Research Article

Statistics in Medicine

Received 23 February 2009, Accepted 4 December 2009 Published online in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3840

The statistics of phase 0 trials

Larry V. Rubinstein,[a]*† Seth M. Steinberg,[b] Shivaani Kummar,[b] Robert Kinders,[c] Ralph E. Parchment,[c] Anthony J. Murgo,[d] Joseph E. Tomaszewski[a] and James H. Doroshow[a,b]

The PD-driven phase 0 trial is a new form, designed to be a first-in-man study, often of a new agent, conducted to assess drug effect on a molecular target, by means of a pharmacodynamic (PD) assay, in a very small number (10–15) of patients. Such a study is meant to be a proof of principle trial to determine whether the agent yields the PD effect predicted by pre-clinical studies. The dosage is meant to be pharmacologically active, but is neither toxic nor likely to yield clinical benefit. Such a trial may be used to serve as a very early test of an agent's biologic effect, allowing for early weeding out of ineffective agents, or as an early means of determining the most promising of competing analogue agents. This manuscript will present designs for such PD-driven studies that are statistically efficient and rigorous, focusing on non-comparative trials. The phase 0 trial promises to become an increasingly important tool for facilitating and speeding the development of new therapeutic agents, particularly in oncology. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: first-in-man study; PD-driven phase 0 trial; pharmacodynamic assay; phase 1 clinical trial; phase 2 clinical trial

1. Introduction and statement of the concept—measuring biological effectiveness (with a PD endpoint) as a very early screening and drug development tool

Currently, only 10 per cent of investigational new drug (IND) applications to the Food and Drug Administration (FDA) result in clinically approved agents, and in oncology it is only 5 per cent [1, 2]. This is a very serious problem, since the development of a new agent is a lengthy and expensive process and many of these agents fail relatively late in that process. The fact that an increasing proportion of IND agents is molecularly targeted, which suggests testing the agent for effectiveness against the target by means of a PD assay very early in the drug development process. This is particularly useful and important since the pre-clinical tests of such effectiveness are often misleading, yielding both false positive and false negative results. For this reason, the FDA issued a new Exploratory IND Guidance in 2006, to allow for such studies as small first-in-man trials, conducted at dose levels and administration schedules not expected to result in significant clinical toxicity, and generally restricted to at most approximately one week per patient [1, 2]. Conducting studies under this guidance requires substantially less pre-clinical toxicology work than that required for standard IND phase 1 studies [1, 2]. Therefore, phase 0 studies can be administered while the toxicology studies preparatory to filing a standard IND are being conducted, and they will not postpone the time until the phase 1 trial can be initiated.

Phase 0 studies can be very effective tools for determining very early in the drug development process whether an agent potentially has the anticipated biologic effect. They can also be used to prioritize among analogs or agents designed to have the same molecular target by means of comparing pharmacokinetic (for example, oral bioavailability) and/or PD characteristics (although we will not deal explicitly with such comparative designs in this paper). They are an opportunity for developing and validating clinical PD assays very early in the drug development process, to enable more reliable usage of such assays in phase 1 and phase 2 trials [3]. Finally, they can contribute to better defining the appropriate dose range or administration schedule to take into phase 1 and phase 2 testing by determining a minimum dose necessary to achieve a biological effect and by providing pharmacokinetic and PD data taken over time.

[a] *Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD 20892, U.S.A.*

[b] *Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, U.S.A.*

[c] *Laboratory of Human Toxicology and Pharmacology, Applied/Developmental Research Support Directorate, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, U.S.A.*

[d] *Food and Drug Administration, Office of Oncology Drug Products, Silver Spring, MD, U.S.A.*

*Correspondence to: Larry V. Rubinstein, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD 20892, U.S.A.

