

Opportunities of enrichment designs in the era of precision medicine

Pei-Fen Kuan* Bo Huang[†]

Abstract

Traditional clinical development of an experimental therapy utilizes the one-size-fits-all approach by testing the regimen in an unselected or untargeted patient population with a specific disease. The assumption is that response in the population with the disease is homogeneous. With the advent of targeted therapies, selection of treatment can be tailored to the genetic makeup of each individual. Therefore, these targeted therapies may benefit only a subset of the entire population and traditional statistical designs may no longer be appropriate or efficient. Statistical designs involving predictive biomarkers generally fall into 2 categories: classical designs and adaptive designs. We give a brief overview of the literature, and discuss the challenges and opportunities in the era of biomarker-based personalized medicine from a pharmaceutical industry perspective with some recent examples and case studies.

Key Words: Predictive biomarker, Enrichment, Adaptive design

1. Introduction

Traditional clinical development of an experimental therapy utilizes the "one-size-fits-all" approach by testing the regimen in an unselected or untargeted patient population with a specific disease. The assumption is that response in the population with the disease is homogeneous. With the advent of molecularly-targeted therapies, genetic engineering such as DNA sequencing and mRNA transcript profiling now makes a finer taxonomy of disease possible, which enables the development of precise diagnostic, prognostic, and therapeutic paradigms for specific subsets of patients, i.e. personalized medicine. Therefore, these targeted therapies may benefit only a subset of the entire population and may not benefit or even harm the rest of the population. On the other hand, proteomic and genetic biomarkers have the potential to provide substantial added value to the current medical practice. For instance, in oncology, some biomarkers provide the possibility to integrate an accurate predictor of efficacy with a specific mechanism-based therapy using the genetic makeup of the tumor and the genotype of each individual patient to guide the selection of cancer treatment.

As a result of these new opportunities and challenges, the traditional paradigm of drug development not taking into account response heterogeneity may be suboptimal. To embark on the mission of personalized medicine, innovative statistical designs beyond the interaction test of a traditional fixed design are becoming increasingly attractive, which allow assessment of treatment effects due to phenotypic or genomic heterogeneity of patients. Furthermore, focused clinical trials using a biomarker strategy may result in smaller study sizes, higher probability of trial success, enhancement of the benefit-risk relationship, and potentially mitigating ever-escalating development costs.

In 2012, FDA released a draft guidance on enrichment strategies for clinical trials to support approval of human drugs and biological products. In their guidance, the term enrichment is defined as "the prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than

*Department of Applied Math and Statistics, Stony Brook University, Stony Brook, NY 11794

[†]Pfizer Inc., 445 Eastern Point Rd, Groton, CT 06340

it would be in an unselected population”. Further, enrichment strategies fall into three broad categories (excerpts from the guidance):

- **Strategies to decrease heterogeneity** These include selecting patients with baseline measurements in a narrow range (decreased inter-patient variability) and excluding patients whose disease or symptoms improve spontaneously or whose measurements are highly variable (decreased intra-patient variability). The decreased variability provided by these strategies increases study power.
- **Prognostic enrichment strategies** choosing patients with a greater likelihood of having a disease-related endpoint event (for event-driven studies) or a substantial worsening in condition (for continuous measurement endpoints). These strategies will increase the absolute effect difference between groups but will not alter relative effect.
- **Predictive enrichment strategies** choosing patients more likely to respond to the drug treatment than other patients with the condition being treated. Such selection can lead to a larger effect size (both absolute and relative) and permit use of a smaller study population. Selection of patients could be based on a specific aspect of a patients physiology or a disease characteristic that is related in some manner to the study drugs mechanism, or it could be empiric (e.g., the patient has previously appeared to respond to a drug in the same class).

In this paper, we give an overview of population enrichment designs that are growing in the statistical literature, with a focus on designs of predictive enrichment strategies so as to increase the power and efficiency to detect an effective therapy with regard to a predictive biomarker in a clinical trial. A predictive biomarker is type of biomarker that identifies patients who are likely to benefit from a particular treatment, in contrast to a prognostic biomarker which is associated with only the disease outcome. In addition, the focus of this chapter is on the efficacy endpoint of interest, design issues with respect to predictive biomarkers on safety will not be discussed.

