Wiley StatsRef:
Statistics Reference Online

Enrichment Design

By *Valerii V. Fedorov*[1] *and Tao Liu*[2]

Keywords: *Clinical trials, design of experiments, mixed population, population enrichment, randomized discontinuation*

Abstract: Enrichment designs usually consist of two or three stages, where the first stage serves as a screening process for selecting a certain subpopulation, and the succeeding stages serve to distinguish the treatment effect from the placebo effect, within the selected (enriched) subpopulation. The efficiency of its efficacy (response) detection comes at the expense of the ability to make inferences for the general population; that is, the results are valid only for the selected subpopulation. At the same time, if a valid partitioning of the population is proven, then it might be viewed as evidence of "latent" classifiers, which can be targeted for a search. Statistical analysis commonly makes explicit use of the outcomes from only the selected subpopulation (final stage). Some simple models can take advantage of the outcomes from all stages. Most published studies using these designs have been conducted in the field of clinical oncology, but enrichment designs have become increasingly popular in various therapeutic areas.

Enrichment designs for evaluating certain treatments or drugs had been used for decades before Hallstrom and Friedman [1], Temple [2], and Pablos-Mendez et al [3] gave them formal discussion and definition in 1990s. In such designs, a subpopulation is selected or screened out from the general population for an experimental study. The procedure for selection of such a subpopulation is called *enrichment* [2]. The goal of the design is to enhance the signal of an external intervention in the enriched subpopulation and separate it from the interference of many other undesired factors. The discussion and employment of such designs can be traced back to the 1970s [4]. Recent years have seen these designs gain great popularity in many different disciplines, particularly in the field of clinical oncology.

The enrichment intentions can be roughly classified into the following categories [2].

[1] GlaxoSmithKline Pharmaceuticals, Biomedical Data Sciences, Collegeville, Pennsylvania
[2] University of Pennsylvania School of Medicine, Department of Biostatistics & Epidemiology, Philadelphia, Pennsylvania

This article was originally published online in 2007 in Wiley Encyclopedia of Clinical Trials, © John Wiley & Sons, Inc. and republished in Wiley StatsRef: Statistics Reference Online, 2014.

- **Variance suppressing selection.** Selecting the most homogeneous subpopulation, such as those patients with the greatest tendency to adhere to the study protocol, those whose cholesterol level falls within a certain range, or those with a similar tumor size and health condition [5].
- **Response enhanced selection.** Identifying the subpopulation with the strongest potential magnitude of response, such as improvement in mental condition, extension of survival, or reduction in tumor growth rate [6–9].
- **Target responder selection.** Choosing the subpopulation that is more likely to respond or to experience an event than the general population, such as people who responded to a treatment at the initial stage, or those with a recurrent history of a certain event [10, 11].

These selection maneuvers are not mutually exclusive. An enrichment process can often achieve more than one of the above objectives. An effective enrichment can greatly increase the **power** of detecting the target treatment effect (if only for the selected subpopulation). However, a very strict, multistage (*See* **Multistage design**) enrichment process may lead to a small subpopulation size, and consequently to lower statistical precision (*See* **Accuracy and Precision**)or to a prolonged recruitment period. An ideal enrichment design should be based on a careful trade-off between these two aspects.

There are many variants of enrichment designs. Figures **1**, **2**, and **3** show some relatively simple schemes.

The randomized discontinuation trial (RDT) (Figure 1) was first proposed by Amery and Dony [4] as an alternative to the classic **placebo** (or comparator)-controlled, randomized clinical trial (RCT) (*See* **Randomized Controlled Trials**) to reduce the trial's duration and the degree of the patients' exposure to inert placebo. In this design, after all of the eligible population have provided informed consent for randomization, they are assigned to an experimental treatment at the first stage. This stage is called the *open stage* [4]; at the end of the open stage, the individuals' responses (often **surrogate endpoints**) are collected and evaluated by the study clinician. The individuals who had no response or showed **serious adverse effects** are excluded from the study. The rest (open-stage responders) are then randomized (*See* **Randomization: Overview**) to a placebo or to the experimental treatment (or a comparator) in a double-blind fashion (*See* **Single and Double-Blind Procedures**).