†E-mail: rubinsteinl@ctep.nci.nih.gov

2. Statistical design of a phase 0 trial

The challenge presented by the PD-driven phase 0 study is to assess the change in the PD endpoint effected by the agent, with very few patients, each treated over a short period of time, but maintaining a certain amount of statistical rigor. Kummar *et al.* [1] and Murgo *et al.* [2] suggest several statistical designs to address this challenge in different clinical contexts, three of which we present here. Typically, a phase 0 trial will encompass several escalating dose levels for the experimental agent. In general, the approach taken is to mimic the design of a phase 2 study [4], and to design the phase 0 study as a phase 2 study in miniature for each separate dose level. Thus, the first step is to define what is meant by a PD 'response' for each individual patient, which is analogous to defining what constitutes an objective tumor response for a patient in a phase 2 trial. The second step is to define what constitutes a promising observed PD response rate for each dose level—in other words, how many patients must demonstrate a PD response for the dose level to be declared biologically effective. This is analogous to setting a threshold for observed response rate in a phase 2 trial, in order that the agent be deemed sufficiently promising for further testing [4]. Further details of the approach given in Kummar *et al.* [1] and Murgo *et al.* [2] are given in the following sections.

2.1. Determining statistical significance of a PD effect at the patient level—defining a PD response

In oncology, generally, the PD endpoint is assessed both in tumor tissue and in an easily assayed surrogate tissue such as blood (peripheral blood mononuclear cells (PBMCs)). The tumor tissue assay is considered to be more reliable with respect to reflecting the biological effect of the agent in what is generally the target tissue of interest [1, 3]. However, the number of tumor biopsies usually is severely limited for ethical reasons [1, 2]. Therefore, the PBMC assay, for example, is used as a surrogate, since multiple PBMC assays can be performed both in pre-treatment and post-treatment, thus allowing for the assessment of both the pre-treatment variability at the patient level and the post-treatment PD effect over time [1–3]. Generally, there are only two tumor biopsies, one taken shortly before treatment with the agent, and one taken at the post-treatment time point of greatest interest, often when the PD effect is anticipated to be at its maximum. The measure of treatment effect for the tumor PD assay is the difference between the pre-treatment and the post-treatment values (often measured on the log scale rather than on the original). Generally, there are multiple pre-treatment PBMC assays for both pre-treatment and post-treatment. The primary measure of treatment effect for the PBMC assay is the one that corresponds in time to that of the tumor assay—the difference between the most immediate pre-treatment PBMC assay and the post-treatment PBMC assay closest in time to that of the tumor biopsy. The other pre-treatment PBMC assays should, ideally, cover a time span comparable to that of the pre-treatment versus post-treatment biopsies. In that way, they provide a measure of the natural variation of the assay, for an individual patient, over that time span. The other post-treatment PBMC assays provide a means of assessing the post-treatment PD effect over time, as a secondary set of PD endpoints.

Defining a PD 'response', both for the tumor assay and for the PBMC assay, usually involves both a biologic criterion and a statistical criterion for what is significant. The biologic criterion generally depends upon characteristics of the biologic target of the agent. For example, in the recent National Cancer Institute (NCI) phase 0 trial of ABT-888 [5, 6], the criterion chosen was that the reduction in the assay value had to be at least 50 per cent. The statistical criterion may be either 90 per cent confidence or 95 per cent confidence (generally one-sided, since the anticipated treatment effect is generally in one direction) that the observed treatment effect is not a result of the sort of natural random variation in the assay, for an individual patient, which would be seen in the absence of a true treatment effect. (In this respect, there is an opportunity to employ more statistical rigor than that is typical for the analogous phase 2 study design.) For the PBMC assay, this natural variation can be assessed by the pooled intra-patient standard deviation (SD) of the pre-treatment values. However, for the tumor assay, multiple pre-treatment assays per patient will generally not be available. Therefore, the inter-patient SD of the pre-treatment values must be used instead. Details concerning the definition of a PD response are illustrated in Figure 1. The thresholds for declaring the PD effect (pre-treatment