2. Issues to Consider

There are a number of issues one needs to consider before designing a trial with a predictive biomarker component. Its an important step as we are armed with alternative design options in a rapidly growing literature. First, we need to evaluate the strength of preclinical evidence for a potential predictive biomarker. If there is compelling preliminary evidence that the experimental therapy does not provide benefit to all the patients, and the benefit is restricted to a subset of patients expressing a molecular or genetic value, an enrichment strategy may be adopted. Otherwise, an unselected or all-comers strategy may be wise so that there is no missed opportunity for drug development, patients and the marketplace. The second issue is whether the prevalence (percentage) of the biomarker positive group is high, moderate or low. If the prevalence is high, population enrichment may be redundant and a traditional design could render the greatest commercial value and market reach. Likewise, an all-comers design may not be feasible for a low-prevalance situation. Thirdly, the accuracy of the measurements used to identify the enrichment population and the sensitivity and specificity of the enrichment criterion in distinguishing responders and non-responders are also critical issues. An accurate, reproducible and adequately validated assay is essential for establishing desired therapeutic activity and clinical validation of the biomarker (usually realized by a companion diagnostic kit from a central lab to identify the patients) in a prospective manner. In addition, the feasibility and timing to obtain a biopsy (de-novo

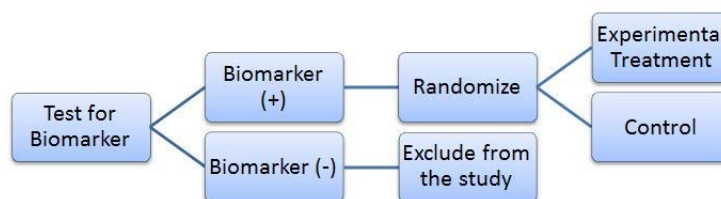


Figure 1: Biomarker enrichment design.

or archived) or serum sample at baseline prior to randomization determines whether the biomarker can be prospectively validated.

3. Classical Designs

3.1 Biomarker Enrichment Design

It is also referred to as biomarker-enrichment design or gene-enrichment design. All patients in the trial may not generally benefit from the study treatment under consideration. The goal of the enrichment designs is to study the clinical benefit in a subgroup of the patient population defined by a specific biomarker status. In this design, the patients are screened for the presence or absence of a biomarker(s) profile. After extensive screening, only patients with the presence of a certain biomarker characteristic or profile are enrolled in the clinical trial (Sargent et al., 2005; Freidlin et al., 2010). In principle, this design essentially consists of an additional criterion for patient inclusion in the trial (Figure 1). The following considerations should be taken into account in this design 1) a smaller sample size is usually required but the screening may still take the same amount of time (or even longer as explained below) as with an all-comers designs given the extensive pre-screen testing that will be conducted before enrollment; 2) the marketing label will be restricted; 3) there may still be a potential subset of patients who may benefit with the new treatment; 4) restricted enrollment does not provide data to establish that treatment is ineffective in biomarker negative patients; 5) a low prevalence of the marker may be challenging operationally and financially. Operationally, the biggest challenge is in recruitment and financially, it may not be commercially attractive.

The efficiency of this design is a function of the percent of biomarker positive patients who are likely to be benefitted by the target treatment and the reliability and reproducibility of the assay also plays a pivotal role. This design is appropriate when the mechanistic behavior of drug is known and there is compelling preliminary evidence of benefit to a subset population.

3.2 Biomarker Stratified Design

This design is chosen when there is no preliminary evidence to strongly favor restricting the trial to patients with specific biomarker profile that would necessitate a biomarker-enrichment design. This design is prospective and leads to a definitive marker validation strategy. In such cases, marker by treatment or stratified design is more informative than

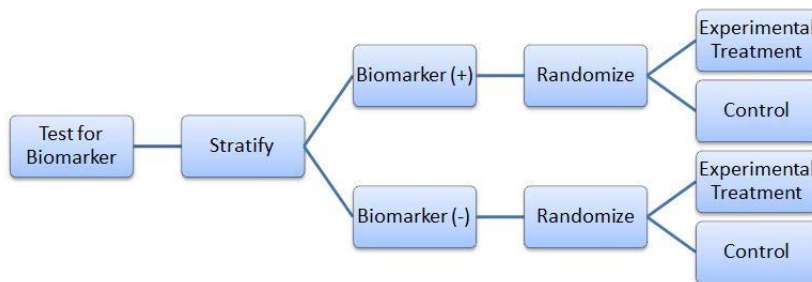


Figure 2: Biomarker stratified design.