The first stage serves as a filter for removing those who are unlikely to respond to treatment. The rationale is that the nonresponders would contribute little information about the population for whom the treatment can be useful. The second stage serves to distinguish whether the treatment adds anything over the placebo effect.

A commonly accepted assumption with an RDT is that "the treatment will not cure the condition during the open stage" [4]. For this reason, an RDT is generally applied under conditions that require sustained use of a therapy [6, 12, 13], such as stabilizing tumor growth or treating some chronic disease. Another often accepted assumption with this design is that the treatment effect at the open stage will not carry over to the second stage. In **oncology**, this might mean that the tumor growth rate is uniquely defined by the treatment and is changed as the treatment is changed [6].

Traditionally, the statistical analysis of an RDT uses only the outcomes from the second stage, treating it as an RCT rendered on the enriched subpopulation. Capra [14] compared the power of an RDT with that of an RCT when the primary endpoints are individuals' survival times. Kopec et al [7] evaluated the utility and efficiency of the RDT when the endpoints are binary (*See* **Binary data**); they compared the relative sample size required for the desired power of the RDT versus the RCT under various scenarios and parameter settings. Fedorov and Liu [10] considered **maximum likelihood estimation** of the treatment effect for binary endpoints. With some moderate assumptions, they incorporated the information from the open stage into their inference.

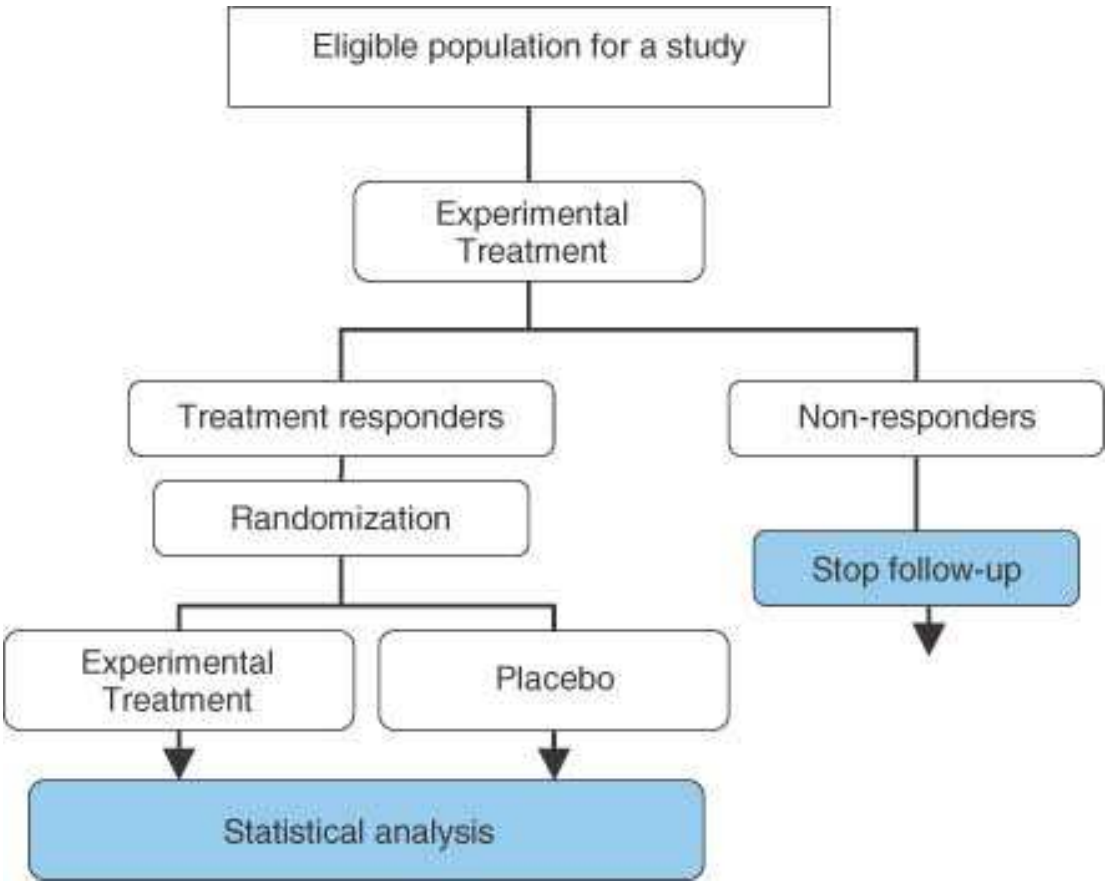

Figure 1. The diagram of a randomized discontinuation trial.