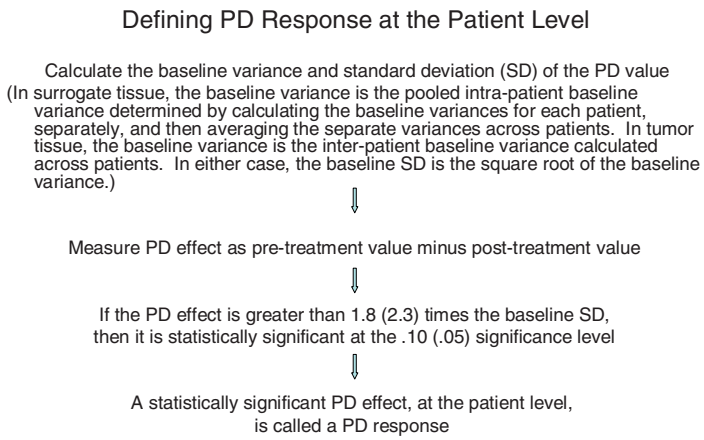

Figure 1. This figure illustrates the defining of PD 'response' for an individual patient. Multipliers of the baseline SD are derived from asymptotic normal distribution theory. Significance levels are one-sided.

value minus post-treatment value, for the case where the agent is anticipated to reduce the assay value, as in the NCI phase 0 trial [5, 6]) statistically significant (at the one-sided 0.10 or 0.05 significance levels) are calculated from the variance of the difference of two normally distributed variables. (If the number of samples from which to estimate the pre-treatment variability of the assay is very limited (under 20) consideration should be given to using *t*-distribution, rather than normal distribution, cut-off values.)

2.2. Determining statistical significance of a PD effect for a given dose level

For each dose level, the investigators may set a threshold for the number of patients, among the total, that must demonstrate a PD response, in order for the dose level to be judged as yielding a promising biologic effect. As the false positive rate for a PD response, for an individual patient, has been determined (as given above), the false positive rate for declaring a dose level effective, for each assay separately and for the two combined, can be calculated from the binomial distribution. Similarly, for a targeted PD response rate across patients, the power to declare the dose level effective, for each of the two assays, can be calculated. The investigators may employ a one-stage or two-stage design to assess the PD response rate at each dose level, just as in phase 2 studies [4], and the calculations of power and false positive rate are done in an identical fashion. Examples are given below of designs to target 80, 60, or 40 per cent PD response rates across patients.

2.3. Three trial designs—designs to detect an 80, 60 or 40 per cent PD response rate across patients

To target a true 80 per cent PD response rate at each dose level, a one-stage design may be used. Three patients are treated and the dose level is declared effective with respect to either PD assay if at least two of the patients demonstrate a PD response which is significant at the 0.10 level. This design yields 90 per cent power to detect a true 80 per cent PD response rate, across patients, for either assay, with an overall 6 per cent false positive rate for the two assays combined, under the null hypothesis that the agent has no biologic effect. This is the design that was used in the NCI phase 0 trial of ABT-888 [5, 6], and it is illustrated in Figure 2.

To target a true 60 per cent PD response rate at each dose level, a two-stage design may be used. Three patients are treated and the cohort is expanded to five patients if exactly one patient, for either PD assay, demonstrates a PD response which is significant at the 0.05 level. The dose level is declared effective with respect to either PD assay if at least two of the patients demonstrate a PD response which is significant at the 0.05 level. This design yields 89 per cent power to detect a true 60 per cent PD response rate, across patients, for either assay, with an overall 4 per cent false positive rate for the two assays combined, under the null hypothesis that the agent has no biologic effect. This design is illustrated in Figure 3.

To target a true 40 per cent PD response rate at each dose level, a similar two-stage design may be used. Five patients are treated and the cohort is expanded to eight patients if exactly one patient, for either PD assay, demonstrates a PD response which is significant at the 0.05 level. The dose level is declared effective with respect to either PD assay if at least two of the patients demonstrate a PD response which is significant at the 0.05 level. This design yields 87 per cent power to detect a true 40 per cent PD response rate, across patients, for either assay, with an overall 10 per cent false positive rate for the two assays combined, under the null hypothesis that the agent has no biologic effect. This design is illustrated in Figure 4.