biomarker-enrichment design. In this design, the patients are tested for biomarker status and then separately randomized according to their positive or negative status of the marker. Thus, the randomization is done using marker status as the stratification factor; however only the patients with a valid measurable marker results are randomized. Patients in each marker group are then randomized to two separate treatments (Figure 2). Two separate hypotheses tests are conducted to determine the superiority of one treatment over the other separately within each marker group. The sample size is calculated separately to power the testing within each marker sub-group. Another variation to the hypothesis test within the same design is to conduct a formal marker by treatment interaction test to see if the treatment effect varies within each marker status subgroup. In this case, the study is powered based on the magnitude of interaction. This design can be viewed as two stand-alone trials, however it is different from a large clinical trial by the calculation of the sample size and restriction of the randomization to only patients with a valid marker result.

3.3 Biomarker based Strategy Design

In this design, patients are randomly assigned to treatment dependent or independent of the marker status (Figure 3). All patients randomized to the non-biomarker based arm receive the control treatment. In the biomarker based arm, the patients receive the targeted or experimental therapy if the marker is positive and control treatment if the marker is negative. The outcome of all of the patients in the marker based sub-group is compared to that of all patients in the non-marker based sub-group to investigate the predictive value of the marker.

One downside of this design is that patients treated with the same regimen are included on both the marker-based and the nonmarker-based subgroup, resulting in a substantial redundancy leading to many patients receiving the same treatment regimes in both subgroups. Hence, this design can reduce the treatment effect especially if the prevalence of the marker is high requiring a large sample size. This is illustrated in the following example, in the ERCC1 trial (Cobo et al., 2007) and presented in Freidlin et al. (2010), about 57% of the biomarker-based strategy arm patients were assigned to the same regimen of cisplatin+docetaxel as done in the standard of care arm. Thus the comparison weakens the between-arm treatment effect difference and reduces the statistical power to reject null hypotheses. This can lead to either getting incomplete information or may miss a valuable biomarker in addition to delaying the evaluation of the biomarker due to the increased sam-

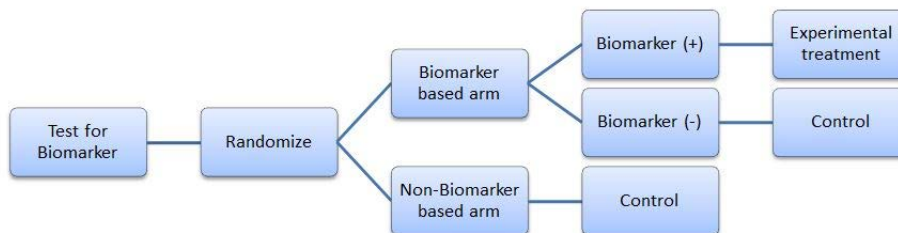


Figure 3: Biomarker strategy design.

ple size required to achieve a desired power. One other disadvantage of this design is the inability to examine the effect of targeted therapy in patients in the negative marker status group as none of these patients receive it. Even if the patients in the negative marker status group respond to the targeted therapy, this cannot be assessed. The treatment difference between the new treatment and the control treatment can be diluted by marker-based treatment selection and sometimes can be a poor choice as compared to the randomized design.

4. Novel Designs

In the past decade, the application of adaptive design methods in clinical trials has become very popular due to its flexibility and efficiency. This efficiency is at times gained at cost of increased complexity, logistical challenges especially for adequate drug supply and other operational issues. In most cases, intensive planning must be undertaken to carefully define go/no-go criteria is likely to delay the start of the trial. The sponsor should have the adequately trained resources to implement and execute adaptive designs.

4.1 Adaptive Biomarker Design

Dr. Richard Simon and his colleagues from the National Cancer Institute have developed several adaptive biomarker designs including adaptive signature design (Freidlin and Simon, 2005), cross-validated adaptive signature design (Freidlin et al., 2010) and Biomarker adaptive threshold design (Jiang et al., 2007), with the aim to Screen/select/develop biomarker/gene signature and confirm in the same trial with family-wise type I error control. These designs are simple to implement and may greatly enhance the capability of developing and validating biomarkers in an efficient manner.

