sponsors may be able to identify the reasons for missing data, protocol violations, or protocol deviations and take corrective actions to minimize the likelihood of these occurring during the remainder of the clinical investigation.

Q5. Should the risk-based monitoring approach include processes to ensure that appropriate blinding is maintained?

Yes. As identified in the RBM guidance, for investigations in which blinding will be used for interventions and/or outcome assessments, ensuring that the investigation blind is maintained is a critical process that sponsors should consider in their risk assessment.

Specific risks to the maintenance of the blind that are identified during the risk assessment should be mitigated in advance of investigation initiation, when feasible. In addition, identifying and tracking deviations during investigation conduct that could result in unintentional unblinding of treatment assignment should be considered as a part of the monitoring plan to ensure that appropriate blinding is maintained at CI sites and by the sponsor. For example, in a blinded investigation that requires a site staff member to be unblinded to administer the test article, the site processes for maintaining the blind for the remainder of the site staff and the sponsor should be monitored.

FDA recognizes that Data Monitoring Committees (DMC) may access unblinded data as described in the DMC Charter. (For additional information about DMC, see the guidance for clinical trial sponsors *Establishment & Operation of Clinical Trial Data Monitoring Committees* (March 2006.)

B. Monitoring Plan Content

Q6. What elements should sponsors include in monitoring plans?

The following elements (discussed in detail in section IV.D of the RBM guidance) are summarized here to assist sponsors in developing monitoring plans:

- A synopsis of the study

- Study objectives

- Identification of critical data and study procedures
- Trial-specific risks to be addressed by monitoring
- Monitoring methods and rationale for use of the monitoring methods, including how the methods address identified risks
- Criteria for determining the timing, types, frequency, and extent of monitoring activities
- Specific activities necessary for each monitoring method used
- Definitions of events or results that would trigger changes in monitoring activities (for example, how protocol deviations may be monitored as events that would trigger changes in monitoring activities)
- Identification of protocol deviations and failures that, if occurred, would affect study integrity, and how they will be recorded, tracked, and reported
- Format, content, timing, and archiving requirements for documentation of monitoring activities
- Processes for communicating routine monitoring results to appropriate parties
- Processes for immediate reporting of significant monitoring issues to appropriate parties
- Processes for appropriate communication from study management and other stakeholders to monitors
- Processes to address unresolved or significant issues identified by monitoring
- Processes to ensure that root cause analyses are conducted where important deviations are discovered and that corrective and preventative actions are implemented
- Other quality management practices applicable to the clinical investigation (for example, reference to other documents describing appropriate actions regarding non-compliance)
- Training for personnel who carry out monitoring activities
- Planned audits of monitoring activities
- Process for updating monitoring plans

In addition, FDA recommends that monitoring plans also include the following items, which will help explain how the sponsor intends to address the risks that could affect the clinical investigation.

- A description of the investigation design and the blinding and randomization procedures, if applicable

- Processes for confirming that randomization is performed according to the protocol and investigational plan when randomization is identified as a risk to be addressed by monitoring

- Sampling plan(s) that will be used to identify the specific records and data that will be monitored, including the rationale for how the sampling plan provides a representative picture of the overall information, and how the sampling plan will be implemented

- A description of the types of significant issues identified through monitoring that would trigger immediate issue escalation

- An approach for determining if issues identified at a site also exist at other CI sites and an approach for correcting these issues

The monitoring plan should describe each of these items in sufficient detail. Sponsors also should reference related documents, when appropriate. Sponsors are encouraged to develop monitoring plans that emphasize critical risks that have the greatest potential to adversely affect investigation quality, including the rights, safety, and welfare of investigation subjects, and the collection or analysis of clinical data such as investigation safety and efficacy endpoints.

C. Follow-Up and Communication of Monitoring Results

Q7. How should sponsors follow up on significant issues identified through monitoring, including communication of such issues?

Significant issues identified through monitoring (for example, significant non-compliance with the protocol) should be thoroughly evaluated in a timely manner at the appropriate level (for example, sponsor, CI site(s)) as described in the monitoring plan. Appropriate corrective and preventative actions should be taken. Deviations from the investigational plan should be documented, tracked, and escalated to relevant personnel, as appropriate. Related systemic issues should be identified and resolved promptly to ensure that investigation quality, including the rights, safety, and welfare of investigation subjects and data integrity, is maintained.

Although not an exhaustive list, some examples of corrective and preventive actions that may be needed include retraining CI and site staff; clarifying protocol requirements through protocol amendment(s); or revision(s) to informed consent documents or procedures.

Significant issues identified through monitoring and the actions to be taken should be documented and communicated to the appropriate parties, which may include, but are not limited to, the following: (1) sponsor management, (2) sponsor teams, (3) CI sites, (4) institutional review board(s), (5) other relevant parties (for example, DMCs and relevant contract research organizations), and (6) FDA, when appropriate.

(See Q6 for a description of elements that sponsors should include in monitoring plans regarding follow up and communication of significant issues.)

Q8. How should centralized monitoring activities and the results of these activities be documented and shared with those involved in the investigation?

As described in the RBM guidance, documentation of monitoring activities should generally include the following: (1) the date of the activity; (2) the individual(s) conducting and participating in the activity; (3) a summary of the data or activities reviewed; (4) a description of any noncompliance, potential noncompliance, data irregularities, or other deficiencies identified; and (5) a description of any actions taken, to be taken, or recommended (see section V of the RBM guidance for details). Such documentation should include the results of centralized monitoring activities in sufficient detail to allow verification of adherence to the monitoring plan describing those activities.

Reports of centralized monitoring activities should be provided to appropriate management, including sponsor staff responsible for investigation and site oversight, in a timely manner for review and follow up. In addition, sponsors should inform a CI of monitoring findings from centralized monitoring activities that are relevant to the CI's activities.