Guidance for Industry

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

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.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be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance.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 comments to the Division of Dockets Management (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) Robert Temple, 301-796-2270, (CBER) Office of Communication, Outreach and Development, 301- 827-1800, or (CDRH) Robert L. Becker, Jr., 301-796-6211.

Guidance for Industry
Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products

Additional copies are available from:

Office of Communication
Division of Drug Information, WO51, Room 2201
Center for Drug Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave.
Silver Spring, MD 20993
Phone: 301-796-3400; Fax: 301-847-8714
druginfo@fda.hhs.gov
http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, suite 200N
Rockville, MD 20852-1448
(Tel) 800-835-4709 or 301-827-1800
ocod@fda.hhs.gov
http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm

or

Office of Communication, Education, and Radiation Programs
Division of Small Manufacturers, International, and Consumer Assistance, HFZ-220
Center for Devices and Radiological Health
Food and Drug Administration
1350 Piccard Drive, Rockville, MD 20850-4307
DSMICA E-mail: dsmica@cdrh.fda.gov; DSMICA Fax: 301-443-8818
(Tel) Manufacturers Assistance: 800-638-2041 or 301-443-6597
(Tel) International Staff: 301-827-3993
http://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 Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2012
Clinical Medical

TABLE OF CONTENTS

Guidance for Industry[1]
Enrichment Strategies for Clinical Trials to Support Approval of Human Drug and Biological Products

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I.   INTRODUCTION

This document provides guidance to industry on enrichment strategies that can be used in clinical trials intended to support effectiveness and safety claims in new drug applications (NDAs) and biologics license applications (BLAs). Similar strategies may also be useful in studies conducted in earlier stages of development.

FDA's guidance documents, including this guidance, 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 FDA guidances means that something is suggested or recommended, but not required.

II.   BACKGROUND

Clinical trials are not designed to demonstrate the effectiveness of a treatment in a random sample of the general population. Instead, sponsors use a variety of strategies to select a subset of the general population in which the effect of a drug, if there is one, can more readily be demonstrated. Some of these selection strategies are obvious (e.g., patients are enrolled only if they have the disease that the drug being studied is intended to treat), but there are many more ways in which patients are typically chosen to make detection of a treatment effect more likely. Examples include selecting patients whose disease does not spontaneously disappear or exhibit a large degree of variability, who are likely to comply with treatment, who are likely to have a high rate of disease progression, or who have some

---

[1] This guidance was developed by the Center for Drug Evaluation and Research in coordination with the Center for Biologics Evaluation and Research and the Center for Devices and Radiological Health at the Food and Drug Administration.

41 characteristic that suggests they can respond to the treatment. All of these selection strategies can be
42 described as *enrichment* of the study population.

43
44 For the purposes of this guidance, the term *enrichment* is defined as the prospective
45 use of any patient characteristic to select a study population in which detection of a
46 drug effect (if one is in fact present) is more likely than it would be in an unselected
47 population.
48
49

50
51 Among others, demographic, pathophysiologic, historical, genetic or proteomic, clinical, and
52 psychological characteristics have been used for enrichment. Enrichment may also refer to the
53 population to be analyzed within a broader population; that is, a study could include patients both with
54 and without the enrichment characteristic, but the primary analysis would be of the subset with the
55 characteristic, an approach that increases the study's ability to detect a drug effect, but that can also
56 provide some information about patients without the enrichment characteristic. Although this guidance
57 focuses on enrichment directed at improving the ability of a study to detect a drug's effectiveness,
58 similar strategies can be used in safety assessments.
59
60 Enrichment strategies fall into three broad categories:
61
62   1. *Strategies to decrease heterogeneity* − These include selecting patients with baseline
63   measurements in a narrow range (decreased inter-patient variability) and excluding patients
64   whose disease or symptoms improve spontaneously or whose measurements are highly
65   variable (decreased intra-patient variability). The decreased variability provided by these
66   strategies increases study power (see section III).

67   2. *Prognostic enrichment strategies* − choosing patients with a greater likelihood of having a
68   disease-related endpoint event (for event-driven studies) or a substantial worsening in
69   condition (for continuous measurement endpoints) (section IV). These strategies will increase
70   the absolute effect difference between groups but will not alter relative effect.

71   3. *Predictive enrichment strategies* − choosing patients more likely to respond to the drug
72   treatment than other patients with the condition being treated. Such selection can lead to a
73   larger effect size (both absolute and relative) and permit use of a smaller study population.
74   Selection of patients could be based on a specific aspect of a patient's physiology or a disease
75   characteristic that is related in some manner to the study drug's mechanism, or it could be
76   empiric (e.g., the patient has previously appeared to respond to a drug in the same class)
77   (section V).

78
79 This guidance describes and illustrates important enrichment strategies within these categories;
80 discusses study design options for different strategies, including advantages and disadvantages of the
81 various designs; and addresses issues of interpretation of the results of enrichment studies.
82
83 The enrichment strategies described in this guidance are discussed primarily in the context of
84 randomized controlled trials. In almost all cases, the strategies affect patient selection before
85 randomization (with a few exceptions for adaptive strategies to be noted later). These strategies,

86 therefore, generally do not compromise the statistical validity of the trials or the meaningfulness of the
87 conclusions reached with respect to the population actually studied.
88

89 The principal concerns with the use of enrichment strategies relate to the generalizability and
90 applicability of the study results. When considering use of an enrichment design, it is important to
91 consider whether the enrichment strategy can be used in practice to identify the patients to whom the
92 drug should be given and whether the drug might be useful in a broader population than will be
93 studied. The extent to which patients who do not meet the selection criteria for enrichment should be
94 studied (section VII.B.) is therefore a critical consideration. In addition, the accuracy of the
95 measurements used to identify the enrichment population and the sensitivity and specificity of the
96 enrichment criterion in distinguishing responders and non-responders are also critical issues.

97
98
99 **III.   DECREASING HETEROGENEITY**
100
101 Approaches to increasing study power − the ability of a clinical trial to demonstrate a treatment effect
102 if one is present − by decreasing heterogeneity (non-drug related variability) are widely practiced. The
103 following strategies are useful and generally accepted ways to decrease heterogeneity:
104

105 • Defining entry criteria carefully to ensure that entered patients actually have the disease that is
106   being studied and training investigators to adhere to protocol-specified entry definitions and
107   criteria.

108 • Identifying and selecting patients likely to comply with treatment to decrease variability in
109   drug exposure. *Note*: removing poor compliers identified after randomization is generally not
110   acceptable because such patients are not likely to be a random sample of the study population
111   and because compliance itself has been linked to outcome, even compliance in taking a
112   placebo.

113 • Using placebo-lead in periods prior to randomization to eliminate patients who improve
114   spontaneously or have large placebo responses.

115 • Decreasing intra-patient variability by enrolling only patients who give consistent baseline
116   values (e.g., for blood pressure measurements, treadmill exercise tests, pulmonary function
117   tests, or patient reported outcome (PRO) measures).

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

120 • Excluding patients unlikely to tolerate the drug.

121 • Excluding patients who are likely to drop out for non-medical reasons (e.g., because they have
122   difficulty getting to the study site).

123 Other strategies sometimes employed include eliminating patients in outcome trials with concomitant
124 illness likely to lead to early death or early drop-out and broad exclusions of patients on concomitant
125 therapies. There are concerns, however, that these strategies can result in studies that provide too little
126 information about the full range of people who will receive a drug in clinical practice, such as the
127 elderly, people with multiple illnesses, and those taking multiple drug therapies. It is not clear that

128   concomitant illnesses that do not affect survival or other endpoint measurements or concomitant drugs
129   unrelated to a test drug really do interfere with assessment of a treatment effect. Therefore, the
130   implications of using these strategies should be carefully considered before they are used. Two of
131   these strategies – encouraging compliance and reducing the numbers of spontaneous improvers or
132   placebo responders − warrant further discussion (see below).
133

134   **A.**   **Encouraging Compliance**
135

136   Practices such as encouraging good compliance by making patients aware of the conditions and
137   demands of the trial, avoiding too-rapid titration of drugs that could cause intolerable early side
138   effects, using adherence prompts and alert systems, and counting pills (or using "smart bottles" to
139   monitor drug use) so that non-compliant patients can be encouraged to perform better have become
140   standard. There have also, on occasion, been more specific efforts to identify and enroll good
141   compliers into clinical trials.
142

143   1.   The VA Cooperative hypertension studies of the late 1960s and early 1970s[2,3] gave prospective
144   patients placebo tablets containing riboflavin during a single-blind placebo period, then
145   examined patients' urine for fluorescence and randomized only patients whose urine fluoresced
146   (evidence that the patients had been taking the riboflavin tablets) on two consecutive visits
147   during a 2-4 month observation period.

148   2.   The Physicians' Health Study[4] used an 18-week, pre-randomization placebo run-in during which
149   patients (all physicians) self-reported compliance. About one third of the screened patients were
150   not randomized because of self-reported poor compliance. Compliance during the randomized
151   study was reported as a very satisfactory 90% over the 5 years of the study, greatly increasing its
152   power.[5]

153

154   **B.**   **Decreasing Placebo Responses and Spontaneous Improvement**
155

156   In placebo-controlled trials of drugs for symptomatic conditions (e.g., depression, anxiety, angina) or
157   laboratory abnormalities (e.g., dyslipidemia, hypertension), it is relatively common to have a single-
158   blind, placebo lead-in period and to randomize only patients whose signs or symptoms remain above
159   some entry level value. This is in part to identify and not randomize patients who would have had an
160   improvement for a reason other than a response to the test treatment (spontaneous improvement, a
161   placebo response, or effect of expectations or observations) that resolved the patient's symptoms or
162   signs, making the patient incapable of showing a response to treatment. Also, many signs and
163   symptoms vary spontaneously and, therefore, initial screening values that would support enrollment
164   may represent *random highs* of the disease course that will be followed by regression to the mean,
165   leaving the patient without the condition to be treated.
166

167

168   **IV.**   **PROGNOSTIC ENRICHMENT STRATEGIES—IDENTIFYING HIGH-RISK**
169   **PATIENTS**
170

171   A wide variety of prognostic indicators have been used to identify patients with a greater likelihood of
172   having the event (or a large change in a continuous measure) of interest in a trial. These indications
173   include clinical and laboratory measures, medical history, and genomic or proteomic measures.

174   Selecting such patients allows a treatment effect to be more readily discerned. For example, trials of
175   prevention strategies (reducing the rate of death or other serious event) in cardiovascular (CV) disease
176   are generally more successful if the patients enrolled have a high event rate, which will increase the
177   power of a study to detect any given level of risk reduction. Similarly, identification of patients at
178   high risk of a particular tumor, or at high risk of recurrence or metastatic disease can increase the
179   power of a study to detect an effect of a cancer treatment. Prognostic enrichment strategies are also
180   applicable, or potentially applicable, to the study of drugs intended to delay progression of a variety of
181   diseases, such as Alzheimer's disease, Parkinson's disease, rheumatoid arthritis, multiple sclerosis,
182   and other conditions, where patients with more rapid progression could be selected; it is possible, of
183   course, that such patients might be less responsive to treatment (i.e., that rapid progression would be a
184   negative predictor of response), and that would have to be considered.

185

186   For any given desired power in an event-based study, the appropriate sample size will depend on effect
187   size and the event rate in the placebo group. Prognostic enrichment does not increase the relative risk
188   reduction (e.g., percent of responders or percent improvement in a symptom), but will increase the
189   absolute effect size, generally allowing for a smaller sample size. For example, reduction of mortality
190   from 10% to 5% in a high-risk population is the same relative effect as a reduction from 1% to 0.5% in
191   a lower risk population, but a smaller sample size would be needed to show a 5% vs. 0.5% change in
192   absolute risk. It is common to choose patients at high risk for events for the initial outcome study of a
193   drug and, if successful, move on to larger studies in lower risk patients.

194

195   **A.**   **Experience with Prognostic Enrichment Strategies**

196

197   *1.*   *Cardiovascular Studies*

198

199   In CV disease, the severity of the illness being studied, as well as other factors that can indicate
200   increased risk, such as a history of recent myocardial infarction or stroke; the presence of concomitant
201   illness such as diabetes, hypertension, or hyperlipidemia; and certain blood markers, such as very high
202   LDL cholesterol, low HDL cholesterol and high C-reactive protein (CRP), have been used to identify
203   patients at greater risk for cardiovascular events, considerably reducing the sample sizes needed to
204   show an effect in outcome studies. Outcome studies using ACE inhibitors in congestive heart failure
205   (CHF) and HMG CoA reductase inhibitors in hyperlipidemia (the enalapril and statin trials,
206   respectively) illustrate this approach.

207

208   In the enalapril trials, mortality reduction and decreases in morbid events (such as hospitalization)
209   were first assessed in a very ill CHF population of NYHA Class IV patients (CONSENSUS),[6] then in
210   less ill patients (SOLVD treatment[7]) and eventually in asymptomatic patients (SOLVD prevention[8]).
211   In the later studies, composite endpoints were needed because the number of early deaths was too low
212   to allow an effect to be demonstrated, about 15% at one year on placebo in SOLVD treatment and
213   about 5% in SOLVD prevention, far lower than the 44% placebo 6 month mortality in CONSENSUS.
214   The very high early mortality in CONSENSUS, together with the large effect size (40% reduction),
215   allowed demonstration of a survival effect in just 253 patients, while the studies in less ill patients
216   required studies of 2,000-4,000 patients. The higher risk patients enrolled as a result of prognostic
217   enrichment showed, as expected, a larger absolute effect size, but relative effect size was also greater
218   in the more ill patients, suggesting that severity was also a predictive marker (see section V).