The *placebo run-in trial* (PRIT) (*See* **Run-In Period**) is another often used enrichment design [15,16]. It is very similar to the RDT in its setup structure, except that the block "experimental treatment" is replaced by "placebo treatment" and "responders" by "compliers" (Figure 2).

An assumption with the PRIT design is that the participants behave coherently throughout the trial. If a patient's adherence to the protocol is poor during the placebo run-in, then his adherence will be poor during the second stage, and vice versa. This design can be more efficient than a conventional RCT when the compliance of the general population is known or expected to be poor [5] or "when poor adherence is associated with a substantial reduction of therapy" [17]. Davis et al [18] examined the efficiency of the PRIT design through empirical evaluations, in the setting of evaluating a cholesterol-lowering drug for elderly patients. The analyses were carried out using the outcomes from the second stage only, as if it were a conventional RCT.

Both RDT and PRIT designs are fairly simple schemes. In reality, researchers often employed these designs with certain modifications to meet each study's requirements. Sometimes RDT and PRIT are even used in combination. For example, Fava et al [11] proposed a study design they named the 'Sequential Parallel Comparison Design' for their psychiatric disorder study (see Figure 3 for the design diagram). The first stage of the design consists of three double-blinded (DB) arms: two placebo arms and one treatment arm with unequal randomization (usually more patients are on the placebo arms). Only the