3. Example of a phase 0 trial—the NCI PARP inhibitor trial—and further discussion of phase 0 statistical issues

The NCI selected ABT-888, an inhibitor of the DNA repair enzyme poly (ADP-ribose) polymerase (PARP), for the first ever phase 0 trial for two reasons [5, 6]. First, it was anticipated to have a wide margin of safety relative to target modulating doses in pre-clinical models. This is an essential characteristic for a phase 0 agent. Phase 0 trials cannot promise any benefit for the patients who participate, so there must be reasonable assurance that toxicity will be minimal. Second, it was anticipated to have a wide therapeutic applicability if demonstrated effective. Elevated PARP levels are characteristic of tumors and can result in resistance to both chemotherapy (CT) and radiotherapy (RT). Therefore, PARP inhibitors hold promise of wide applicability as CT and RT sensitizers. The NCI trial demonstrated statistically significant reduction in the PAR levels (a surrogate for PARP inhibition) in both tumor and PBMCs [5, 6]. It also gave an opportunity to explore the correlation between blood and tissue marker levels, to determine to what extent blood levels could be used as a surrogate for the more difficult to obtain tissue levels. It is interesting to note that biological activity was seen at the second dose level tested. In the future, investigators may not be so fortunate—dose escalation requires careful planning, and, potentially, pre-clinical work to assure that an excessive number of dose levels are not tested, interfering with the goal of keeping small the sample size of the phase 0 trial.

There are a number of statistical issues relating to phase 0 trials that deserve further mention:

(1) In the NCI phase 0 trial it was found that the variance of the pre-treatment PD assay values was reduced if the logs of the values were used instead. It is often appropriate to log-transform the PD assay values since geometric, rather than arithmetic, changes in value are thought to be qualitatively similar along the assay scale.
(2) It will often be the case that assessing the PD treatment effect can be done with greater statistical power if the mean effect is measured across patients and then a test applied to the null hypothesis that the mean effect is equal to 0. Analogously, there have been proposals that phase 2 trials be assessed by testing whether the mean tumor shrinkage is statistically

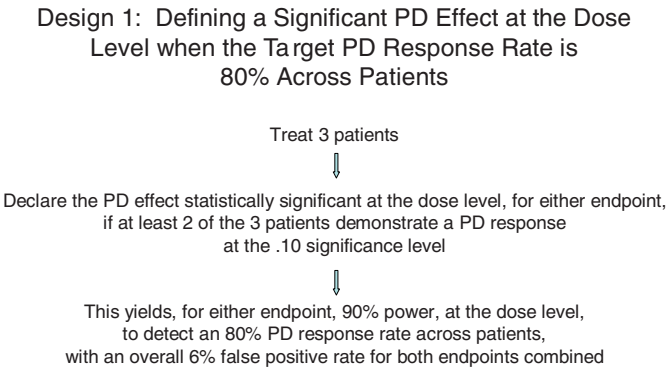

Figure 2. This figure illustrates the defining of what constitutes a promising observed response rate for a dose level. The target (true) PD response rate, across patients, is 80 per cent. Power and false positive rate are derived from the binomial distribution.

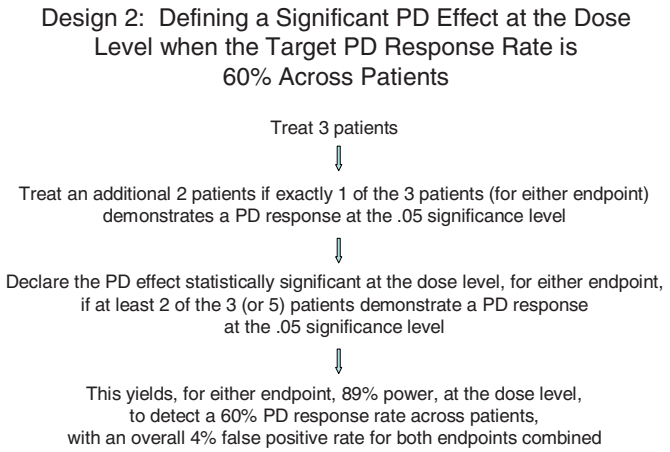

Figure 3. This figure illustrates the defining of what constitutes a promising observed response rate for a dose level with a two-stage design. The target (true) PD response rate, across patients, is 60 per cent. Power and false positive rate are derived from the binomial distribution.