219

220   A similar strategy was used in the statin trials. The early CV outcome trials with statins were able to
221   evaluate the effects of the drugs on mortality because they enrolled patients with a history of heart
222   disease and very elevated cholesterol levels, patients whose mortality risk was substantial.[9,10] As the
223   benefit of statins became established in high-risk patient populations, subsequent CV outcome trials
224   had to enroll patients with less marked LDL cholesterol elevations and without known coronary artery
225   disease − populations that had not yet been shown to benefit from LDL cholesterol lowering and who
226   could still be ethically studied, but who were identified as high risk because of some other illness (e.g.,
227   type 2 diabetes mellitus) or risk factor (e.g., low HDL cholesterol, elevated high-sensitivity CRP). As
228   the population's risk became lower, sample sizes increased considerably but prognostic factors made
229   the studies possible. For example, in the recent JUPITER study[11] (n=17,802), a statin was shown to
230   have an effect on outcome in patients with *normal* LDL cholesterol, but who were at higher CV risk
231   based on factors other than LDL cholesterol, including age, one additional CV risk factor, and a high-
232   sensitivity CRP ≥ 2 mg/L. As the magnitude of risk declined in these study populations, it often also
233   became necessary to rely on composite endpoints, as the mortality rate was too low to allow a
234   mortality trial of reasonable size.
235
236   Choosing patients at relatively high risk of cardiovascular events can also be critical to the success of
237   safety studies intended to rule out a given level of cardiovascular risk. This approach is now
238   recommended for new antidiabetic treatments[12] and has been a consideration in the design of studies to
239   evaluate the cardiovascular risk of non-steroidal anti-inflammatory drugs.
240
241   *2.*   *Oncology*
242
243   It is clear that chemoprevention studies are most likely to be successful in people identified as high
244   risk (e.g., by genetic or other characteristics) for developing a cancer or recurrence of cancer. For
245   example, adjuvant therapy studies of tamoxifen showed that the drug not only delayed development of
246   metastases in patients with breast cancer, but also reduced the risk of contralateral tumors (new
247   primary tumors) in this high-risk group (high risk because they had already had breast cancer).
248   Tamoxifen was then studied in 13,000 high-risk people (calculated using the Gail model) without a
249   prior diagnosis of breast cancer who were followed for 4 years (NSABP P-1).[13] The study showed a
250   44% relative reduction in risk of invasive breast cancer, and tamoxifen was labeled for use in reducing
251   the risk of breast cancer in high-risk individuals identified using the Gail model calculator. A study in
252   people at lower risk would have required a substantially larger sample size. For example, a study of
253   people with a risk that was 25% of the risk of the NSABP P-1 study population would have needed
254   about 20,000 people to detect an effect of the size observed in the NSABP P-1 study with 90% power.
255
256   **B.**   **Potential Strategies for Prognostic Enrichment**
257
258   There may be additional approaches to identifying high-risk CV patients. A report in 2005[14] pointed
259   out that a higher resting heart rate, a small increase in exercise heart rate, and delayed recovery of
260   heart rate were all strong predictors of sudden death, suggesting a potential enrichment strategy in
261   studies of drugs to prevent sudden death. More recently, the potential for risk prediction based on
262   genetic factors has been examined,[15] as has the predictive value of coronary artery calcium score.[16,17]
263   The actual usefulness of these potential predictors as enrichment tools has yet to be established.
264

265 Published reports suggest additional strategies for identifying high-risk breast cancer patients,
266 especially for participation in studies of adjuvant treatment of cancer, where the likelihood of
267 recurrence and survival are critical to demonstrating an effect on these endpoints.
268
269   *1.   Prostate Cancer*
270
271 It has been reported[18] that in men with localized prostate cancer following radical prostatectomy, high
272 *PSA velocity* (PSA increase > 2 ng/mL during prior year) strongly predicted prostate cancer recurrence
273 and mortality over a 10-year period. Prognostic markers such as PSA velocity, if validated in future
274 studies, could be used to identify high-risk patients. Studies of adjuvant treatment for prostate cancer
275 would be better able to show an effect on survival if they enrolled patients with a high risk of death.
276
277   *2.   Breast Cancer*
278
279 Many investigators have reported gene expression profiles that appeared to predict the likelihood of
280 breast cancer recurrence after surgery. Selection of a population with a high rate of recurrence and
281 poor survival is critical if an adjuvant therapy study is to be successful. In a report on the use of five
282 different gene-expression profiling approaches in a non-randomized 285-patient sample treated with
283 local therapy, tamoxifen, tamoxifen plus chemotherapy, or chemotherapy alone, Fan, et al.,[19] found
284 that four of the five approaches for classifying patients had high concordance and a striking ability to
285 predict large differences in both recurrence and mortality, as illustrated in Figure 1, showing the
286 difference between patients with *good* and *poor* 70-gene profiles on relapse-free survival and overall
287 survival. These data need to be verified in actual trials, but gene expression profiling may provide a
288 means of identifying higher risk patients for adjuvant trials.
289
290   **Figure 1: Relapse Rates and Survival in Breast Cancer Patients Based on 70-Gene Profile**
291
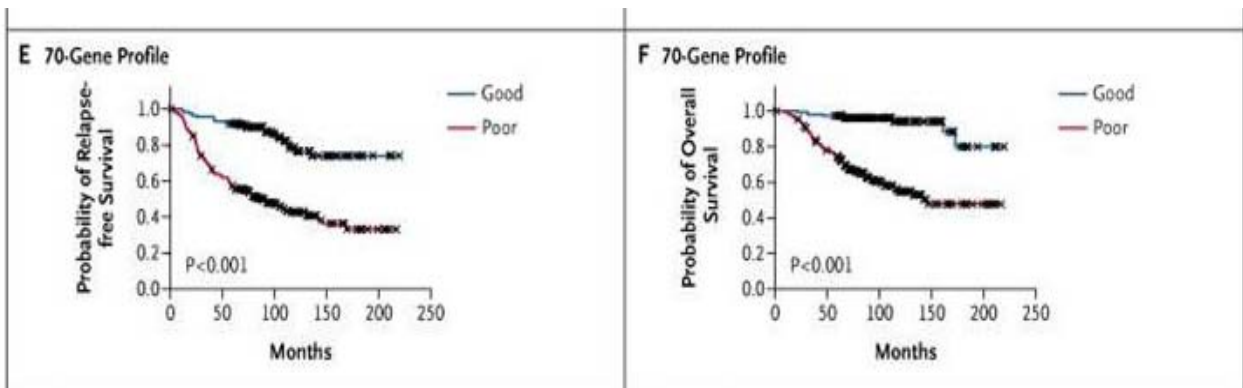
292
293
294
295 Recently, MammaPrint, an in vitro diagnostic test using the gene expression profile of fresh breast
296 cancer tissue samples to assess a patient's risk for distant metastasis, was cleared by FDA as a
297 prognostic test for certain breast cancer patients. As noted, use of such a diagnostic test represents a
298 potential enrichment strategy for adjuvant trials to identify a population at higher risk for recurrence.
299
300 Women with a deleterious BRCA 1 or 2 mutation have a lifetime incidence of breast cancer and
301 ovarian cancer of 60% and 15-40%, respectively, compared to a risk of 12% and 1.4%, respectively, in

302 women without a BRCA mutation.[20] Selecting women with such prognostic markers for a primary
303 prevention trial in breast cancer or ovarian cancer would increase the likelihood of cancer events,
304 thereby permitting a smaller sample size and a shorter study.
305
306
307 **V.   PREDICTIVE ENRICHMENT**
308
309 There are many possible ways to identify patients more likely to respond to a particular intervention,
310 and these have long been used in clinical trials when selection of patients has been based on a specific
311 aspect of pathophysiology, past history of response, or a disease characteristic that is related in some
312 manner to the study drug's mechanism (e.g., genomic or proteomic factor). For example:
313

314 • CHF can result from either systolic or diastolic dysfunction. Presumably, a population with
315   systolic dysfunction would be more likely to respond in a study of an inotrope.

316 • High-renin status predicts a greater anti-hypertensive response to beta-blockers, ACE
317   inhibitors, and angiotensin receptor blockers. A population with high renin hypertension
318   would be more responsive than a general hypertension population in studies of drugs in these
319   classes.

320 • For some indications, antibacterial drug effects are best analyzed in patients whose organism is
321   sensitive to the antibacterial drug. Most commonly, patients are randomized before sensitivity
322   is known, but only those patients whose organism is subsequently found to be sensitive to the
323   test antibacterial are analyzed for effectiveness.

324 • An initial screening for response −a biomarker measurement (e.g., radiographic response or
325   reduction of ventricular premature beats), early clinical response, or full-fledged clinical
326   response − in an open pre-randomization period can be used to identify a responder population
327   to be randomized into the controlled study. This approach is of particular value when
328   responders constitute only a small fraction of the overall population to be treated.

329 • A population of non-responders, or people intolerant to another drug, can be randomized to the
330   new drug or the original drug. The comparison is enriched because the population is expected
331   to have a poor response or a high rate of intolerance to the original drug compared to the test
332   drug. These designs cannot be used where effectiveness is critical or the intolerance is
333   dangerous.

334 • Proteomic markers, such as the HER 2/neu marker in breast cancer indicating potential for
335   response to trastuzumab, tumor EGFR markers, or genetic markers related to a drug's
336   mechanism of action can be used to identify potential responders, a rapidly growing
337   enrichment strategy in oncology.

338 Identifying a responder population (i.e., a subset of the overall population with a larger than average
339 response to treatment) and studying this population in a clinical trial can provide two major
340 advantages: increased study efficiency or feasibility and an enhanced benefit-risk relationship for the
341 people in the subset compared to the overall population.
342
343 *Increased Efficiency or Feasibility*
344
345 Identification of a high treatment response population greatly increases the chance that a study of an
346 effective drug will be able to detect a treatment effect and allows a study to succeed with a smaller
347 sample size than a study in an unselected population. The strategy can be particularly useful for early
348 effectiveness studies because it can provide clinical *proof of concept* and contribute to selection of
349 appropriate doses for later studies. When the treatment responder population constitutes only a small
350 fraction of all patients, say 20% (a common situation in oncology settings), enrichment can permit a
351 showing of effectiveness when a study in an overall population may have difficulty showing any
352 effect. Unlike prognostic enrichment, which leads to a larger absolute effect but no change in relative
353 effect, a predictive enrichment population will show both a larger absolute effect and a larger relative
354 effect than the general, unselected population.
355
356 The extent to which the sample size needed to adequately power a study can be reduced using a
357 predictive enrichment strategy is a function of the prevalence of the enrichment marker and the
358 relative effectiveness of the drug in the marker-positive and marker-negative populations. Table 2
359 illustrates how sample size ratios − the ratio of the number of subjects needed in an unselected
360 population versus the number needed if only the marker positive population is studied − change with
361 varying prevalence of marker-positive patients and different magnitudes of treatment effect in marker-
362 negative patients (treatment effect in marker-negative patients of either 0% or 50% of the effect in
363 marker-positives). Table 1 assumes the classification of patients into positive versus negative is 100%
364 accurate.
365
366 **Table 1: Sample Size Ratios as a Function of the Prevalence of Marker-Positive Patients**
367

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

368
369 In general, the lower the prevalence of marker-positive patients and the smaller the relative effect size
370 in the marker-negative population, the more the sample size can be reduced in a study of marker-
371 positive patients compared to a study in an unselected population. For example, when the prevalence
372 of marker-positives in a population is only 25% and no treatment effect is expected in the 75% of

373 patients who are marker-negative, the required sample size in a study of an unselected population
374 would be 16 times the sample size needed for a study that included only marker positive patients. A
375 more detailed description of these results and the conditions under which the results were obtained has
376 been presented by Simon and Maitournam.[21]
377
378 *Enhanced Benefit–Risk Relationship*
379
380 Identification of a responder population can enhance the benefit–risk relationship of a drug by
381 avoiding exposure and potential toxicity in people who cannot benefit from the drug. For drugs with
382 significant toxicity and a low overall response in a general population − factors that could deter further
383 development − identifying a responder population could make a risk more acceptable and facilitate
384 continued development and approval. For example, the significant survival advantage (approximately
385 5 months) seen with trastuzumab in the treatment of metastatic breast cancer in patients with high
386 HER 2/neu expressing tumors (about 25% of breast cancers) ultimately supported use of the drug in
387 the marker-selected population despite the significant cardiotoxicity that emerged (see further
388 discussion in section V.B.3). The much smaller mean effect (less than 2 months) that would have
389 been observed in an unselected population and the fact that only about one-fourth of patients would
390 have benefited might have made approval difficult to support in the face of the observed cardiotoxicity
391 of the drug.
392
393 Identifying a more responsive population does not necessarily imply that there is no benefit in the
394 remaining population. It is therefore generally desirable to have some data in the non-selected (non-
395 enrichment) population to determine whether they respond less well, or indeed do not respond at all.
396 These data also can provide an assessment of safety in the non-selected population in the event that
397 such patients are exposed postapproval. The data need not be obtained in the controlled trials
398 supporting effectiveness but could be obtained in earlier studies. A qualitative estimate of
399 effectiveness might also be based on pharmacologic or even pre-clinical data. A strong mechanistic
400 rationale can make study of the non-enriched population unnecessary (e.g., study of effects in an
401 infection caused by a resistant organism).
402
403 A trial intended to provide evidence of effectiveness to support approval could include a broad range
404 of patients, but be prospectively designed to have as its primary endpoint the effect in the enriched
405 population subset. This is a standard (and unavoidable) approach when the baseline characteristic can
406 only be determined after randomization (e.g., the infectious organism or tumor characteristic), but the
407 approach (preferably with stratified randomization) is also valuable in other settings to gather some
408 information on the marker-negative population.
409
410 The following discussion considers predictive enrichment strategies in 5 categories:
411
412 • empiric strategies
413 • pathophysiologic strategies
414 • genomic strategies
415 • randomized withdrawal studies
416 • studies in non-responders or patients intolerant to other therapy
417
418