For example, the cross-validated adaptive signature design was proposed to optimize the efficiency of both the development and validation components of the adaptive signature design. This design enables complete cross-validation of the entire sample available in the study. This procedure preserves the overall study wise type I error rate and increases the statistical power to detect a clinically meaningful treatment difference in the sub-group of patients who benefit from the targeted therapy.

Similar to the adaptive signature design, the initial null hypothesis is to test the benefit of the targeted therapy against the control is conducted in the overall population which is conducted at a slightly lower significance level α_1 than the overall alpha α . If this hypothesis is rejected, then the targeted therapy is declared superior than the control treatment for

the overall population and analysis is completed. If the first hypothesis is not rejected; then the signature component of the design is used select a potentially promising biomarker sub-group also known as the sensitive subset. The sensitive subset is determined by developing the classifier using the full population. It is done by the following steps:

1. The study population is split into k sub-samples.
2. One of the k sub-samples is omitted to form the training sub-sample. Similar to the adaptive signature design, develop a model to predict the treatment difference between targeted therapy and control as a function of baseline covariates. This is carried out for each subject in this training sub-sample to classify the subject has sensitive or non- sensitive.
3. Repeat the same process leaving out a different sample from the k sub-sample. After k iterations, all the subjects in the full dataset will be classified as sensitive or non-sensitive.
4. Now, compare the treatment difference within the subgroup of patients classified as sensitive using a test statistic (T). Then, generate the null distribution of T by permuting the two treatments and repeating the entire k iterations of the cross-validation process. Perform the test at $\alpha - \alpha_1$. If the test is rejected, then the superiority is claimed for the targeted therapy in the biomarker positive sub-group.

4.2 Adaptive Enrichment Design

Adaptive enrichment designs (e.g., Wang et al., 2007, 2009; Jenkins et al., 2011; Brannath et al., 2009; Friede et al, 2012) were developed for confirmatory pivotal studies with the assumption that the biomarker of interest has already been established and its clear what patients are marker positive and what patients are marker negative so that a prospective design is feasible. An adaptive enrichment design is often a two-stage design. In Stage 1, all patients are enrolled in the trial regardless of biomarker status. An interim analysis is performed to determine whether to continue with all patients or only with biomarker selected population in Stage 2. In Stage 2, a hypothesis test is performed either on the full population, the biomarker-defined sub-population or both. A closed testing procedure with a p-value combination method (Marcus et al, 1976; Bauer and Kieser, 1999) is implemented to control the familywise type I error rate.

For example, Jenkins et al. (2011) proposed an adaptive seamless Phase II/III design with subpopulation selection. It allows the trial to select the appropriate subgroup at the interim analysis or continue in all patients with both the subgroup and the full population as co-primary populations. In this method, a combination of test statistics for the final endpoint from each stage is used for hypothesis testing. The decision to extend to the second stage is based on intermediate or surrogate endpoint correlated to the final endpoint. This method is presented in the context of oncology using the progression free survival (PFS) and overall survival (OS) as interim and final endpoints.

4.3 Basket and Umbrella Trial Design

Increasing knowledge about the genetic causes of disease is prompting intense interest in the concept of precision medicine. This is particularly the case in oncology, which researchers view as the field most advanced with the strategy. The science is prompting researchers to develop treatments that target the mutations regardless of where a patients cancer is located in the body.

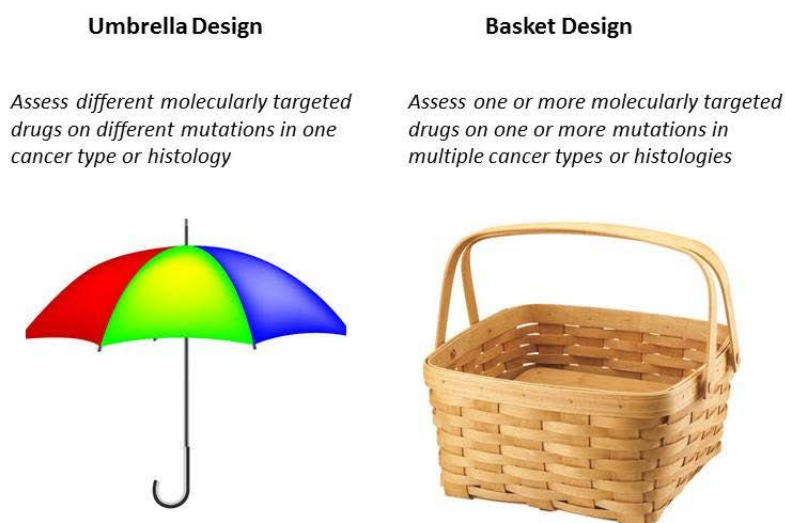


Figure 4: Oncogenic Drivers in Lung Adenocarcinoma.