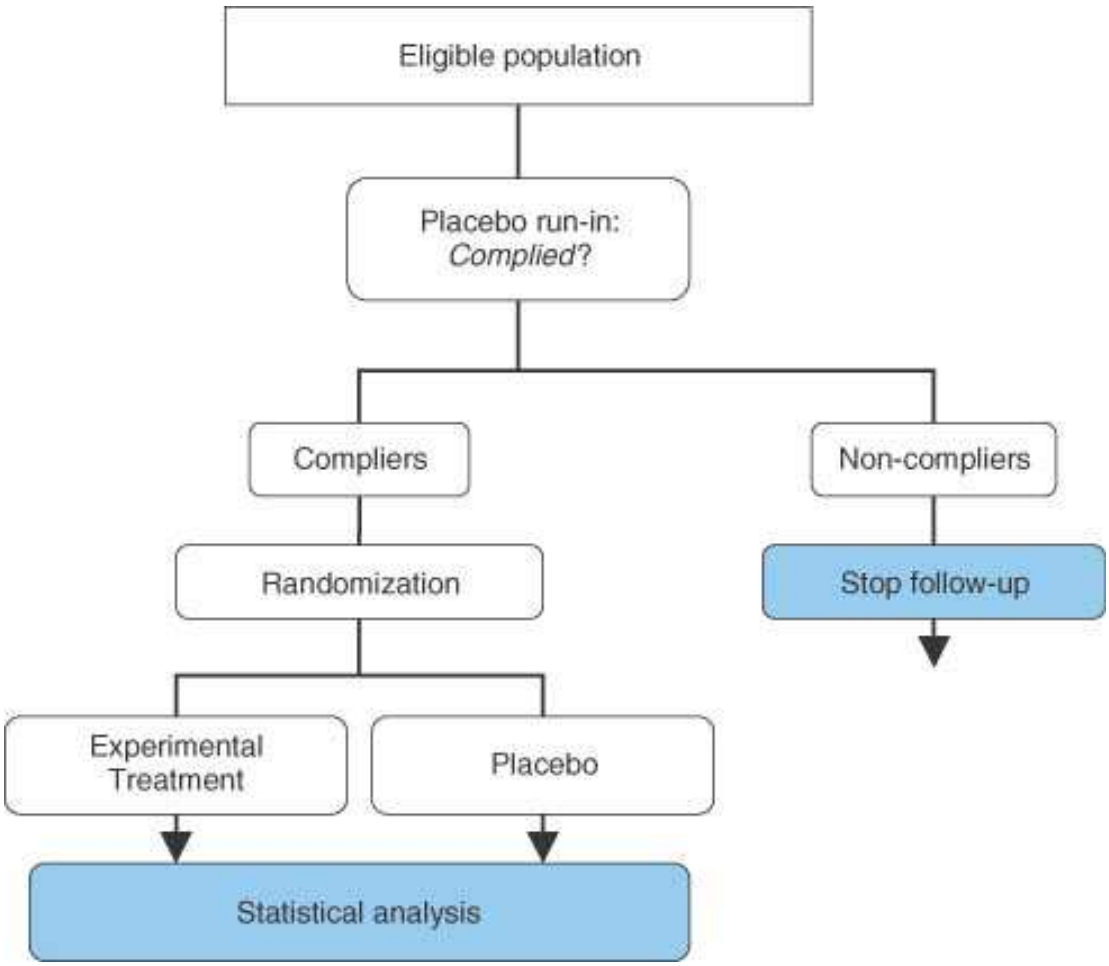

Figure 2. The diagram of a placebo run-in trial.

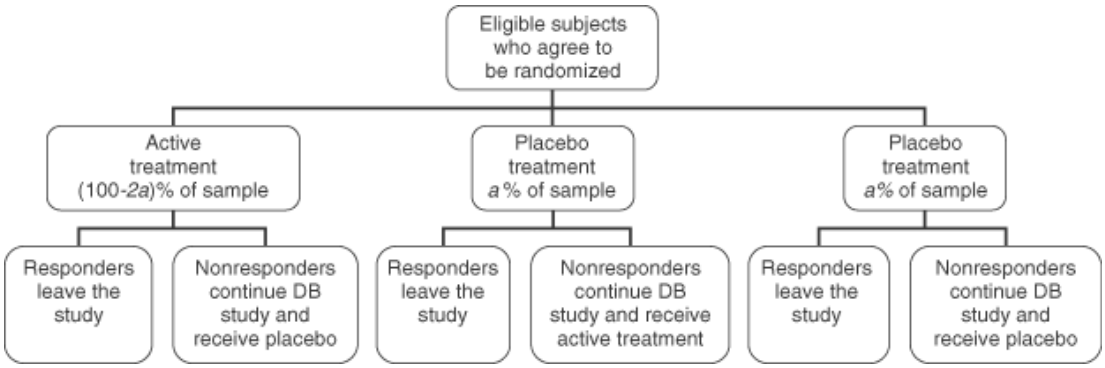

Figure 3. Study design for a major depressive disorder trial. DB, double blind. (From Fava et al [11].)

nonresponders of the first stage continue to the second stage, and they are assigned in a double-blinded way to the active treatment or placebo, depending on the arm they were on at the first stage. The rationale behind the design is that "since patients on the second stage have already 'failed placebo', their placebo response will be reduced." The data analysis of this design is similar to RDT and PRIT. Fava et al [11] have discussed the statistical model for the design and the design optimization.

1 Models

In most publications, it is (implicitly and explicitly) assumed that the general population consists of K subpopulations:

$$\bar{\phi}(x,\theta) = \sum_{k=1}^{K} \pi_k \phi(x,\theta_{kj}), \quad j = 1, 2, \ldots, J, \tag{1}$$

where x is the endpoint of interest (which can be a collection of several variables), π_k is a fraction of the k-th subpopulation, $\phi(x, \theta_{kj})$ is the distribution density of x in the k-th subpopulation under the j-th treatment, θ_{kj} are unknown parameters defining the distribution of x in each subpopulation, $\bar{\phi}(x,\theta)$ is the marginal distribution of x, and the vector θ comprises all π_k and θ_{kj}. The goal can be the estimation of all components of θ, or typically a subset of θ, such as the fraction of responders $\pi_k{}^*$ and the parameters θ_{k^*j}. In popular settings for continuous x, $\phi(x, \theta_{kj})$ is a normal density with $\theta_{1kj} = \mu_{kj} = E(x \mid k)$ and $\theta_{2kj} = \sigma_{kj}{}^2 = \mathrm{Var}(x|k)$. In this case, the parameters of interest can be μ'_{k^*j}, μ''_{k^*j} and $\pi_k{}^*$, where j' and j'' denote two comparative treatments, and the other parameters can be viewed as nuisance (*See* **Nuisance Parameters**). Often the parameter estimation is complemented or replaced by hypotheses testing.