10

419   **A.   Empiric Strategies**
420
421   With an empiric strategy, the selection of likely responders for a study is not based on any
422   understanding of the basis for differences in response between patients, but on observations of a
423   response during screening periods, or prior experience with the drug or related drugs.
424
425   *1.   Open Trial Followed by Randomization*
426
427   A straightforward enrichment strategy, in cases when a treatment response can be identified shortly
428   after treatment initiation, is to give open-label drug to all patients, identify apparent responders, either
429   on the planned study endpoint or on a biomarker or other short-term response thought to be predictive
430   of clinical response; withdraw the treatment; and then randomize only responders into a placebo-
431   controlled trial. This strategy is particularly useful when there is a low response rate. This strategy
432   has been discussed in the past[22] and was used throughout the 1970s[23] to develop new anti-arrhythmic
433   drugs. Patients were titrated on the investigational agent until they had an acceptable reduction of
434   ventricular premature beats (VPBs). Only the responding patients were then randomized into placebo-
435   controlled trials, often fixed-dose, dose-response studies.
436
437   In the mid-1980s, controlled studies to evaluate topical organic nitrate patches randomized only people
438   who were shown immediately before the study to have an exercise angina response, or a blood
439   pressure response, to sublingual nitroglycerin. See also section 2 below (History of Response to a
440   Treatment Class).
441
442   The Cardiac Arrhythmia Suppression Trial (CAST) − a study of the mortality effect of suppressing
443   VPBs in patients with a recent acute myocardial infarction (AMI) and at least 6 VPBs per hour − is
444   one of the best known studies conducted in apparent responders. [24] It was known that people with > 10
445   VPBs per hour after an AMI had a 4-fold increase in the rate of sudden death. Previous failed
446   attempts to show survival benefits with antiarrhythmics had been criticized for the low rate of VPB
447   suppression achieved because many included patients did not respond. The CAST used an open-label
448   screening period to identify responders to two drugs, encainide and flecainide, shown to be very
449   effective in suppressing VPBs in a previous study.[25] Only patients who had at least a 70% VPB
450   reduction were randomized. Unfortunately, despite the enrichment effort, these anti-arrhythmic agents
451   did not decrease mortality but instead more than doubled it. This result reflects the inadequacy of the
452   surrogate endpoint of VPB reduction as a predictor of an effect on mortality, not the study design. The
453   enrichment design did, however, yield a study capable of showing an effect of VPB suppression and
454   allowed clear interpretation of the study, which showed, contrary to expectations, that even in VPB
455   responders, the drugs were not helpful and, indeed, were harmful.
456
457   *Note:* Use of an initial open-label phase without a control group does raise some concerns that need to
458   be considered when such a design is used. For example, in the CAST study there were deaths during
459   the screening period (not surprising given the recent infarction) that were difficult to interpret in an
460   open, uncontrolled setting where all patients received active drug. In CAST II[26] (ethmosin vs.
461   placebo), the initial screen for VPB suppression used a randomized comparison of drug to placebo,
462   with the responders then randomized into the placebo-controlled trial. This strategy showed that the
463   drug used in the screen was itself lethal (19 deaths on ethmosin vs. one on placebo) and the study was
464   stopped early.

465
466   A similar problem was described in outcome trials of carvedilol in the treatment of chronic CHF.[27]
467   These studies, unlike CAST, clearly showed a benefit of treatment. In two large studies, some patients
468   were excluded during a run-in period because they could not tolerate carvedilol. Some of those
469   patients died. The drop-out rates during the subsequent controlled trials were undoubtedly decreased
470   by the screening procedure that excluded patients intolerant to beta-blockers, and the results made
471   carvedilol seem better tolerated than it would actually be in patients starting therapy. The randomized
472   comparisons and the benefits demonstrated are fully valid in these trials for the populations studied,
473   but the benefits and risks facing unselected patients, who would be treated in clinical use of the drug,
474   may be different from those observed in the clinical trial, requiring close attention to the screening
475   period results.
476
477   There are many other outcome study settings in which it would be possible to select people more
478   likely to benefit from treatment. Patients with a lipid abnormality might be given the planned
479   treatment in a screening period to evaluate their biochemical response. For the randomized trial, only
480   people with a response of a certain size might be randomized, giving a greater mean effect on the lipid
481   level and, presumably, a larger effect on outcome. That approach could be useful in an early outcome
482   trial, but it would also be possible to randomize a broader population stratified by such an initial
483   response with the intent of making the primary study endpoint the result in the high-response subgroup
484   while gaining some information about the less responsive group. Again, the response in such a
485   selected group would not describe the response in an unselected population, an important issue for
486   labeling.
487
488   Active, open screening for empiric responders is particularly advantageous when a population is made up
489   up of subsets (not identifiable pathophysiologically or genetically prior to treatment) with potentially
490   very different responses to interventions. It is hard to know in advance when this is true, but certain
491   difficult-to-study conditions, such as irritable bowel syndrome or fibromyalgia, might be candidates
492   for this approach.[22]
493
494   The overall strategy (open trial followed by randomization) is a very efficient way to document
495   effectiveness, but it cannot be used prospectively to identify the responder population when the drug is
496   used in clinical practice. In some cases, however, an early response could be used to determine who
497   should stay on the drug, essentially how all symptomatic treatments are used.
498
499   *2.   An Individual's History of Response to a Treatment Class*
500
501   Information about prior response to a drug in a pharmacologic class, if available, can be used to
502   identify potential responders for a study of a new member of that class. As is the case with an open
503   label trial followed by randomization, use of patient history of response to a drug class can greatly
504   increase the efficiency of a trial in demonstrating effectiveness. In most cases, however, it will not
505   help identify the population to be treated in clinical practice.
506
507   A study enriched with prior responders to a pharmacologic class can be useful in demonstrating
508   effectiveness at the proof-of-concept stage. This design may be particularly advantageous for
509   randomized, fixed-dose, dose-response studies (the preferred dose-response study design described in
510   the ICH guidance *E4, Dose-Response Information to Support Drug Registration*[28]). A responder

511 population provides a larger overall treatment effect and, therefore, a steeper dose-response curve,
512 which generally allows for easier interpretation of the curve (identifying the steep area and plateau of
513 the curve) and more precise characterization of dose-response, especially for doses providing near-
514 maximum effects. For example, a dose-response study of indapamide in known responders to
515 diuretics demonstrated mean decreases of 29/12 mmHg (systolic/diastolic) for the 2.5 mg dose and
516 37/15 for the 5 mg dose, an increase in effect with the 5 mg dose considerably larger than that seen in
517 studies of unselected patients, where 2.5 mg and 5 mg gave similar results.
518
519   *3.   Factors Identified in Results from Previous Studies*
520
521 Analyses of results of previous trials can sometimes point to a substantially greater effect in a
522 specific subset of the overall population and provide a basis for studying that subset in a
523 subsequent study, either as the sole population studied or as the identified primary endpoint subset
524 in a study of a broader population. For example, BiDil, a treatment for severe CHF, was
525 approved[29] on the basis of a placebo-controlled study carried out entirely in self-identified blacks
526 (the African-American Heart Failure Trial (A-HeFT)).[30] The selection of a black population was
527 based on two previous studies (the Vasodilator-Heart Failure Trials (V-HeFT) I and II[31,32]) of a
528 hydralazine hydrochloride – isosorbide dinitrate combination vs. placebo in a racially mixed
529 population that strongly suggested effectiveness in blacks. In those studies, the combination had
530 not shown an overall benefit, but post-hoc analyses revealed a nominally significant effect in black
531 patients in V-HeFT I and apparent equivalence to enalapril in V-HeFT II. In contrast, there was
532 little or no effect of the combination in whites in V-HeFT I and nominally significant inferiority of
533 the combination to enalapril in whites in V-HeFT II. The replication of the observed effect in
534 blacks was strong, with only a suggestion of a modest effect in whites, perhaps a third of the effect
535 in blacks. A trial to establish this small effect in a white population would have required 16,000
536 patients. The product was approved for "self-identified blacks" only.
537
538 **B.   Pathophysiological Strategies**
539
540 These strategies involve selection of likely responders based on the patient's individual physiology or
541 on assessment of disease pathophysiology that suggests that only certain patient subgroups will
542 respond to a particular therapy or that certain subgroups will respond better than others.
543
544   *1.   Metabolism of the Test Drug*
545
546 For a drug that acts through an active metabolite, as is the case for the antiplatelet drug clopidogrel,
547 patients may differ in their ability to metabolize the prodrug to its active metabolite. Some patients
548 may not form the active metabolite at all and others may not make enough to respond to the dose
549 selected. Including these patients in a trial will dilute the overall drug effect and can also lead to
550 inefficient or inappropriate use of the drug in practice if the two subsets of patients are not identified
551 and treated differently. In some cases, it will be possible to adjust (increase) the dose in the poor
552 metabolizers, but patients who cannot make the active metabolite at all should probably be excluded
553 from the trial or from the planned primary analysis. A closely related approach is the assessment of
554 uptake of the test drug by a tumor.[33] Historically, before treatment of thyroid tumors with I-131, a low
555 dose was given to determine whether the tumor did, in fact, take up iodine and to what extent, so that
556 the needed dose could be estimated.

557
558   2.   *Effect on Tumor Metabolism*
559
560 It may be possible to select patients for a cancer trial by screening for an effect on a tumor metabolic
561 response, as assessed by a positron emission tomography (PET) scan. For example, response to the
562 tyrosine kinase inhibitors imatinib and sunitinib in patients with gastrointestinal stromal tumors
563 (GIST) has been shown to correlate well with metabolic responses (decreased tumor glucose
564 utilization) assessed by 15F-fluorodeoxyglucose (FDG) PET imaging.[34] The clinical trial could enter
565 only the identified metabolic responders or enter all patients, stratified by metabolic response. In the
566 second case, the primary hypothesis could be the treatment effect in the metabolic responder stratum.
567
568   3.   *Proteomic Markers and Genetic Markers Linked to a Proteomic Marker*
569
570 Increasingly, cancer treatments are directed at enzymatic, hormonal, or other functions that are tied to
571 tumor surface intracellular receptors. The following examples illustrate use of proteomic markers, or
572 genetic markers that are linked to a proteomic marker, that are known to be essential for the activity of
573 the drug.
574
575   • Trastuzumab was developed to bind to the Her-2-neu receptor, which is present on normal
576   and malignant cells but is over-expressed in about 25% of breast cancers. Binding of
577   trastuzumab to the Her-2 neu receptor blocks receptor-mediated growth-stimulating
578   intracellular signaling, decreasing cellular repair after chemotherapy and radiation therapy
579   and also increasing apoptosis. In activity-estimating trials, anti-tumor activity in patients
580   with lower levels of Her-2-neu receptor expression (1+ by immunohistochemical staining)
581   was minimal, so that definitive efficacy trials in patients with metastatic disease were
582   limited to patients with Her-2-neu 2+ or 3+ over-expression. In the treatment of metastatic
583   disease, when added to either of two background regimens, trastuzumab increased survival
584   by a mean of about 5 months, about 3 to 4 times the effect that would have been expected
585   in an unselected population, assuming no response (which a modest amount of testing
586   showed was the case) in the Her-2-neu negative patients. Enrichment thus allowed a
587   modest-sized study to show a striking effect and directed treatment to the population that
588   could benefit. In addition, because the drug was shown to be moderately cardiotoxic in the
589   metastatic breast cancer trials, in designing adjuvant studies it was considered critical to
590   focus on potential responders (i.e., patients with Her-2-neu receptor over-expression).[36]
591
592   • Imatinib was developed to treat patients with gastrointestinal stromal tumor (GIST), a
593   tumor not previously responsive to antineoplastic therapy. Imatinib inhibits c-Kit, a
594   receptor tyrosine kinase that is mutated and activated in a large majority of GIST patients,
595   resulting in abnormal proliferation of tumor cells. In a small study (N=147) in patients
596   with a pathologic diagnosis of c-Kit-positive unresectable and/or metastatic malignant
597   GIST, 56 patients responded and 55 of the responding patients demonstrated a durable
598   partial response of 7-38 weeks (median 13 weeks). This study was followed by a placebo-
599   controlled, adjuvant therapy trial using imatinib in GIST patients with c-Kit expression in
600   whom complete gross resection of GIST had been performed. The study showed a
601   substantial increase in recurrence-free survival at a median follow up of 14 months.[37]
602

- 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 (ER) and progesterone receptors (PR) have a response rate of approximately 70%; patients whose tumors express either ER or PR, but not both, had a 40% 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[38] and would be expected in any trials of endocrine therapy.