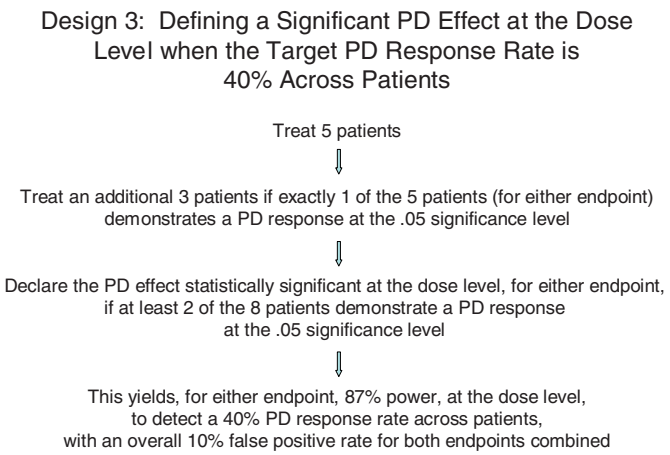

Figure 4. This figure illustrates the defining of what constitutes a promising observed response rate for a dose level with a two-stage design. The target (true) PD response rate, across patients, is 40 per cent. Power and false positive rate are derived from the binomial distribution.

significant [7]. The problem with this approach is that a statistically significant mean treatment effect does not necessarily imply a biologically relevant treatment effect for a meaningful proportion of the patients [8]. For this reason, the NCI phase 0 trial investigators chose to impose the additional criterion of a biologically relevant level of PAR reduction for the individual patients. Likewise, it was felt appropriate to determine, for the individual patients, whether the PAR reduction observed was statistically significant. This follows the standard phase 2 model of determining what would constitute a response, for the individual patient, suggestive of benefit for that patient, and then assess the proportion of patients demonstrating such a

response [8]. There may be phase 0 situations where this approach is too statistically demanding, and it is appropriate to resort to assessing the mean treatment PD effect.

(3) For the tumor biopsy assay, multiple pre-treatment assays per patient will generally not be available for ethical reasons. Therefore, the inter-patient SD of the pre-treatment values must be used instead of the intra-patient SD, which cannot be determined. The inter-patient variability will often be substantially greater than the intra-patient variability. This can seriously limit the ability to declare statistically significant an observed treatment effect measured by the tumor assay. For example, in the NCI phase 0 trial, an observed 95 per cent post-treatment reduction in the tumor assay value was required for statistical significance, whereas an observed 55 per cent post-treatment reduction was sufficient for the PBMC assay [2].

4. Conclusions

The fact that an increasing proportion of IND agents is molecularly targeted, as opposed to being generally cytotoxic (to rapidly growing cells, in particular) provides the new opportunity of testing the agent for effectiveness against the target by means of a PD assay (which may require development in conjunction with the phase 0 trial) very early in the drug development process. Phase 0 trials provide an excellent opportunity to establish feasibility and further refine target or biomarker assay methodology in a limited number of human samples before initiating larger trials involving patients receiving toxic doses of the study agent. We have demonstrated that, despite the small sample size, the nature of the PD assay values allows for a reasonable degree of statistical rigor and, especially in the case of the surrogate assays (which can be repeated multiple times), a reasonable degree of statistical power.

As far as we know, the NCI phase 0 trial is the first and the only one completed so far. It received favorable editorial notice [9, 10] and we expect more such trials to follow soon. We feel that phase 0 trials will have broad applicability. Also, in rare diseases, where the molecular or genetic pathway may be known, we feel that this may prove an invaluable tool to enable early proof-of-principle testing where the available patient population is severely limited.