The population model [1] should be complemented by models that describe the evolution of the response to treatment, and the observation processes. For instance, the enrichment process of an RDT is often achieved using only surrogate endpoints at the end of the first stage, which are less accurate measures than the primary endpoints (*See* **Primary Efficacy Endpoint**), and can lead to misclassification of treatment responders (*See* **Misclassification error; Misclassification models**) and nonresponders (Figure 1). Fedorov and Liu [10] proposed to model such an imperfect enrichment process through the introduction of false-positive and false-negative detection rates. This model builds the connection between the outcomes at the first and the second stages, and hence makes it possible to use the observed information from both stages by constructing the complete data **likelihood**.

In other settings [6], the outcome at the first stage is $x(t_1)$ (e.g., tumor size) at moment t_1, while at the final stage it is $x(t_1 + t_2)$. Thus, a model describing the relationship between $x(t_1)$ and $x(t_1 + t_2)$ is needed. In oncology, $x(t)$ can be a **tumor growth** model. With this model in place, the optimal selection of t_1 given $t_1 + t_2$ can be considered [6, 8].

2 Design and Efficiency

In terms of design, the choice of randomization scheme, rates of randomization, and selection of the length of the first stage can be diversified in many ways. Let us consider a simple RDT design (Figure 1) with binary outcomes as an example. Figure **4** shows the efficiency comparison between an RDT design and an RCT with two equal arms [10]. Suppose that, at the end of the open phase, 10% of the responders to an active treatment are misclassified as nonresponders, and that 10% of the nonresponders are misclassified as responders. The x-axis represents the population response rate to placebo, and the y-axis the increase in the response rate due to the treatment, the estimation of which is the primary interest. Numbers next to each curve indicate the fraction of patients randomized to the placebo arm in