- 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 an initial study in 49 patients with melanoma, 11 of 16 patients with $BRAF^{V600E}$ 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 in 675 patients with metastatic or unresectable melanoma who had the $BRAF^{V600E}$ mutation compared vemurafenib to dacarbazine. 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.[39]

The examples of pathophysiologic selection just described reflected, at least initially, tumor receptor variables that could be described as proteomic variables, but that were in many cases later identified as tumor genetic markers (EGFR and BRAF genetics, for example). In such cases the genetic marker defines a pathophysiologic effect.

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 non-inferiority of the test treatment.[40] It is also important to gain as much information as possible about the marker-response relationship (sensitivity and specificity).

C.   Genomic Strategies

To date, most genomic enrichment strategies have involved tumor genomics. 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 responders without providing a pathophysiologic basis for the difference in response.

Any genetic differences that predict response must in the end have some pathophysiologic basis, but enrichment strategies to identify responsive patients could be used before recognition of a mechanism. Studies directed at tumors with specific genomic patterns that appear to predict outcome (e.g., tumors with mutations in a target gene, genome-wide expression profiles, SNP arrays) are becoming increasingly common. One difficulty is that relationships between genetic patterns and outcomes are often found only after study results are known. Markers discovered this way will have credibility problems related to the post-facto nature of the finding and will almost always need confirmation in a

649 prospectively planned enriched study. Such findings are of particular concern when overall study
650 results show no effect. As a general matter, preservation of study specimens and analysis of results by
651 various genomic markers are increasingly prevalent practices and are of great potential value, whether
652 they lead to an enriched study following a failed trial, or to better targeted therapy after a successful
653 trial.
654
655 Freidlin and Simon,[41] however, have proposed a novel, prospective approach that makes use of
656 genomic data collected on all randomized subjects during the conduct of a study:
657

658 1. Design the study as usual, but divide it into first and second halves. Prospectively allocate the
659   overall study alpha as 0.04 for the whole population and 0.01 for a patient subset to be
660   identified in the first half of the study.

661 2. Run the first half of the study and conduct unblinded data analyses, searching for a genetic
662   predictor of response. There would be no limit to the number of such analyses conducted. A
663   single genetic subset appearing to predict response may be identified.

664 3. Complete the remainder of the study, entering patients according to the original eligibility
665   criteria (both the predicted responders and predicted non-responders) as before.

666 4. At the conclusion of the study, the effect in the entire study population is tested at an alpha of
667   0.04, and the genetically identified subset is tested only in the second half at an alpha of 0.01.

668 5. The study shows evidence of effectiveness if either analysis is positive. When the responder
669   population is a small fraction of the total population, but exhibits a large response, this design
670   can improve the chance of detecting a treatment effect. It also retains good power for the
671   overall study if the drug is more broadly effective.

672 A similar approach that avoids delay while the genomic marker is being evaluated might also be
673 considered[42]:
674

675 1. Run the planned study, again prospectively dividing the available alpha between the overall
676   study population and a subset to be identified later.
677

678 2. Whether the overall population analysis is positive or negative, take a random sample of the
679   study population (50%, 33% or another fraction) and search for a subset with a genomic
680   pattern (or, for that matter, any other subset) showing a substantial differential effect. This first
681   analysis would be considered wholly exploratory. It would be important to have genomic
682   assessments on a substantial portion of all subjects.[43]
683

684 3. Examine the remaining data, considering only the genomic subset found positive in the first
685   sample. As long as the second part was kept blind until after the first analysis and only the
686   marker-positive subset is tested, study-wide alpha error will be preserved. Any such approach
687   would need scrupulous attention to maintaining the blind, perhaps by using an independent
688   group to do the genomic search.
689

690 For anti-viral drugs it can be anticipated that there will be cases in which either host or viral genomics
691 plays a role in determining drug response. For example, genome-wide association studies have
692 identified host IL28B gene variants associated with the likelihood of a viral response to

693   ribavirin/interferon regimens to treat Hepatitis C virus (HCV). This association has been
694   independently substantiated in multiple studies and can distinguish patients with a low (about 25%) or
695   high (about 80%) likelihood of sustained viral response,[46] to a pegylated interferon plus ribavirin.
696
697   It is also well documented that HCV viral genotype determines the needed duration of therapy and
698   predicts likelihood of response. Six genotypes have been identified. Genotype 1 is the most prevalent
699   genotype in the United States and is more resistant to treatment than genotypes 2 and 3. Patients with
700   genotype 1 treated with a pegylated interferon and ribavirin for 48 weeks had a sustained virologic
701   response of 40 to 50%. Patients with genotypes 2 and 3 had sustained responses of 80% or more with
702   only 24 weeks of treatment with pegylated interferon and ribavirin.[57] Two protease inhibitors, direct-
703   acting antiviral agents boceprevir and telaprevir, have been shown to provide greatly improved
704   sustained viral response and the potential for reduced duration of total treatment in patients with
705   genotype 1 virus, when added to ribavirin/interferon.
706
707   **D.   Randomized Withdrawal Studies**
708
709   In a randomized withdrawal study, patients who have an apparent response to treatment in an open-
710   label period or in the treatment arm of a randomized trial are randomized to continued drug treatment
711   or placebo. As such trials generally involve only patients who appear to have responded, this is a
712   study enriched with apparent responders, an empiric strategy. The study evaluation can be based on
713   signs or symptoms during a specified interval (e.g., blood pressure or angina rate), on recurrence of a
714   condition that had been absent (e.g., depression), or on the fraction of patients developing a rate or
715   severity of symptoms that exceeds some specified limit (a failure criterion).
716
717   The randomized withdrawal design was proposed as a way to establish long-term effectiveness of
718   drugs in settings in which long-term use of a placebo would not be acceptable (e.g., most psychiatric
719   and antihypertensive drug treatments).[46] Even in settings in which long-term placebo use may be
720   acceptable, however, it is generally difficult to recruit patients, and drop-out rates are often high,
721   posing difficult analytic problems. A randomized withdrawal design in which the study population is
722   on treatment for an extended duration followed by blinded, randomized withdrawal of treatment for a
723   short duration could provide evidence of prolonged effectiveness with only brief exposure to placebo.
724   The design allows a patient to be removed from the study (for having reached an endpoint) when the
725   condition returns at specified severity, avoiding long-term exposure to an ineffective treatment.[48]
726
727   The randomized withdrawal design can also be used as an initial trial to show effectiveness when there
728   is an existing population of patients in an open-label treatment setting (e.g., under an IND or as an off-
729   label use of an approved drug), as illustrated by the cases of nifedipine and gamma-hydroxybutyrate.
730
731   The approval of nifedipine for vasospastic angina (the first drug approved for this condition) illustrates
732   the utility of this design. An open-label, historically controlled trial[47] was considered inadequate to
733   support approval because the natural history of vasospastic angina was not well-established.   A
734   randomized withdrawal design (see Figure 2) was conducted in patients already receiving the drug,
735   with a primary endpoint of recurrence of severe vasospastic angina leading to study withdrawal. A
736   total of 28 patients participated in the study. One-third of the patients randomized to placebo
737   withdrew early, as compared to no withdrawals in patients randomized to nifedipine (see Table 2).
738

739   **Figure 2: Nifedipine Randomized Withdrawal Trial in Vasospastic Angina**
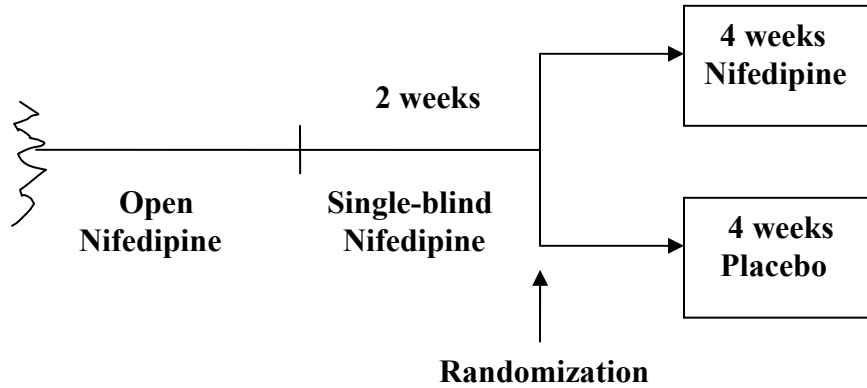
740
741
742   **Table 2: Results of Nifedipine Randomized Withdrawal Study**
743

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

744   * Statistically significant at $p \leq 0.05$

745
746
747   Another example in which patients already using a drug were studied was gamma-hydroxybutyrate
748   (GBH, sodium oxybate), which was approved for treatment of cataplexy on the basis of a single
749   placebo-controlled study of conventional design and a second, small, randomized withdrawal study in
750   55 long-term (7 to 44 months) users randomized to 2 weeks of continued treatment with GBH or
751   placebo. The second study produced a clinically and statistically impressive result, as shown in Table
752   3, and needed little time for recruitment.
753
754
755   **Table 3: Randomized Withdrawal Study of GHB in Cataplexy**
756

	Median Attacks/2 weeks	
Treatment Group	Baseline	Change in Rate
Placebo (n=29)	4.0	+21.0
GHB (n=26)	1.9	0
	P < 0.001	

757
758
759   By randomizing patients to different doses, the randomized withdrawal design can also be used to
760   obtain long-term dose-response data. For example, this design was used to demonstrate effectiveness
761   of a single weekly dose of fluoxetine in preventing recurrence of depression; patients on 20 mg/day
762   were randomized to placebo, fluoxetine 70 mg/week as a single dose, or continued 20 mg per day.
763   Both fluoxetine groups were superior to placebo in reducing the rate of recurrence.

764
765 A randomized withdrawal design may also be useful in establishing the duration of benefit of long-
766 term treatments, an issue of great interest for drugs with long-term toxicity concerns and a potential for
767 long, post-treatment benefit. For example, the original node negative adjuvant tamoxifen breast
768 cancer trial (NSABP B-14)[48] randomized 2,892 women with node-negative, hormone receptor-positive
769 breast cancer to receive 5 years of tamoxifen versus placebo. Tamoxifen reduced the risk of
770 recurrence by 41% and the risk of death by 33% and was rapidly adopted as the standard of care. A
771 second randomization was then performed in which the 1,166 patients who had received tamoxifen
772 were assigned to receive either an additional 5 years of tamoxifen or placebo. The trial was halted by
773 the data monitoring committee after an interim analysis revealed that continuing the study could not
774 demonstrate an advantage for continued tamoxifen use and suggested that tamoxifen was harmful.
775 There was a non-statistically significant trend in both disease free survival (DFS) and overall survival
776 (OS) favoring the placebo group and an increase in new malignancies in the tamoxifen group. Similar
777 designs have been used to examine long-term effects of bisphosphonates.
778
779 **E.   Studies in Non-responders or Patients Intolerant to Other Therapy**
780
781 A study can be enriched by selection of patients who failed to respond to an existing drug, or who
782 were intolerant of that drug. Although non-responders or treatment intolerants are not more likely
783 than an unselected population to respond to or tolerate the new drug, they would generally be less
784 likely to respond to or tolerate the existing drug, giving the test drug an enrichment advantage.
785 Because patients in a trial sometimes respond to a drug to which they had previously failed to respond,
786 in most cases studies in non-responders are informative for the between-drug comparison of
787 effectiveness only if patients are randomized to both the new and failed drug (i.e., not simply placed
788 onto the new drug in a single arm study or randomized to new drug versus placebo). This approach
789 can also provide important information to practitioners; it is critical to know whether another member
790 of a pharmacologic class or a member of a different class can be useful in patients who fail on a
791 previous treatment. The approach may be useful in two settings:
792
793 • To demonstrate the effect in non-responders to previous therapy of a drug that may not be
794 more effective overall than existing therapy in an unselected population, but that has a different
795 responder population. A drug that acts through a mechanism different from that of existing
796 treatments might be effective in non-responders to the existing drug.
797
798 • To efficiently demonstrate the treatment effect of a new drug that is actually moderately
799 superior to the existing drug, but where a very large study would be needed to show superiority
800 if the study included unselected patients, many of whom would respond to the less effective
801 drug. For example, if the new drug response rate is 90% and the existing drug response rate is
802 80%, a study with 90% power to detect that 10% difference would require about 600 patients.
803 In contrast, if only non-responders to the existing drug were randomized (20% of the patients
804 treated with the existing drug), few would respond to the existing drug, and at least half of the
805 patients would respond to the new drug, a difference detectable with fewer than 40 of the non-
806 responders.
807
808 *Note*: In neither case would showing an advantage for the new drug in non-responders to previous
809 therapy establish overall superiority of the new drug.

810
811   *1.   Studies in Non-Responders*

812
813   a.   Captopril

814
815   To support a claim in severe hypertension unresponsive to other agents, a study was designed
816   to evaluate patients who had not responded to standard triple therapy (propranolol 320 mg,
817   hydrochlorothiazide 100 mg, and hydralazine 200 mg) (there were many escape pathways for
818   non-responding patients so that they would not be untreated). Patients who had failed triple
819   therapy and had diastolic pressure that was severely elevated were observed for 1 to 2 weeks
820   on the same regimen (triple therapy lead-in) and, if their diastolic pressure did not exceed a
821   defined limit, randomized to the same standard triple therapy they had failed on, or to captopril,
822   with a 2:1 captopril to triple therapy randomization ratio. The number of responders (diastolic
823   pressure less than 90 mmHg or fall of at least 10 mmHg, but not to below 90 mmHg diastolic)
824   clearly favored captopril in this difficult-to-treat population (Table 4).