A key driver of the strategy is the fact that the same cancer-causing molecular traits are often found in a variety of tumor types, raising hope that a drug effective against the target in, say breast cancer, would be effective in a tumor originating in another organ. Indeed, Roche Holding AGs breast-cancer drug Herceptin, which targets a receptor called Her2, turned out to be effective and was eventually approved for gastric tumors that have high levels of Her2. But the drug Zelboraf, which is especially effective against the skin cancer melanoma with a certain mutation in a gene called BRAF, turns out to have essentially no effect against colon cancer harboring the same mutation, raising the issue that it is much more complicated and researchers should have some caution toward broad success in the approach.

Another major issue in the clinical development of precision medicines is that genetically characterizing tumors breaks common cancers such as lung or breast into a dozen or more much rarer diseases. That poses a challenge to drug companies, which in recruiting for a single-drug trial could have to screen as many as 10,000 patients to find enough patients to test a drug against a rare mutation. Screening patients for a trial involving 10 or 20 drugs instead is expected to be much more efficient, and more quickly provide patients with access to potentially beneficial treatments.

Umbrella trial design and basket trial design (Figure 4) are proposed in recent years to meet these challenges and to develop novel targeted therapies in a faster and more efficient manner.

An umbrella trial assesses different molecularly targeted drugs on different mutations in one cancer type of histology. Examples are Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And molecular Analysis 2 (I-SPY TRIAL 2, I-SPY 2, NCT01042379; ref. Barker et al., 2009), the FOCUS4 study in advanced colorectal cancer (Kaplan et al., 2007), and the phase II adaptive randomization design Biomarker-integrated

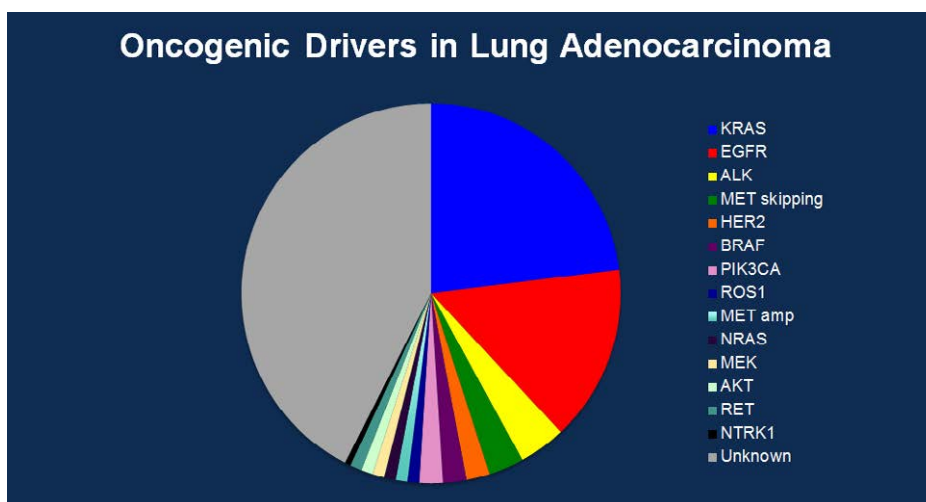


Figure 5: Oncogenic Drivers in Lung Adenocarcinoma.

Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE; ref. Kim et al., 2011) in NSCLC (NCT00409968).

A basket trial assesses one or more molecularly targeted drugs on one or more mutation-s regardless of cancer types or histologies. This design facilitates a particular targeted therapeutic strategy (i.e., inhibition of an oncogenically mutated kinase) across multiple cancer types. Examples are NCI's Molecular Analysis for Therapy Choice (MATCH) and the Molecular Profiling based Assignment of Cancer Therapeutics (MPACT, NCT01827384) trials (Conley et al., 2014). These designs are quite powerful because they can screen and test multiple treatments, multiple biomarkers in multiple indications simultaneously.