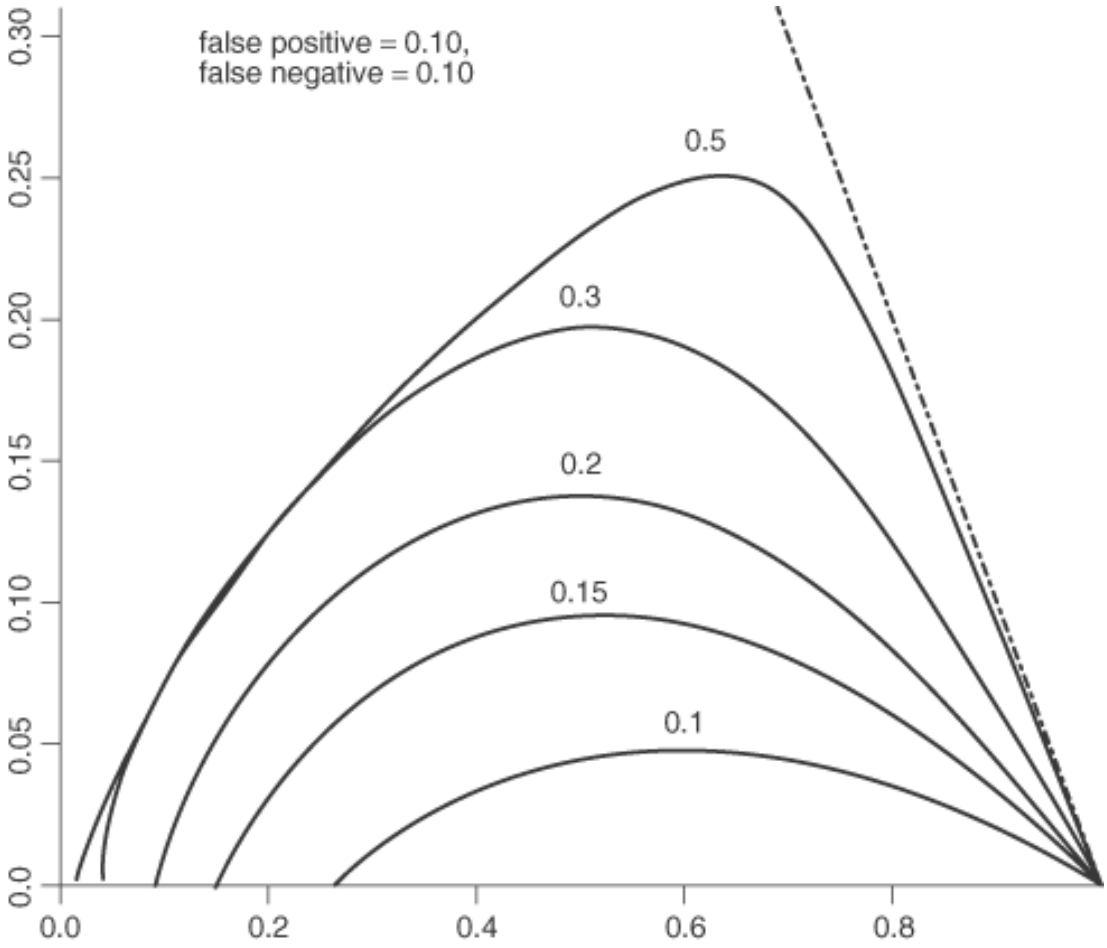

Figure 4. The area below each curve corresponds to the rate for which a randomized discontinuation trial (RDT) is superior to a randomized controlled trail (RCT) given the fraction of patients randomized to placebo.

the second stage of an RDT (usually the smaller value is more ethical). In the region under each curve, the RDT dominates the RCT in terms of efficiency. This figure demonstrates that an RDT has better performance under some scenarios (particularly when the treatment effect is small).

However this efficiency gain is not universal, even in this simple, idealized setting. For more realistic cases, careful selection of models, combined with ethical and logistical considerations, is essential. In some publications [6, 8, 9, 14], investigators use the **Monte Carlo simulations** to compare different designs.

3 Applications

Majority of applications of enrichment designs are in the field of clinical oncology. Typical examples include a study with enrichment screening based on preliminary data on erlotinib [19], a study of a cytostatic antineoplastic agent [9], and a study of the putative antiangiogenic agent carboxyaminoimoidazole in renal cell carcinoma [20]. Temple [21] provided a review and discussion of enrichment designs in cancer

treatments. For the early development of molecularly targeted anticancer agents, RDTs were employed to select sensitive subpopulations when an effect assay for such separation was not available [6, 8]. Other applications of enrichment designs can be found in children's health research [22, 23], clinical research in **psychiatry** [11, 24, 25], cardiac arrhythmia suppression study in cardiology (*See* **Cardiology and cardiovascular disease**) [26], and a few other therapeutic areas [27–30].

Similar enrichment strategies can also be found in some two-stage surveys [31]. Freidlin and Simon [8] evaluated cytostatic drugs using a design in which they expected only certain patients to be sensitive to their target treatment.

4 Discussion

Because the goal of enrichment is to separate certain subpopulations from the general population and randomize only the selected subpopulation in the trial, the enrichment design is usually only capable of detecting the efficacy rather than effectiveness. This type of efficacy distinction is often of main interest in oncology, which is the reason that RDT is frequently used for screening for a treatment activity. As certain assumptions are satisfied, the efficacy detection can be greatly enhanced. The work by Kopec et al [7] illustrated that, when compared with an RCT, the sample size required for an RDT can be reduced by more than 50%. Fedorov and Liu [10] showed that the increase in efficiency of efficacy detection can be even higher if additional assumptions are made and the information from the open stage can be seamlessly included.

However, enrichment designs are not always superior to other designs, even when all of the ethical and logistical conditions are acceptable. For example, studies that have compared the relative efficiency between RDTs and the classic RCT [7, 8, 10, 32] have found that the RCT can be more efficient under certain conditions, even when the separation of the subpopulation at the first stage is perfect (i.e., no misclassifications).

Other limitations of enrichment designs include:

1. The benefits of using enrichment designs come at the expense of the applicability of the study results to the general population. The Coronary Drug Project Research Group study (*See* **Coronary Drug Project**) [22] illustrated the effect of compliance on the conclusions for the enriched subpopulations and the general population.
2. The recruitment process for an enrichment design could last much longer than for a conventional RCT.
3. The use of surrogate endpoints of RDT at the first stage can affect the performance of enrichment process. Fedorov and Liu [10] showed the consequences of misclassification on design efficiency in the first stage of an RDT.
4. The screening phases (the run-in phase for PRIT and the open phase for RDT) are not free; they cost the **sponsors** time and money and come with errors [33, 34]. In general, an enrichment design should undergo a cost-benefit analysis [35] in which efficiency is not the only contributor to the utility function.