825
826   **Table 4: Results of the Captopril Severe Hypertension Trial for the Group**
827   **Randomized to Captopril or Triple Therapy**

828

		Captopril	**Triple Therapy**
Total number analyzed		66	30
Week 4	Normalized DBP ≤ 90	21 (33%)	5 (17%)
	Reduction in DBP ≥ 10	8 (13%)	3 (10%)
Week 8	Normalized (DBP ≤ 90 mmHg)	22 (34%)	4 (14%)
	Reduction in DBP > 10 mmHg	14 (22%)	3 (10%)

829
830   *Note*: approximately 25% of the triple therapy *non-responders* did respond to the previously
831   failed therapy in the new trial. This finding reinforces the need for randomization to the new
832   and reportedly failed therapy in a study in non-responders. A study that had merely switched
833   patients from the failed therapy to the new one and found 25% responders might have been
834   interpreted as showing an effect of the new drug in the non-responders to prior therapy when,
835   in fact, it would not have demonstrated that.

836
837   b.   Clozapine

838
839   Clozapine is an antipsychotic agent associated with serious toxicity, a greater than 1% rate of
840   potentially fatal agranulocytosis. For clozapine to be approved, it was crucial to show that it
841   offered a clear advantage over safer alternatives. To show this, a study was conducted in
842   hospitalized schizophrenic patients with a history of poor response to neuroleptics who, in
843   addition, had failed to respond to 6 weeks of treatment with haloperidol. These patients were
844   randomized to 4 weeks of treatment with clozapine or chlorpromazine plus benztropine. The
845   results showed a striking advantage for clozapine on Clinical Global Impression (CGI) and
846   British Psychological Rating Scale (BPRS) standard measures in antipsychotic drug trials (see
847   Table 5). Despite its serious risk, clozapine was approved for use in patients not responding to
848   other anti-psychotic agents.

849

850
851
852

Table 5: Results of Clozapine Study in Non-responders to Standard Psychotropic Agents

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

853   * $p < 0.05$
854   CGI = Clinical Global Impression
855   BPRS = British Psychological Rating Scale
856
857   c.   Rofecoxib
858
859 It is widely believed that individual patients respond differently to different NSAIDs. To
860 examine this belief, a controlled trial was conducted in which osteoarthritis patients identified
861 as non-responders to celecoxib were randomized to celecoxib or rofecoxib. In fact, there was
862 no observed difference (Figure 3).
863
864 **Figure 3: Study Comparing Rofecoxib to Celecoxib in Celecoxib Non-Responders**
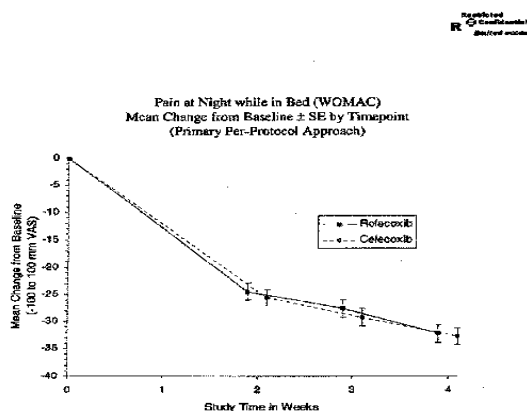
865
866 *Note*: there was considerable and prompt improvement in pain in both groups. A baseline-
867 controlled, single-arm trial of rofecoxib would have led to a clearly erroneous conclusion, and
868 even a placebo-controlled trial of rofecoxib in this population might have shown an effect that
869 would have been incorrectly interpreted as an effect in celecoxib non-responders.

870
871   2.   *Study in Intolerants: Angiotensin Receptor Blockers (ARBs) in People Who Cough on*
872   *Lisinopril*
873
874 Studies of the tolerability of a new drug in people who do not tolerate a previous treatment are also
875 informative and efficient. Comparative studies in an unselected population could provide some
876 information on relative tolerability, but a very large study would be needed to show small differences.
877 For example, if the true rates of cough for an angiotensin converting enzyme inhibitor (ACEI) and an
878 angiotensin II receptor antagonist (ARB) were 5% and 1%, respectively, a study with 90% power to
879 show a difference in an unselected population would need about 800 patients. In contrast, a study in
880 patients known to cough on ACEIs would need fewer than 20 patients, if, for example, the cough rate
881 were > 90% in the ACEI arm and 20% in the ARB arm.
882
883 This approach was used in a study of 84 elderly hypertensive patients with a history of coughing on an
884 ACEI.[49] They were then withdrawn from their ACEI and given 8 weeks of the ACEI lisinopril, which
885 had to cause at least moderate coughing for patients to continue on the study. Lisinopril was then
886 withdrawn for 4 weeks and coughing had to disappear. The patients were then randomized to losartan
887 50 mg, lisinopril 10 mg, or metolazone (diuretic active control that does not induce coughing) for 10
888 weeks. The study achieved a very persuasive result with this small population (see Table 6).
889
890   **Table 6: Comparison of Coughing Rates with ARB, ACEI, and an Additional Active**
891   **(Non-Cough-Inducing) Control**
892

	Lisinopril	**Losartan**	**Metolazone**
	N = 28	N = 28	N = 28
Any Cough	97%	18%	21%
	P < 0.001		

893
894
895
896 **VI.**   **ENRICHMENT STUDY DESIGN AND OTHER CONSIDERATIONS**
897
898 **A.**   **General Considerations**
899
900 In general, enrichment studies should incorporate the established principles of well-controlled studies,
901 controlling bias (randomization and blinding), and studywise type I error (see section VI.C).
902
903 An enrichment design should be explicitly described in the protocol and study report and should fully
904 detail the enrichment maneuvers and their impact on interpretation of results. For example, if only
905 half of patients screened meet the selection criterion, the implications of this finding on expected
906 response rate in the overall population and generalizability of the results should be evaluated. Such
907 descriptions are particularly important for trials in which high-risk patients (prognostic enrichment)
908 and probable responders (predictive enrichment) have been selected, when the description is critical to
909 knowing to which patients the results apply.
910
911

912   *1.   Performance Characteristics of a Screening Strategy for Selecting Patients*
913

914   Some prognostic and predictive enrichment strategies depend heavily on a screening measurement for
915   selecting the enriched population. For less well-established measures, such as use of proteomic and
916   genomic markers to identify responder populations, there are generally more performance
917   uncertainties than for phenotypically well-known prognostic factors such as blood pressure or
918   cholesterol. It is critical to understand, to the extent possible, the accuracy of the test used for
919   enrichment, as well as the performance characteristics (sensitivity, specificity) of the measurement
920   used. If a selection criterion does not accurately distinguish patients, the effect of enrichment is
921   weakened. If a classifier is overly inclusive (lacks specificity), the estimated difference between the
922   effect in the enriched and non-enriched populations will be attenuated, defeating the goal of the
923   enrichment strategy. If a classifier is overly exclusive (i.e., lacks sensitivity), patients who could
924   benefit from the drug will not be studied, and study subjects will be needlessly difficult to find. In a
925   non-inferiority study, an inaccurate classifier would bias the study toward a finding of no difference.[40]
926

927   Apart from the adverse impact of a poor classifier on a study, the performance characteristics of the
928   classifier remain important if the classifier is to be used in selection of patients after approval of the
929   drug. In some cases, the performance characteristics of a classifier will be well-documented in earlier
930   studies, but the enriched trial may itself constitute an important source of information about the
931   sensitivity and specificity of the classifier. For example, by including patients above and below a
932   presumed classifier cut off and examining results by classifier status, it is possible to assess the
933   quantitative relation of the enrichment factor to response (e.g., the relationship of a receptor level such
934   as HER 2/neu to response). This relationship can be explored even if the primary analysis defines the
935   marker-positive population of interest on the basis of a pre-specified cut-off. The classifier is pertinent
936   not only to selection of patients for the study, but to selection of patients for treatment once the drug is
937   marketed. Planned interim evaluations could be used to narrow selection criteria, especially if there
938   were a measure of effect that could be assessed early (e.g., short-term pharmacologic effect).
939

940   *2.   When Should a Classifier Be Developed and Characterized?*
941

942   During a development program using an enrichment strategy, it is not always clear when the
943   performance of a classifier should be characterized, nor how well the classifier and assay
944   methodologies must perform if they do in fact define a population with increased response. Answers
945   to these questions depend in part on when during development it becomes clear that an enrichment
946   strategy should be used and how critical enrichment will be to successful drug development and
947   approval. When there is not strong pre-existing information defining a responder subset, early
948   exploratory studies that include patients over a broad range of values of the enrichment factor can be
949   used to develop criteria for classifying patients for subsequent enrichment studies. If the need for
950   classifier-based patient selection does not become apparent until late in development, it is also possible
951   to first characterize a classifier in a phase 3 trial. When the classifier is not based on an intrinsically
952   binary measurement, the trial could explore a range of threshold values for a classifier that would be
953   used to identify patients for the primary analysis, but this must be planned in advance to control the
954   Type I error rate of the study. With few exceptions, the enrichment characteristics used in
955   confirmatory studies should be measured at baseline, and patients who are classified as having, or not
956   having, the predictive marker should be stratified and randomly assigned to treatments if both
957   subgroups of patients are to be included. If the classifier is known only after randomization, but is a

958 baseline characteristic, randomization should be effective even without stratification, as long as the
959 sample sizes in the treatment groups within each marker-defined subgroup are large enough to balance
960 important prognostic baseline factors.[43]
961
962 **B.   Which Populations to Study**
963
964 As will be described in more detail below, trials can be designed (1) to include only patients with the
965 enrichment factor or (2) to include patients with and without the enrichment factor, but with an intent
966 to analyze only the patients with the enrichment factor as one of the primary study hypotheses.
967 Studies including both populations need not include a wholly unselected population of patients with
968 the disease to be treated, but can designate separate sample sizes for patients with and without the
969 enrichment characteristic to collect sufficient information to demonstrate effectiveness in the enriched
970 subgroup and also to allow a reasonable estimate of effect in the non-enriched group. Many design
971 alternatives have been discussed in the literature.[50,51]
972
973 A critical question in all settings in which enrichment is used is therefore the extent to which the
974 enrichment marker-negative population should be studied, an issue that may bear importantly on how
975 a drug would be labeled. In some cases, study of the general population (one including the marker-
976 negative population) would not be expected. For example, if prognostic enrichment is used to ensure
977 that there are sufficient events to make a trial feasible, even if it is thought that the treatment effect
978 would also be present in the lower-risk population without the marker (but at a lower absolute effect
979 size), it may not be possible to design a trial that includes a significant fraction of the marker-negative
980 population without greatly increasing the sample size needed − a strategy that may make the trial
981 impractical and defeat a major purpose of prognostic enrichment. The Clinical Studies section of
982 labeling and sometimes the Indications section would identify the population studied and comment, as
983 appropriate, on use in other populations. Advice regarding use in the untested marker-negative
984 population would depend on the particular circumstances. For example, the presence of significant
985 toxicity could lead to doubts about the advisability of using the drug in the lower-risk population. The
986 heterogeneity-reducing factors discussed in Section III would not ordinarily call for study of the
987 population lacking the enrichment factor (e.g., poor compliers).
988
989 It is principally in the area of predictive enrichment, especially predictive enrichment using a
990 pathophysiological or genomic marker, that the question of studying the population without the
991 enrichment factor is most germane. Experience suggests that the selected enrichment factors often do
992 not prove to precisely dichotomize patients into subpopulations that will and will not respond, so that
993 it is usually desirable to obtain some information on the marker-negative population to assess
994 performance of the factor. However, even an imperfectly characterized predictive marker can greatly
995 increase the power and likelihood of success of a study. Moreover, in treating serious and life-
996 threatening illnesses, especially when there are alternative treatments, using the test treatment in
997 patients thought unlikely to respond raises critical ethical issues.
998
999 Efforts to use predictive enrichment thus offer a number of design choices. The study designs
1000 illustrated below are fixed sample size designs that can be used with predictive enrichment strategies
1001 (also see section VI.D below concerning adaptive enrichment and non-fixed sample size). The
1002 examples all describe trials intended to show superiority of the test treatment to a control (placebo,
1003 standard of care), but non-inferiority studies would present similar issues.