5. Case Studies

5.1 Development of Crizotinib in ALK+ NSCLC

Lung cancer is currently the leading cause of cancer death in both men and women. Historically, lung cancer was categorized as 2 types of diseases: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), with NSCLC accounting for about 85% to 90% of lung cancer cases. NSCLC can also be classified according to histological type: adenocarcinoma, squamous-cell carcinoma and large-cell carcinoma. Such classification is important for determining management and predicting outcomes of the disease.

However, with the rapid advance in biological and genetic science in the past 2 decades, researchers find there are various oncogenic drivers behind the progression of lung cancer caused by the inactivation of the so-called tumor suppressor genes. Figure ?? and Table 1 shows the potential oncogenic drivers in NSCLC based on the current knowledge, such as the EGFR mutation and the ALK mutation.

The MTA crizotinib (XALKORI) is a potent, selective, small-molecule competitive inhibitor of anaplastic lymphoma kinase (ALK), MET, and ROS-1 (Christensen et al., 2007; Shaw et al., 2014). The first-in-human phase 1 trial started in December 2015 to estimate the MTD opening to all-comer patients with solid tumors. The EML4-ALK translocation in NSCLC was discovered in 2007. In the same year, the study was amended to add patients with EML4-ALK mutation to the MTD cohort, and the first clinical response was observed in ALK+ tumors in early 2008. Subsequently, the clinical development program progressed rapidly, and crizotinib was approved in 2011 by the FDA for NSCLC that is

Table 1: Oncogenic Drivers in Lung Adenocarcinoma.

Oncogenic drivers	Prevalence
KRAS	20% - 25%
EGFR	13% - 17%
ALK	3% - 7%
MET skipping	~ 3%
HER2	~ 2%
BRAF	~ 2%
PIK3CA	~ 2%
ROS1	~ 1%
MET amp	~ 1%
NRAS	~ 1%
MEK	~ 1%
AKT	~ 1%
RET	~ 1%
NTRK1	~ 0.5%

Table 2: Clinical studies that led to accelerated approval and full approval.

Protocol	Setting	Trial Design	Primary Endpoints
A8081001 (n=119)	All lines, solid tumors, ALK-positive NSCLC	Single-arm, open-label study of crizotinib	Safety, response pharmacokinetics
A8081005 (n=136)	≥ 2nd line ALK-positive NSCLC	Single-arm, open-label study of crizotinib	Safety, response
A8081007 (confirmatory phase 3) (n=318)	2nd line ALK-positive NSCLC	Crizotinib versus (pemetrexed or docetaxel), randomized, open-label study	PFS

ALK-positive as detected by an FDA-approved companion diagnostic test, a commercially available break-apart fluorescence in situ hybridization (FISH) probes for detecting ALK gene rearrangement to detect the rearrangement in NSCLC (Kwak et al., 2010). It took only 6 years from FIH to registration.

Table 2 summarizes the clinical studies and their trial designs and endpoints that led to the accelerated and full approval by the global health authorities. Classical enrichment designs were used for these studies that allowed for the investigation of this novel drug in an efficient and rapid way because patients with ALK mutation only account for approximately 5% of NSCLC population. High response rates (55%-60%) in the two single-arm enrichment studies led to accelerated approval by the FDA. Full approval was granted after positive readout of randomized confirmatory study A8081007.

In the absence of comparative data, it was unclear whether the distinct clinicopathologic characteristics of patients with ALK-positive NSCLC noted above might be contributing to the observed antitumor activity of Crizotinib. Extensive retrospective statistical analyses were conducted using bootstrapping (covariate-matched) and modeling (covariate-adjusted) to simulate outcomes of randomized controlled studies of crizotinib versus standard advanced NSCLC treatment (Selaru et al., 2016). These analyses utilized data from the control arms of three Pfizer-sponsored phase III studies evaluating first-line paclitaxel-carboplatin or gemcitabine-cisplatin and second- or later-line erlotinib regimens in patients with advanced unselected NSCLC. These analyses demonstrated clinically meaningful and statistically significant effect of Crizotinib despite the lack of a concurrent active control

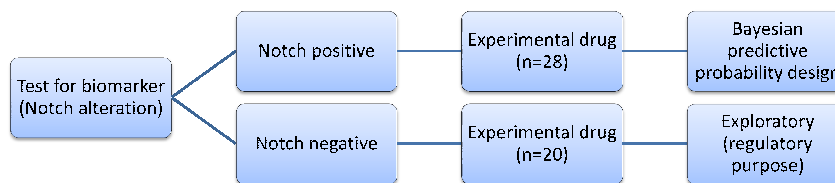


Figure 6: Study Design of an Experimental Notch Inhibitor.

arm.