The enrichment (selection) process is not universally applicable to all scenarios, and may prove ethically controversial [36–38], even though its rationale is well supported and it has been natural to apply to the most commonly reported applications. Researchers must scrutinize the **validity** of the assumptions and consider the possible ethical issues (*See* **Medical Ethics and Statistics**; **Ethics in Research**) associated with enrichment before the design is carried out.

5 Related Articles

Subgroup
Run-In Period
Outcome measures in clinical trial
Multiple endpoints in Clinical Trials; Clinical trials, overview

References

[1] A. P. Hallstrom and L. Friedman, Randomizing responders. *Control Clin Trials.* **1991**; *12*: 486–503.
[2] R. J. Temple, Special study designs: early escape, enrichment, studies in non-responders. *Commun Stat Theory Methods.* **1994**; *23*: 499–531.
[3] A. Pablos-Mendez, A. G. Barr, and S. Shea, Run-in periods in randomized trials. *JAMA.* **1998**; *279*: 222–225.
[4] W. Amery and J. Dony, Clinical trial design avoiding undue placebo treatment. *J Clin Pharmacol.* **1975**; *15*: 674–679.
[5] E. Brittian and J. Wittes, The run-in period in clinical trials: the effect of misclassification on efficacy. *Control Clin Trials.* **1990**; *11*: 327–338.
[6] G. L. Rosner, W. Stadler, and M. J. Ratain, Discontinuation design: application to cytostatic antineoplastic agents. *J Clin Oncol.* **2002**; *20*: 4478–4484.
[7] J. Kopec, M. Abrahamowicz, and J. Esdaile, Randomize discontinuation trials: utility and efficiency. *J Clin Epidemiol.* **1993**; *46*: 959–971.
[8] B. Freidlin and R. Simon, Evaluation of randomized discontinuation design. *J Clin Oncol.* **2005**; *23*: 5094–5098.
[9] L. V. Rubinstein, E. L. Korn, B. Freidlin, S. Hunsberger, S. P. Ivy, and M. A. Smith, Design issues of randomized phase II trials and proposal for phase II screening trials. *J Clin Oncol.* **2005**; *23*: 7199–7206.
[10] V. V. Fedorov and T. Liu, Randomized discontinuation trials: design and efficiency. GlaxoSmithKline Biomedical Data Science Technical Report, 2005-3.
[11] M. Fava, A. E. Evins, D. J. Dorer, and D. A. Schoenfeld, The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom.* **2003**; *72*: 115–127.
[12] C. Chiron, O. Dulac, and L. Gram, Vigabatrin withdrawal randomized study in children. *Epilepsy Res.* **1996**; *25*: 209–215.
[13] E. L. Korn, S. G. Arbuck, J. M. Pulda, R. Simon, R. S. Kaplan, and M. C. Christian, Clinical trial designs for cytostatic agents: are new approaches needed? *J Clin Oncol.* **2001**; *19*: 265–272.
[14] W. B. Capra, Comparing the power of the discontinuation design to that of the classic randomized design on time-to-event endpoints. *Control Clin Trials.* **2004**; *25*: 168–177.
[15] 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.
[16] J. E. Buring and C. E. Hennekens, Cost and efficiency in clinical trials: the US Physicians' Health Study. *Stat Med.* **1990**; *9*: 29–33.
[17] K. B. Schechtman and M. E. Gordon, A comprehensive algorithm for determining whether a run-in strategy will be a cost-effective design modification in a randomized clinical trial. *Stat Med.* **1993**; *12*: 111–128.
[18] C. E. Davis, W. B. Applegate, D. J. Gordon, R. C. Curtis, and M. McCormick, An empirical evaluation of the placebo run-in. *Control Clin Trials.* **1995**; *16*: 41–50.
[19] Tarceva (erlotinib). Tablet package Insert. Melville, NY: OSI Pharmaceuticals, December **2004**.
[20] W. M. Stadler, G. Rosner, E. Small, D. Hollis, B. Rini, S. D. Zaentz, and J. Mahoney, Successful implementation of the randomized discontinuation trial design: an application to the study of the putative antiangiogenic agent carboxyaminoimidazole in renal cell carcinoma—CALGB 69901. *J Clin Oncol.* **2005**; *23*: 3726–3732.
[21] R. J. Temple, Enrichment designs: efficiency in development of cancer treatments. *J Clin Oncol.* **2005**; *23*: 4838–4839.
[22] The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med.* **1980**; *303*: 1038–1041.
[23] P. Casaer, J. Aicardi, P. Curatolo, K. Dias, M. Maia, et al, Flunazirizine in alternating hemiplegia in childhood. An international study in 12 children. *Neuropediatrics.* **1987**; *18*: 191–195.
[24] F. M. Quitkin and J. G. Rabkin, Methodological problems in studies of depressive disorder: utilities of the discontinuation design. *J Clin Psychopharmacol.* **1981**; *1*: 283–288.