1004
1005   *1.   Studying Marker-Positive Patients Only*
1006
1007   A study randomizing only marker-positive patients is shown in Figure 4. Because a study that uses a
1008   marker-positive only population will provide no direct information about the marker-negative
1009   population, its use should generally be limited to situations in which information about the marker-
1010   negative population is not needed or is not feasible given the objectives of the study. For example, if it
1011   appears clear, based on mechanistic, pre-clinical, or early clinical data, that the marker-negative
1012   patients will have no or minimal response or would be exposed to unreasonable risk, inclusion of the
1013   marker-negative patients would, in most cases, not be justified.
1014
1015
1016   **Figure 4**
1017

Prospective, Screened - no possible effect in the marker-negative group
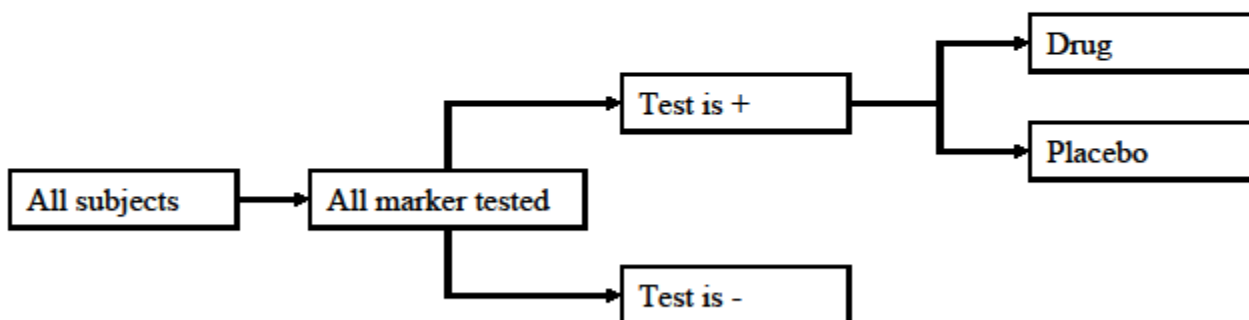
1018
1019
1020
1021   The study shown in figure 4 would support an effectiveness claim for the enriched population, but it
1022   would overstate the actual effectiveness for an unselected population, so that the fraction of patients
1023   with the marker would be important information. The study would provide no new clinical evidence
1024   with respect to the marker negative population and would not further characterize the predictiveness of
1025   the marker because there would be no ability to compare effectiveness. Because there will be no
1026   effectiveness or safety information on the enrichment-marker-negative patients, it is implicit in this
1027   approach that the selection process would be fully described in the labeling and that in clinical
1028   practice, in most cases, all patients would be tested for the enrichment marker before exposure.
1029   Moreover, because assessment of marker status is critically important to effectiveness in patients in
1030   clinical practice, it would generally be expected that the enrichment marker would be measured after

1031   approval using an established, FDA approved, laboratory test explicitly labeled for this purpose as a
1032   companion diagnostic.[55]
1033
1034   *2.   Studying Both Marker Positive and Negative Patients*
1035
1036   We encourage inclusion of some predictive marker-negative patients in most trials intended to provide
1037   primary effectiveness support, unless it has been well established in earlier studies that the marker-
1038   negative patients do not respond, or there is a strong mechanistic rationale that makes it clear that they
1039   will not respond. In general, the greater the uncertainty about the marker cut-off and responsiveness
1040   of marker-negative patients, the more important it is to include a reasonable sample of marker-
1041   negative patients. When substantial incentive exists to use the drug in the marker-negative population
1042   (e.g., for serious diseases with few alternative therapies), characterization of the response in the
1043   marker-negative population is more important, especially if the drug has important safety concerns.
1044
1045   There are two cases to consider in studies that include marker-positive and marker negative patients:
1046   (1) when the marker can be assessed before randomization and (2) when the marker can be assessed
1047   only after randomization. Figures 5A and 5B provide sample study designs for these two cases. [51]
1048
1049   **Figure 5A**
1050

Patients Prospectively Screened and Stratified - where there is possible effect in the marker-negative group
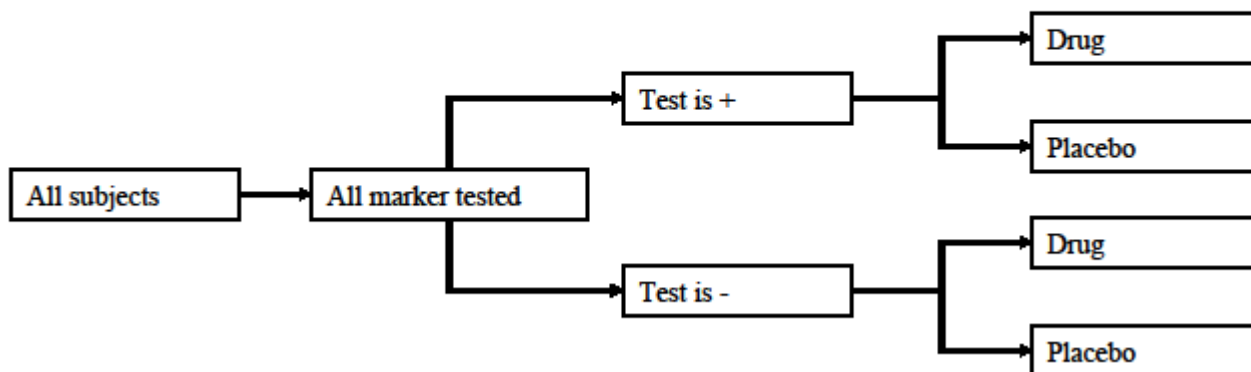
1051
1052
1053
1054   In the first case, marker status is determined for all patients, and randomization is stratified by marker
1055   status. The primary study objective would usually be a statistically rigorous demonstration of the
1056   treatment effect in the marker-positive patients, and the study would therefore be powered for the
1057   effect in that group. The size of the marker-negative group would be determined separately (i.e., it
1058   would not be necessary to randomize all marker-negative patients). Because the treatment effect
1059   would be expected to be much smaller (if there were any effect) in the marker-negative population, the
1060   size of the marker-negative population would usually be too small to give a definitive answer on the
1061   effect in that population; however, it would provide at least some estimate of the effect in that
1062   population. The design could also provide an overall risk–benefit assessment for the drug in a general
1063   population, which would be advantageous if some exposure in marker-negatives is anticipated in

1064 clinical practice (e.g., because the test is not widely available). When there is substantial uncertainty
1065 about whether a marker is predictive, (i.e., can select a population in which treatment is effective), the
1066 primary endpoint could be the effect in the overall population, or study alpha could be divided
1067 between the two endpoints (overall population and marker-positive population).[52]
1068
1069 **Figure 5B**
1070
1071

All patients randomized because marker
cannot be assessed prior to randomization
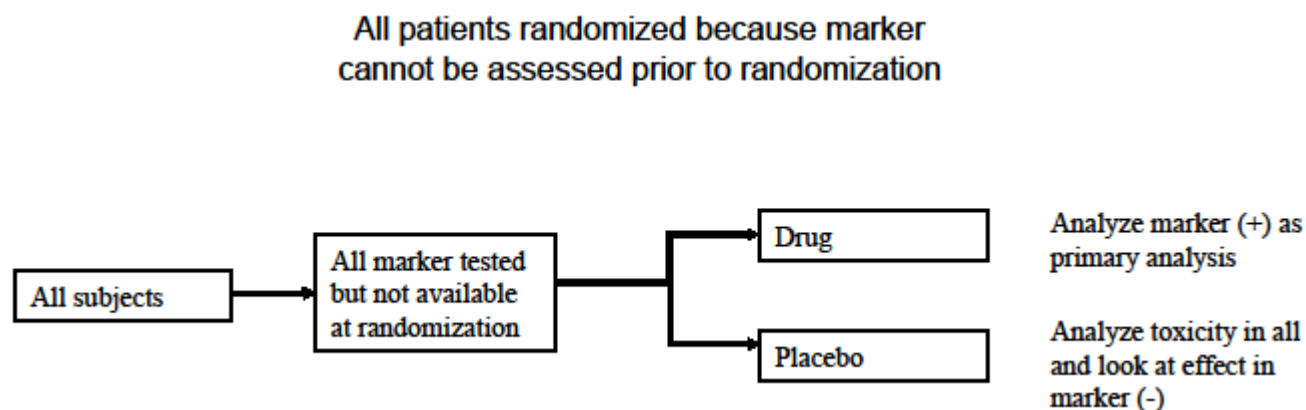
1072
1073
1074 The second case (Fig 5B) is one in which a drug that is expected to be effective only in the marker-
1075 positive subset (e.g., only in patients with a sensitive organism) must nonetheless be given to all
1076 patients because the marker result is not available at randomization. It would still be appropriate to
1077 have the primary study outcome be the effect in the marker-positive subset, but the risk–benefit
1078 assessment would reflect results in the entire population (i.e., the population that would be exposed to
1079 treatment). In these cases, when marker test results for patients using the drug will not be known
1080 before drug administration and if no patient management decisions will be made on the basis of the
1081 test result (e.g., decisions to discontinue treatment in marker-negative patients), FDA approval of the
1082 test prior to approval of the drug is not needed.
1083
1084   *3.   Studies in Alternative Therapy Non-Responders to Support Claim of Effectiveness in*
1085   *Non-Responders*
1086
1087 Knowing that a drug is effective in patients who have not responded to, or who have responded
1088 inadequately, to other therapy is an important clinical finding. The lack of apparent response to prior
1089 treatment, however, is not a reliable predictor of how the patient would respond in a new setting (i.e.,
1090 in a new randomized trial). It is therefore usually essential, in order to demonstrate effectiveness of
1091 the new drug in non-responders, to randomize patients to both the failed drug and the test drug (Figure
1092 6) because in most cases it cannot be assumed that the non-responders to the other therapy outside the
1093 trial will not respond to it in the setting of the new trial, as the examples in Section V illustrate. There
1094 are cases in which previous non-response will almost certainly be repeated (e.g., non-response to an
1095 oncologic treatment) so that re-randomization to the failed treatment may not be necessary. Great care
1096 should be taken in reaching this conclusion, however, as the altered conditions in a new study might
1097 affect the response to an apparently failed prior treatment (e.g., by improving compliance). It would,
1098 therefore, be important to ensure that the failed therapy was properly administered. Finally, it should

1099   be appreciated that re-randomization to a failed treatment could pose significant ethical problems in a
1100   serious or life-threatening condition.
1101
1102
1103   **Figure 6**

Studies in Non-Responders

standard drug

standard drug
non-responder

new drug

1104
1105   4.   *Studies in Patients Intolerant of a Prior Treatment*
1106
1107   As in the case of studies of non-responders, knowing that patients who have experienced important
1108   adverse effects on an available therapy do not have such effects on a new agent is a clinically
1109   important finding. In this case, too (as in studies of non-responders), it is critical to randomize to the
1110   poorly tolerated previous therapy and the new drug (Figure 7) to reach a conclusion that the new drug
1111   has superior safety, because the adverse effects do not always reappear when a treatment is repeated.
1112   For the same reason, it would also be advisable to include a placebo group to be certain that the side
1113   effect was indeed reproduced in the previous treatment group. This study design is not feasible if the
1114   adverse effect was dangerous to the patient. In that case, showing that a new drug lacks the side effect
1115   is probably feasible only in a trial in a previously untreated population, unless it is possible to know
1116   very confidently that the adverse effect would have recurred in the previously treated population that
1117   did not tolerate the initial treatment (i.e., use of a historically controlled design).
1118
1119
1120   **Figure 7**
1121   **Studies in Intolerants**
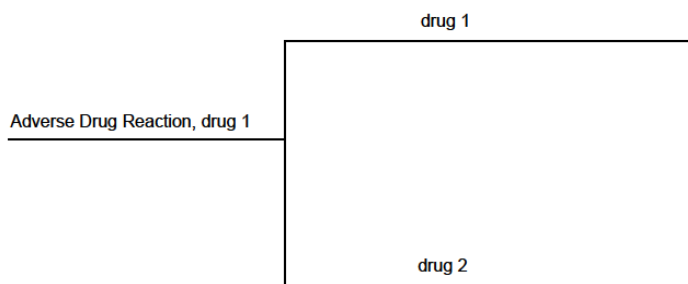
drug 1

Adverse Drug Reaction, drug 1

drug 2

1122
1123
1124
1125

1126
1127   **C.   Type I Error Rate Control for Enriched Study Subpopulations**
1128
1129   Generally, even if patients both with and without an enrichment characteristic are studied, the primary
1130   endpoint is expected to be driven by the result in the enriched subgroup. In some enrichment designs
1131   that enroll patients both with and without the enrichment characteristic, the type-I error rate for the
1132   study can be shared between a test conducted using only the enriched subpopulation and a test
1133   conducted using the entire population.[52] This alpha allocation scheme allows for assessment of the
1134   treatment effect in the entire entered population when there may be some effect in the patients without
1135   the enrichment characteristic while also allowing the assessment in the enriched subgroup.
1136   Determining the required sample size that will provide reasonable power to test the different
1137   hypotheses while controlling type-I error (usually including a pre-specified order of testing or a
1138   multiple testing procedure allowing testing of both hypotheses) is challenging.
1139
1140   **D.   Adaptive Enrichment**
1141
1142   Although an enrichment characteristic should almost always be specified before a study begins, certain
1143   adaptive designs[52] can use enrichment strategies that identify predictive markers during the course of
1144   the study. Specifically, entry criteria or sample size can be modified for later stages of a trial if factors
1145   can be identified that increase event rate or treatment response (e.g., discovery that the enrichment
1146   factor has a greater impact on response than anticipated or that the patients without the enrichment
1147   factor have a very low response or safety concern). Such changes will need appropriate type-I error
1148   rate control to account for interim, unblinded analyses of the accumulating data as well as type-I error
1149   allocation if there were analyses of multiple subgroups, but adjustment may not be needed for the
1150   enrichment if all randomized patients are included.[53,54] However, the issue of whether the statistical
1151   testing results obtained by such an adaptive enrichment strategy are reproducible needs to be
1152   addressed. For a full discussion of adaptive designs see FDA's draft Adaptive Design Guidance.[52]
1153
1154   During a trial, information may be obtained from accruing patients that could inform a pre-planned
1155   opportunity for modification of the trial's size or design (e.g., to be certain that the enriched population
1156   is of adequate size). In a trial that enrolls patients with and without enrichment factors, the relevant
1157   sample size would generally be based on the number of subjects entered with the enrichment
1158   characteristic, so that sample size adjustment based on a blinded analysis of the proportion of marker-
1159   positive patients would usually be recommended. It is possible for enrichment to occur to a greater
1160   degree in the later stages of a trial. For example, a broader population might be studied initially with
1161   altered entry criteria implemented in later stages of the trial based on interim results. Such a procedure
1162   should be described in the study protocol, and the analysis should adequately account for the changes.
1163   A number of examples of such trials have been reported in the literature, but there is little experience
1164   as yet with their use in drug development.
1165
1166   Although there has been little practical experience with enriched study designs whose sample size
1167   changes after the start of the study, or where other changes in the design are pre-planned to be based
1168   on accrued information during a trial, a number of adaptive designs seem potentially applicable:
1169
1170   1.   In a study that includes both marker-positive and negative patients, an interim look could
1171   reveal, either on an early endpoint (e.g., imaging or PD biomarker or tumor response rate) or

1172   later endpoint (e.g., progression-free survival) that the marker-negative population has a much
1173   lower response than the marker-positive group. Additional enrollment of marker-negative
1174   patients could be reduced or stopped entirely.