5.2 Bayesian Predictive Probability for an Enrichment Phase 2 POC Study

Breast cancer is a common type of cancer among women. A diagnosis of triple-negative breast cancer (TNBC) means the 3 most common types of receptors that fuel cancer growth ER, PR and HER2 are not present, which represents 15% of breast cancer patients. In TNBC with Notch genomic alternations (NA+), the inhibition of activation of the Notch pathway using single-agent Notch inhibitor therapy may induce clinical activity (Stylianou et al., 2006). The prevalence of Notch alteration in breast cancer is estimated to be around 10%, so Notch+ TNBC represents only 1-2% of breast cancer, a very rare disease.

This is a phase 2 proof-of-concept (POC) study of an experimental Notch inhibitor, an oral drug given twice daily (BID). The hypothesis is that treatment with this drug response rate can be improved from historical level of $\leq 30\%$ to $\geq 60\%$. However, there are 2 main challenges in designing the trial. First of all, there is no prior clinical data to suggest a high response rate of 60% can be achieved in this rare disease defined by NA+, nor is there any prior data on the analytical validity or clinical utility of the assay. As a result, it is highly desirable to stop the trial early if observed objective response rate (ORR) is low during the trial conduct. Secondly, due to the extremely low prevalence rate of 1% to 2% of the target population in breast cancer, enrollment speed is expected to be slow, albeit 20-25 sites will be opened to screen hundreds of TNBC patients, and the turnaround time of the next generation sequencing assay (2 to 3 weeks) may decrease trial acceptance.

To meet the aforementioned challenges, a Bayesian predictive probability (BPP) design was proposed with multiple interim looks (Lee and Liu, 2008), so that the trial can be stopped early if there is no or low drug effect since it is a costly study with high risk. The Bayesian approach allows greater flexibility in continuously monitoring the trial data to make a go/no-go decision.

Figure 6 illustrates the study design. Patients would be tested for the biomarker status. If its NA positive, patients will be assigned to the experimental drug using the proposed BPP design. It is estimated that at least 28 patients are required to test the hypothesis controlling for the type I and type II error rates. Also 20 NA negative patients would be enrolled to gather some data for exploratory analysis to meet regulatory requirement of health authorities (but not hypothesis to be tested). This is because the treatment effect is expected to be much smaller (if there were any effect) in the marker-negative population, the size of the marker-negative population would usually be too small to give a definitive answer on the effect in that population; But, it would provide at least some estimate of the effect in that population.

The BPP design uses a beta-binomial conjugate distribution for the tumor response rate p . The predictive probability (PP) is the probability of a positive result at the end of the trial, based on the cumulative information in the current stage:

$$PP = \sum_{i=0}^m \{P(Y = i|x) \times I(P(p > p_0|x, Y = i) > \theta_T)\} \quad (1)$$

where x is the current observed number of responses, Y is the number of responses in future patients, m is the total number of patients, θ_T is a prespecified threshold value for the indicator function to determine success at the end of the trial.

During the trial, the PP is compared to some boundary values (θ_L and θ_U) for futility and superiority evaluation.

- If $PP < \theta_L$, stop the trial and reject the alternative.
- If $PP > \theta_U$, stop the trial and reject the null; otherwise continue.

By applying some optimization algorithms, the optimal design that minimizes the maximum sample size can be determined.

With a non-informative prior of beta (0.3, 0.7), it is estimated that 28 patients will be required to have 25 response-evaluable patients so as to control the 1-sided type I error rate at 0.05 with 90% power when the true ORR is 60%. The design has multiple interim looks for potential early stopping, and the decision rules are provided in Table x. At the final analysis, at least 12 responders are required out of 25 evaluable patients to claim the drug efficacious.