[25] D. D. Robinson, S. C. Lerfald, B. Bennett, D. Laux, E. Devereaux, et al, Continuation and maintenance treatment of major depression with the monoamine oxidase inhibitor phenelzine: a double blind placebo-controlled discontinuation study. *Psychopharmacol Bull.* **1991**; *27*: 31–39.

[26] D. S. Echt, P. R. Liebson, and L. B. Mitchell, Morality and morbidity in patients receiving encainide, flecainide, or placebo. *N Engl J Med.* **1991**; *324*: 781–788.

[27] T. D. Giles, G. E. Sander, L. Roffidal, M. G. Thomas, D. P. Mersch, et al, Remission of mild to moderate hypertension after treatment with carteolol, a beta-adrenoceptor blocker with intrinsic sympathomimetic activity. *Arch Intern Med.* **1988**; *148*: 1725–1728.

[28] D. S. Echt, P. R. Liebson, L. B. Mitchell, R. W. Peters, D. Obias-Manno, et al, and the CAST Investigators. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* **1991**; *324*: 781–788.

[29] J. R. Evans, K. Pacht, P. Huss, D. V. Unverferth, T. M. Bashore, and C. V. Leier, Chronic oral amrinone therapy in congestive heart failure: a double-blind placebo-controlled withdrawal study. *Int J Clin Pharmacol Res.* **1984**; *4*: 9–18.

[30] G. H. Guyatt, M. Townsend, S. Nogradi, S. O. Pugsley, J. L. Keller, and M. T. Newhouse, Acute response to bronchodilator. An imperfect guide for bronchodilator therapy in chronic airflow limitation. *Arch Intern Med.* **1988**; *148*: 1949–1952.

[31] J. L. Vazquez-Barquero, J. F. Diez-Manrique, C. Pena, R. G. Quintanal, and M. Labrador Lopez, Two stage design in a community survey. *Br J Psychiatry.* **1986**; *149*: 88–897.

[32] C. Mallinckrodt, C. Chuang-Stein, P. McSorley, J. Schwartz, D. G. Archibald, et al, A case study comparing a randomized withdrawal trial and a double-blind long term trial for assessing the long-term efficacy of an antidepressant. *Pharm Stat.* **2007**; *6*: 9–22.

[33] E. Brittain and J. Wittes, The run-in period in clinical trials: the effect of misclassification on efficiency. *Control Clin. Trials.* **1990**; *11*: 327–338.

[34] R. J. Glynn, J. E. Buring, C. H. Hennekens, Riley, D, T. J. Kaptchuk, et al, Concerns about run-in periods in randomized trials. *JAMA.* **1998**; *279*: 1526–1527.

[35] K. B. Schechtman and M. E. Gordon, A comprehensive algorithm for determining whether a run-in strategy will be a cost-effective design modification in a randomized clinical trial. *Stat Med.* **1993**; *12*: 111–128.

[36] S. J. Senn, A personal view of some controversies in allocating treatment to patients in clinical trials. *Stat Med.* **1995**; *14*: 2661–2674.

[37] S. J. Senn, Ethical considerations concerning treatment allocation in drug development trials. *Stat Methods Med Res.* **2002**; *11*: 403–411.

[38] P. D. Leber and C. S. Davis, Threats to validity of clinical trials employing enrichment strategies for sample selection. *Control Clin Trials.* **1998**; *19*: 178–187.