1175   2.   The design could be used to obtain more precise information on the performance
1176   characteristics of a classifier, for example, by examining an early endpoint using several
1177   different classifier cut-offs to determine the optimal cut-off value. If a cut-off proved too low
1178   (i.e., selected a population with few responses), it could be raised. Such plans would need to
1179   be carefully specified in advance as part of the phase 3 study, but could also be examined in
1180   exploratory studies.

1181   3.   When the optimal marker cutpoint is not well known prior to the study, use of several different
1182   cut-offs creates multiplicity issues; cut-offs should be pre-specified or corrected for the
1183   multiple choices. Jiang, Friedlin, and Simon have suggested a study design that tests the
1184   treatment effect in the overall population and in a subset defined by a marker cut-off point
1185   when the cut-off point is determined after the study is over using a statistical procedure that
1186   controls for multiplicity.[55]

1187   4.   Interim analyses could suggest changing entry criteria to emphasize a better-responding
1188   subgroup; if all randomized patients were included in the final analysis, such a step would not
1189   appear to need alpha adjustment, although describing results could be challenging.[53] It should
1190   be appreciated, however, that unplanned adjustments of entry criteria based on early data can
1191   have unexpected effects that may not enhance the ability of the study to show an effect.

1192   5.   Sample size planning in these designs can be difficult, because such designs are generally used
1193   when there is uncertainty about the prevalence of a marker, its predictiveness, and what sample
1194   size or entry criteria adjustments are contemplated. In some more complex situations, use of
1195   statistical simulations may help calculate study power and the impact of design choices.

1196
1197 **E.   Cautions in Interpretation**
1198
1199 Any use of an enrichment design should be explicit in the protocol and study report and should fully
1200 detail the rationale, specific enrichment maneuvers, and their impact on interpretation of results. For
1201 example, if only half of the patients screened passed the entry test, that should be noted, and the
1202 impact of this selection in terms of the expected response rate in the overall population and on the
1203 generalizability of the results should be evaluated. The importance of such descriptions is obvious for
1204 trials in which high-risk patients (prognostic enrichment) and probable responders (predictive
1205 enrichment) have been selected, where the description is critical to knowing to which patients the
1206 results apply, but such descriptions are important for all types of enrichment studies. Given the
1207 potentially complex interpretation of studies using enrichment designs, we strongly recommend early
1208 discussions with the Agency on plans to use them.
1209
1210 When enrichment depends on a proteomic or genetic test, particularly if the test is intended for use in
1211 practice to identify patients to be treated, the analytical validity of the test is critical. In addition to
1212 assay validity, for any marker used to select patients, even a familiar one, its sensitivity and specificity
1213 and positive and negative predictive values should be well characterized. To the extent an enrichment
1214 strategy successfully identifies patients with high event rates or high response rates and leads to a
1215 successful study, study results could be said to speak for themselves (i.e., the randomized trial *did*

1216 show an effect; the event rate *was* high enough) and certainly support the effectiveness of the drug in
1217 the population studied. Again, however, the enrichment strategies should be clearly described to
1218 indicate how the drug is to be used and to whom the results might apply (groups of patients that do and
1219 do not benefit).
1220
1221 Selection of the optimal predictive enrichment study design, specifically, whether to include both
1222 marker-positive and negative patients, and whether to introduce adaptive elements can be difficult to
1223 determine in the face of uncertainty about the properties of the enrichment marker. Many publications
1224 have addressed these issues.[41,50,51,53] One conclusion is that the greater the uncertainty regarding the
1225 marker cut-off and responsiveness of marker-negative patients, the more sense it makes to include a
1226 reasonable sample of marker-negative patients, perhaps using an adaptive design to exclude such
1227 patients if they are seen not to respond. In general, especially when marker prevalence in the
1228 population studied is relatively low, it would generally be sensible to stratify by marker status.
1229
1230
1231 **VII. ENRICHMENT − REGULATORY ISSUES**
1232
1233 **A. Summary − The Decision to Use an Enrichment Strategy**
1234
1235 The decision to use an enrichment design is largely left to the sponsor of the investigation, but like the
1236 entire research and clinical communities, FDA is very interested in targeting treatments to the people
1237 who can benefit from them (i.e., individualization). FDA's interests also include the adequacy of the
1238 study (Will it successfully assess effectiveness in a defined population and, in so doing, support
1239 marketing approval?) as well as the degree to which study findings can be described in drug labeling.
1240
1241 As discussed above, there are many reasons to use such designs, including an enhanced benefit–risk
1242 relationship if a population with an increased likelihood of response can be identified, and efficiency
1243 in drug development, as smaller studies can often be used to demonstrate effectiveness. There are,
1244 however, two critical considerations when contemplating the use of enrichment designs.
1245
1246 *1. Can the Enrichment Strategy Be Used to Identify the Patients to Whom the Drug*
1247 *Should Be Given?*
1248
1249 When patients with an increased likelihood of response can be defined before treatment by a predictive
1250 marker (a pathophysiologic or genomic characteristic, a short-term screen such as response to a test
1251 dose), a straightforward method is available for selecting patients for treatment. In contrast, some
1252 empiric strategies that provide predictive enrichment (e.g., studying known responders in a
1253 conventional study or in a randomized withdrawal study) can efficiently establish the effectiveness of
1254 a drug in a subset of the population, but provide no way for prescribers to prospectively identify
1255 patients with a greater likelihood of response, or predict the magnitude of response in an unselected
1256 patient. Although this type of untargeted treatment may seem troubling (treatment of many to attain a
1257 response in only some), the reality is that this is generally the case with treatments that are approved
1258 on the basis of conventional studies in a non-enriched population, where there is typically a wide range
1259 of responses, including no effect at all, or even harm in some cases. However, it needs to be
1260 understood and made clear that the magnitude and/or likelihood of a treatment response for an
1261 unselected patient could be substantially less than the mean response observed in a clinical trial that

1262   employed an empiric enrichment strategy. When the prescriber is reasonably able to gauge the
1263   effectiveness of a drug in an individual patient (e.g., pain is relieved, cholesterol is reduced), the pre-
1264   treatment ability to predict the likelihood of a drug response with accuracy may not be as critical.
1265
1266   In some cases, enrichment cannot be used to select patients for study because the enrichment factor is
1267   not known until after the treatment is initiated, but is used to identify the subset of the studied
1268   population to be analyzed − for example, the patients with a sensitive organism in studies of
1269   antimicrobial drugs. Again, the subset analysis documents effectiveness, but the treated population, at
1270   least initially, will be the unselected patients (i.e., a larger group than the population of potential
1271   responders). Such situations are unavoidable, however, if treatment is urgent and must be initiated
1272   before the enrichment measure is available.
1273
1274   Finally, the quality (sensitivity and specificity) of the enrichment strategy may not be critical to
1275   knowing that a drug has an effect in the study subjects (if the selected study population shows an
1276   effect, the drug was in fact effective), but it is important to therapeutic use, as it is plainly undesirable
1277   to inadvertently include non-responders and exclude potential responders because the test has poor
1278   precision or because cut-off points were poorly selected.
1279
1280   The above problems noted, however, if the enrichment strategy allows a drug of value to be developed
1281   and shown to be effective when disease and response variability would make non-enriched studies
1282   unable or unlikely to succeed, there is clearly an important gain from such strategies. Labeling will
1283   reflect limitations and concerns, but it seems clear that a drug shown effective in an enriched study
1284   should be available even if the responder population is not identified as precisely as would be
1285   desirable.

1286   *2.   Might the Drug Be Useful in a Broader Population Than Was Studied?*
1287
1288   The data that should be obtained for the marker-negative patients will be considered below, but it can
1289   be anticipated that less information will be available about them and there will be greater uncertainty
1290   as to their response to the treatment. Studies in unselected patients (i.e., a non-enriched population),
1291   the typical basis for drug approval, simply ignore the question of identifying responders and lead to
1292   treatment of many patients who will not benefit. There would thus seem to be a gain from a process
1293   that seeks to establish the characteristics that predict a drug response, rather than ignoring the varied
1294   responses and overcoming them by simply increasing sample sizes.
1295
1296   In general, then, FDA is prepared to approve drugs studied primarily or even solely in enriched
1297   populations and will seek to ensure truthful labeling that does not overstate either the likelihood of a
1298   response or the predictiveness of the enrichment factor. But the extent of data that should be available
1299   on the non-enriched subgroup should always be considered. Postmarket commitments or requirements
1300   may be requested to better define the full extent of a drug's effect (including efficacy and safety
1301   studies and trials in a broader population).
1302
1303   **B.   Data That Should Be Obtained for the Marker-Negative Patients**
1304
1305   Well-controlled enrichment studies, if successful, provide clear evidence of effectiveness in the
1306   population studied. In many cases, however, questions will remain as to how to identify the patients to

1307 which the data apply and the magnitude of effect in the marker-negative patients. In general, the
1308 heterogeneity-reducing efforts raise few problems of this kind, but the prognostic and, especially, the
1309 predictive enrichment strategies do raise them, and the remedies before or even after approval are not
1310 always clear and may be circumstance-specific.
1311
1312 For studies of serious or irreversible endpoints, showing an effect in a high-risk population (e.g., high
1313 blood pressure, high LDL with a history of MI, severe CHF) has, historically, been a prelude to later
1314 demonstrations of effects in lower-risk patients, but the effect size has sometimes been smaller (CHF),
1315 the endpoints have sometimes changed (mortality versus composite in the lower risk patients), and
1316 consequently, benefit–risk considerations may change. FDA has generally accepted the results from
1317 prognostically enriched studies, approved a claim based on the observed effect, and described the
1318 study, including the patient population, in the Clinical Studies section of labeling, with any enrichment
1319 selection criteria noted. The specific patient population studied has sometimes, but certainly not
1320 always, been given as the indicated population in the Indications section of labeling in addition to its
1321 description in the Clinical Studies section.
1322
1323 The most challenging situation is a finding of benefit in a clinically, demographically,
1324 pathophysiologically, proteomically, or genetically selected population used for predictive enrichment,
1325 because in that case there is inevitably the question of how sure we are that other patients could not
1326 benefit, even if the benefit were smaller. Ideally, therefore, there will be at least some data on the
1327 marker-negative population; the study designs in Figure 5 illustrate this approach, but it must be
1328 appreciated that a study, if sized to show an effect in the enriched population, will have relatively little
1329 capacity to detect or rule out the anticipated substantially smaller effect in the marker-negative
1330 population. Nevertheless, the design does provide some information on that population. It could, for
1331 example, show that the estimated treatment effect is not likely to be as large in the marker-negative
1332 population, but could show a trend suggesting a smaller effect.
1333
1334 In deciding how much information should be available on the marker-negative population, both before
1335 and after approval, risk–benefit assessments of several kinds should be considered. When the
1336 treatment is a critical advance for the enriched group, it would generally be unreasonable to delay
1337 approval for the enriched group, even if few data on the group without the enrichment factor were
1338 available and even if some off-label use were anticipated despite appropriate labeling (although how
1339 much off-label use would be unacceptable would depend on the expected risk of the treatment). For
1340 less important benefits, factors considered would include the expected risks to the marker-negative
1341 population from the use of the drug (related to its observed toxicity), the relative size of the marker-
1342 positive and marker-negative populations, and how convincing it is that there is no useful treatment
1343 effect in the marker-negative population, so that labeling and other information might make off-label
1344 use unlikely. If the risks of the drug are substantial, FDA will want greater confidence that patients
1345 who will not benefit from the drug will not be treated in the course of patient care. Conversely, if the
1346 risks appear low, less assurance regarding the marker-negative population would be sufficient.
1347 Ironically, a very important medical benefit (e.g., survival or prevention of significant disability), one
1348 that realistically and ethically cannot be delayed, could raise the greatest concerns because there would
1349 be great desire to use the drug in the marker-negative population. During development it is important
1350 to discuss with FDA review staff how much information should be available before drug approval.
1351

1352 *Note:* The information on the marker-negative group could come from all of the studies in the clinical
1353 program, including both studies in an unselected population and studies in the marker-negative
1354 population. The information that will describe the effect in the marker-negative group includes direct
1355 empirical evidence of clinical results in patients, pathophysiologic information, or combinations of
1356 various kinds of information.
1357
1358 A number of considerations would support collection of less (sometimes even no) information on the
1359 non-enrichment-factor population (again, this should always be discussed with FDA review staff):
1360
1361 • A clear pathophysiologic basis for concluding that the non-enriched population will not
1362 respond (e.g., because they lack the molecular target of the drug or because they cannot
1363 convert a pro-drug to its active metabolite); this could be supported by pre-clinical or clinical
1364 pharmacologic and biomarker studies
1365 • Early clinical studies that show very marked difference in response between the enrichment
1366 and non-enrichment populations.