6. Conclusion

Clinical researchers face increasing challenges and are presented with unique opportunities in the new era of personalized/precision medicine to meet the needs of patients, payers and the ever increasing costs of health economics. In this paper, we give an overview of population enrichment designs that are growing in the statistical literature, with a focus on designs of predictive enrichment strategies so as to increase the power and efficiency to detect an effective therapy with regard to a predictive biomarker in a clinical trial. The existing designs are categorized into classical designs and novel designs. Two case studies are presented to illustrate the benefits of using an enrichment strategy in oncology drug development. Our focus is on designs that involve 1 or multiple predictive biomarkers that correlate with clinical efficacy endpoints of interest and the experimental treatment. The case of biomarkers that relate to safety is also an important research area of interest and is not covered in this paper.

REFERENCES

- Mandrekar, S.J, and Sargent, D. (2009). Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *Journal of Clinical Oncology*, 27(24): 4027-4034.
- Freidlin, B., McShane, L.M., and Korn, E.L. (2010). Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst*, 102(3):152-160.
- Sargent, D., Conley, B.A., Allegra, C., and Collette, L. (2005). Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol*, 23(9): 2020-2227.
- Cobo, M., Isla, D., Massuti, B., et al. (2007). Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: a phase III trial in non-small-cell lung cancer. *J Clin Oncol*, 25(19): 2747-2754.

- Freidlin, B., and Simon, R. (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res*, 11(21):7872-7878.
- Jiang, W., Freidlin, B., and Simon, R. (2007). Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst*, 99(13):1036-1043.
- Freidlin, B., Jiang, W., and Simon, R. (2010). The cross-validated adaptive signature design. *Clinical Cancer Research*, 16(2): 691-698.
- Barker, A.D., Sigman, C.C., Kelloff, G.J., Hylton, N.M., Berry, D.A., and Esserman, L.J. (2009). ISPY2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharm Ther*, 86:97100.
- Kaplan, R., Maughan, T., Crook, A., Fisher, D., Wilson, R., Brown, L., et al. (2007). Evaluating many treatments and biomarkers in oncology: a new design. *J Clin Oncol*, 31:45628.
- Kim, E.S., Herbst, R.S., Wistuba, II., and Lee, J. (2011). Blumenschein GR Jr, Tsao A, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov*, 1:4453.
- Conley, B.A., and Doroshow, J.H. (2014). Molecular analysis for therapy choice: NCI MATCH. *Semin Oncol*, 41:2979.
- Wang, S.J., O'Neill, R., and Hung, J. (2007). Approaches to Evaluation of Treatment Effect in Randomised Clinical Trials with Genomic Subset. *Pharmaceutical Statistics*, 6(3), 227-244.
- Wang, S.J., Hung, J., and O'Neill, R. (2009). Adaptive patient enrichment designs in therapeutic trials. *Biometrical Journal*, 51: 358374.
- Jenkins, M., Stone, A., and Jennison, C. (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics*, 10: 347356.
- Brannath, W., Zuber, E., Branson, M., Bretz, F., Gallo, P., Posch, M., and Racine-Poon, A. (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine*, 28: 14451463.
- Friede, T., Parsons, N., and Stallard, N. (2012). A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine*, 31: 43094320.
- Marcus, R., Eric, P. and Gabriel, K.R., (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3), pp.655-660.
- Bauer, P., and Kohne, K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics*, 50 (4):10291041. DOI: 10.2307/2533441.
- Christensen, J.G. et al. (2007). Cytoreductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic large-cell lymphoma. *Mol Cancer Ther*, 6, 33143322.
- Shaw, A.T. et al. (2014). Crizotinib in ROS-1 rearranged non-small-cell lung cancer. *N Engl J Med*, 371, 1963-1971.
- Kwak, E.L. et al. (2010). Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*, 363, 16931703.
- Selaru, P., Tang, Y., Huang, B., Polli, A., Wilner, K., Donnelly, E., and Cohen, D. (2016). Sufficiency of Single-Arm Studies to Support Registration of Targeted Agents in Molecularly Selected Patients with Cancer: Lessons from the Clinical Development of Crizotinib. *Clinical and translational science*.
- Stylianou, S., Clarke, R., and Brennan, K. (2006). Aberrant activation of notch signaling in human breast cancer. *Cancer research*, 66(3), 1517-1525.
- Lee, J., Liu, D. (2008). A predictive probability design for phase II cancer clinical trials. *Clinical Trials*, 5(2), 93-106.