Phase 0 trials do not replace phase 1 trials conducted to establish dose-limiting toxicities and define a recommended phase 2 dose. On the other hand, data from phase 0 trials allow phase 1 studies to begin at a higher, potentially more efficacious dose, use a more limited and rationally focused schedule for PD sampling, and use a qualified PD analytic assay for assessing target modulation. Likewise, phase 0 trials, with PD endpoints, will not eliminate the need for phase 2 trials to establish the agent's ability to yield tumor response or clinical benefit, but they will allow for early termination of development of agents that fail to yield the anticipated biologic effect. Therefore, the effort expended to conduct rationally designed phase 0 trials should conserve resources in the long run by improving the efficiency and success of subsequent clinical development.

References

1. Kummar S, Kinders R, Rubinstein L, Parchment RE, Murgo AJ, Collins J, Pickeral O, Low J, Steinberg SM, Gutierrez M, Yang S, Helman L, Wiltrout R, Tomaszewski JE, Doroshow JH. Compressing drug development timelines in oncology using 'phase 0' trials. *Nature Reviews/Cancer* 2007; **7**:131–139.
2. Murgo AJ, Kummar S, Rubinstein L, Gutierrez M, Collins J, Kinders R, Parchment RE, Ji J, Steinberg SM, Yang SX, Hollingshead M, Chen A, Helman L, Wiltrout R, Tomaszewski JE, Doroshow JH. Designing phase 0 cancer clinical trials. *Clinical Cancer Research* 2008; **14**:3675–3682.
3. Kinders R, Parchment RE, Ji J, Kummar S, Murgo AJ, Gutierrez M, Collins J, Rubinstein L, Pickeral O, Steinberg SM, Yang S, Hollingshead M, Chen A, Helman L, Wiltrout R, Simpson M, Tomaszewski JE, Doroshow JH. Phase 0 clinical trials in cancer drug development: from FDA guidance to clinical practice. *Molecular Interventions* 2007; **7**:325–334.
4. Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 1989; **10**:1–10.
5. Kummar S, Gutierrez M, Kinders R, Rubinstein L, Parchment RE, Philips L, Ji J, Chen A, Horneffer Y, Juwara L, Yancey M, Murgo AJ, Collins JM, Tomaszewski JE, Doroshow JH. Phase 0 pharmacodynamic, pharmacokinetic study of ABT-888, an inhibitor of poly (ADP-ribose) polymerase (PARP), in patients with advanced malignancies. *Annals of Oncology* 2008; **19**(Suppl. 3):20.
6. Kummar S, Kinders R, Gutierrez M, Rubinstein L, Parchment RE, Phillips LR, Ji J, Monks A, Pickeral O, DiPaolo T, Low J, Chen A, Murgo AJ, Collins J, Simpson M, Steinberg S, Yang SX, Horneffer Y, Juwara L, Yancey MA, Hollingshead M, Helman L, Wiltrout R, Tomaszewski JE, Doroshow JH, the NCI Phase 0 Working Group. Phase 0 clinical trial of the poly (ADP-ribose) polymerase (PARP) inhibitor ABT-888 in patients with advanced malignancies. *Journal of Clinical Oncology* 2009; **27**:2705–2711.
7. Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non-small-cell lung cancer. *Journal of the National Cancer Institute* 2007; **99**:1455–1461.
8. Rubinstein L, Dancey J, Korn E, Smith M, Wright J. Early average change in tumor size in a phase 2 trial: efficient endpoint or false promise. *Journal of the National Cancer Institute* 2007; **99**:1422–1423.
9. LoRusso PM. Phase 0 clinical trials: an answer to drug development stagnation? *Journal of Clinical Oncology* 2009; **27**:2586–2588.
10. Editors. Phase 0 trials: a platform for drug development? *The Lancet* 2009; **374**:176.