1367 Important toxicity such that use in a less responsive population, even if a small effect could be present,
1368 would not be attractive.
1369
1370 **D. Labeling**
1371
1372 The use of enrichment designs will often have implications for labeling, especially the Indications and
1373 Usage, Dosage and Administration, and Clinical Studies sections. As noted above, prognostic
1374 enrichment will be described in Clinical Studies and has sometimes led to a description of the studied
1375 population in Indications. Use of predictive enrichment will usually lead to an indication directed at
1376 the predictive enrichment population, often with recommended testing, and a description of the
1377 selection in clinical studies. If no marker-negative patients are studied, it will be difficult to describe
1378 the effect of enrichment fully. In some cases, however, earlier data may show clearly that the marker-
1379 negative patients cannot respond. In general, significant toxicity would lead to stronger direction
1380 toward the marker-positive patients and strong directions to avoid the marker-negative patients. When
1381 an enrichment design results in the need for an approved or cleared companion diagnostic device,
1382 coordinated labeling for the therapeutic product and the diagnostic device should be provided.[55,56]

1383
1384 # *APPENDIX*
1385
1386 ## **Additional Guidance Related to Enrichment**
1387
1388 A number of Agency guidances have been issued that provide additional and related information about
1389 clinical trial designs (including enrichment designs) and demonstrating effectiveness. See especially
1390 the following draft and final guidances. Once finalized the draft guidances will represent FDA's
1391 thinking on their respective topics.
1392
1393 Agency guidances are available at http://www.fda.gov/RegulatoryInformation/Guidances/default.htm.
1394
1395 - FDA's draft guidance *Clinical Pharmacogenomics: Premarketing Evaluation in Early Phase*
1396   *Clinical Studies* focuses particularly on use and evaluation of genomic strategies in early drug
1397   development, and highlights identification of enrichment options for later trials.
1398
1399 - FDA's draft guidance for industry and FDA staff on *In Vitro Companion Diagnostic Devices*[56]
1400   defines IVD (in vitro diagnostic) companion diagnostic devices that are essential for the safe
1401   and effective use of their corresponding therapeutic products. The draft guidance describes the
1402   Agency's policies for approval and clearance and for labeling companion diagnostics,
1403   contemporaneously with approval and labeling of the therapeutic product.
1404
1405 - FDA's draft guidance *Adaptive Design Clinical Trials for Drugs and Biologics*[52] considers the
1406   case of enrichment approaches introduced only after randomization and based on interim
1407   evaluations. Such a retrospective finding would have to be carefully implemented and highly
1408   compelling to be accepted without further study.
1409
1410 - FDA's guidance *Providing Clinical Evidence of Effectiveness for Human Drug and Biological*
1411   *Products*[57] describes the amount and type of evidence needed to demonstrate effectiveness, and
1412   is applicable to studies using enrichment designs.
1413

REFERENCES

1. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Eng J Med 1980; 303: 1038-1041.

2. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of Treatment on Morbidity in Hypertension: Results in Patients with Diastolic Blood Pressures Averaging 115 through 129 mm Hg. JAMA 1967; 202: 1028-1034.

3. Veterans Administration Cooperative Study Group on Anti-hypertensive Agents. Effects of treatment of morbidity in hypertension II: results in patients with diastolic blood pressure averaging 90 through 114 mmHg. JAMA 1970; 213: 1143-1152.

4. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. N Eng J Med 1989; 321: 129-15.

5. Lang JM, Buring JE, Rosner B, Cook N, Hennekens CH. Estimating the effect of the run in on the power of the Physicians' Health Study. Stat Med 1991; 10: 1585-1593.

6. The CONSENSUS Study Group. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). N Engl J Med 1987; 316: 1429-35.

7. The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991; 325: 293-302.

8. The SOLVD Investigators. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. N Engl J Med 1992; 327: 685-91.

9. Pedersen TR, Olsson AG, Faergeman O, Kjekshus J, Wedel H, Berg K, Wilhelmsen L, Haghfelt T, Thorgeirsson G, Pyörälä K, Miettinen T, Christophersen B, Tobert JA, Musliner TA, Cook TJ for the Scandinavian Simvastatin Survival Study Group. Lipoprotein changes and reduction in the incidence of major coronary heart disease events in the Scandinavian Simvastatin Survival Study (4S). Circulation 1998; 97: 1453-1460.

10. Shepherd J, Cobbe SM, Ford I, et. al. West of Scotland Coronary Prevention Study Group. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. N Engl J Med 1995; 333: 1301-7.

11. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008; 359: 2195-207.

12. FDA guidance for industry: Diabetes Mellitus—evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. December, 2008.

13. Fisher, B. Constantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 1998; 90: 1371-88.

14. Javier X, Empana JF, Schwartz PJ, Desnos M, Cowben D, Ducimetiere P. Heart-rate profile during exercise as a predictor of sudden death. N Engl J Med 2005; 352: 1957-8.

1453   15. Ripatti S, Tikkamen E, Orho-Melander M, et al. A multi-locus genetic risk score for coronary heart
1454   disease: case-control and prospective cohort analyses. Lancet 2010; 376: 1393-1400.

1455   16. Polonsky TS, McClelland RL, Jorgenson NW, et al. Coronary artery calcium score and risk
1456   classification for coronary heart disease prediction. JAMA 2010; 303: 1610-1616.

1457   17. Ioannidis JP, Tzoulaki I. What makes a good predictor? The evidence applied to coronary artery
1458   calcium score. JAMA 2010; 303: 1646-1647.

1459   18. D'Amico AV, Chen MH, Roehl KA, Catalona WI. Preoperative PSA velocity and the risk of death
1460   from prostate cancer after radical prostatectomy. N Engl J Med 2004; 351: 125-135.

1461   19. Fan C, Oh DS, Wessels L, Welgeh B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM.
1462   Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006; 355:
1463   560-69.

1464   20. www.cancer.gov/cancertopics/factsheet/Risk/BRCA

1465   21. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical
1466   trials. Clin Cancer Res 2004; 10: 6759-6763.

1467   22. Temple R. Special study designs: early escape, enrichment studies in non-responders.
1468   Communication in Statistics. Theory and Methods 1994; 23: 499-531.

1469   23. Roden DM, Reele SB, Higgins SB, Mayol RF, Gaimmons RE, Oates, JA, Woosley RL. Total
1470   suppression of ventricular arrhythmias by encaiside. N Engl J Med 1980; 302: 877-882.

1471   24. Echt DS, Liebson PR, Mitchell LB, Pears RW, Obias-Manno D, Barker AH, Arensberg D, Baller
1472   A, Friedman L, Greene HL, Huther ML, Richardson DW, the CAST Investigators. Mortality and
1473   morbidity in patients receiving encainide, flecainide, or placebo: The Cardiac Arrhythmia
1474   Suppression Trial. N Engl J Med 1991; 324: 781-788.

1475   **25.** The Cardiac Arrhythmia Pilot Study (CAPS) Investigators. Effects of encainide, flecainide,
1476   imipramine and moricizine on ventricular arrhythmias during the year after acute myocardial
1477   infarction. The CAPS. Am J Cardiology 1988; 61: 501-509.

1478   26. The Cardiac Arrhythmia Suppression Trial-II Investigators. Effect of antiarrhythmic agent
1479   moricizine on survival after myocardial infarction. N Engl J Med 1992; 327: 227–33.

1480   27. Pablos-Méndez A, Barr RG, Shea S. Run-in periods in randomized trials: Implications for the
1481   Application of Results in Clinical Practice. JAMA 1998; 279: 222-225.

1482   28. ICH E-4 Dose-Response Information to Support Drug Registration.
1483   http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm121568.ht
1484   m.

1485   29. Temple R, Stockbridge N. BiDil for heart failure in black patients: The US Food and Drug
1486   Administration perspective. Ann Int Med 2007; 146: 57-62.

1487   30. Taylor AL, Ziesche S, Yancy C, Carson P, D'Agostino R Jr, Ferdinand K, et al. Combination of
1488   isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med 2004; 351: 2049-57.

1489   31. Cohn JN, Archibald DG, Ziesche S, Franciosa JA, Harston WE, Tristani FE, et al. Effect of
1490   vasodilator therapy on mortality in chronic congestive heart failure. Results of a Veterans
1491   Administration Cooperative Study. N Engl J Med 1986; 314: 1547-52.

1492   32. Cohn JN, Johnson G, Ziesche S, Cobb F, Frances G, Tristani F, et al. A comparison of enalapril
1493   with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. N Engl J
1494   Med 1991; 325: 303-10.

1495   33. Stroobants S, Goeminne J, Seegers M, et al. [18]FDG-Positron emission tomography for the early
1496   prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec®).
1497   Euro J Cancer 2003; 39: 2012-2020.

1498   34. Prior JO, Montemurro M, Orcurto MV, et al. Early prediction of response to sunitinib after
1499   imatinib failure by 18F-fluorodeoxyglucose positron emission tomography in patients with
1500   gastrointestinal stromal tumor. J Clin Oncol 2009; 27: 439-45.

1501   35. Smith I, Procter M, Gelber RD, et al. 2-Year follow-up of trastuzumab after adjuvant
1502   chemotherapy in HER2-positive breast cancer: a randomized controlled trial. Lancet. 2007; 369:
1503   29-36.

1504   36. Perez EA, Romond EH, Sumam VJ, et al. Updated results of the combined analysis of NCCTG
1505   N9831 and NSABP B-31 adjuvant chemotherapy with/without trastuzumab in patients with HER2-
1506   positive breast cancer. J Clin Oncol 2007; 25: 65 (suppl 185, abst 512).

1507   37. US Food and Drug Administration Drug Approval Summaries: Imatinib Mesylate, Mesna Tablets,
1508   and Zoledronic Acid; The Oncologist 2002; 7: 393-400.

1509   38. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Early Breast Cancer Trialists'
1510   Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast
1511   cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005; 365:
1512   1687-1717.

1513   39. Chapman P, Hauschild A, Robert C, et al. Improved Survival with Vemurafenib in Melanoma with
1514   BRAF V600E Mutation. N Engl J Med 2011; 364: 2507-2516.

1515   40. Wang SJ, Hung HMJ, O'Neill RT. Genomic classifier for patient enrichment: misclassification and
1516   type I error issues in pharmacogenomics non-inferiority trial. Statistics in Biopharmaceutical
1517   Research 2011;3: 310-319.

1518   41. Freidlin B and Simon R. Adaptive signature design: an adaptive clinical trial design for generating
1519   and prospectively testing a gene expression signature for sensitive patients. Clinical Cancer
1520   Research 2005; 11: 7872-7878.

1521   42. Friedlin B, Jiang W, Simon R. The cross-validated adaptive signature design. Clinical Cancer
1522   Research 2010; 16:691-698.

1523   43. Wang SJ, O'Neill RT, Hung HMJ. Statistical considerations in evaluating pharmacogenomics-
1524   based clinical effect for confirmatory trials. Clinical Trials 2010; 7: 525-536.

1525   44. Ge D, Fellay J, Thompson A, et al. Genetic variation in IL28B predicts hepatitis C treatment-
1526   induced viral clearance. Nature 2009; 461: 399-401.

1527   45. Ghany MG, Nelson DR, Strader DB Thomas DL, Seef LB. An update on treatment of genotype 1
1528   chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the
1529   Study of Liver Disease. Hepatology 2011; 54: 1433-1444.

1530   46. Amery W, Dony J. A clinical trial design avoiding undue placebo treatment. J Clinical
1531   Pharmacology 1975, 15: 674-679.

1532   47. Antman E, Muller J, Mac Alpin R, Rubenfire M, Tabatznik B, Liang C, Hempler F, Achuff S,
1533   Reichele N, Geltman E, Kerin N, Nelf R, Paramawalt E. Nifedipine therapy for coronary artery
1534   spasm. N Engl J Med 1980; 302: 1269-1273.

1535   48. Fisher B, Dignam J Bryant J, et al.: Five versus more than five years of tamoxifen for lymph node-
1536   negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel
1537   Project B-14 randomized trial. JNCI 2001; 93: 684-90.

1538   49. Chan P, Tomlinson B, Huang TY, Ko JT, Lin TS, Lee YS. Double-blind comparison of losartan,
1539   lisinopril, and metolazone in elderly hypertensive patients with previous angiotensin-converting
1540   enzyme inhibitor-induced cough. J Clin Pharmacol 1997; 37: 253-6.

1541   50. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical
1542   considerations and practical challenges. J Clin Oncol 2009; 27: 4027-4034.

1543   51. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. JNCI
1544   2010; 102: 152-160.

1545   52. Adaptive Design Guidance.
1546   http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm121568.htm

1547   53. Mehta C, Gao P, Bhatt DL, Harrington RA, Skerjanec S, Ware J. Optimizing trial design:
1548   sequential, adaptive and enrichment strategies. Circulation 2009; 119: 597-605.

1549   54. Wang SJ, Hung HMJ, O'Neill RT. Adaptive patient enrichment designs in therapeutic trials.
1550   Biometrical Journal 2009; 2: 358-374.

1551   55. Jiang W, Friedlin B, Simon R. Biomarker adaptive threshold design: a procedure for evaluating
1552   treatment with possible bio-marker-defined subset effect. JNCI 2007; 99: 1036-43.

1553   56. FDA draft guidance for industry: In Vitro Companion Diagnostic Devices.

1554   57. FDA draft guidance for industry: Providing Clinical Evidence of Effectiveness for Human Drug and
1555   Biological Products.

1556