Issues in Cancer Immunotherapy Trial Design

Richard Simon, D.Sc.
Chief, Computational & Systems Biology Branch
Division of Cancer Treatment & Diagnosis
National Cancer Institute
Immunotherapy is Different
→ Inadequate pre-clinical models
→ More complex mechanism of action
→ Inaplicability of Phase I 3+3 design
→ Necessity of combination regimens
→ Numerous candidate immunomodulating agents
Phase I Design

→ BAD of combination, not MTD of single agent
→ Single agent 3+3 design for MTD generally not needed
Is the "3-3" Dose-Escalation Phase I Clinical Trial Design Suitable for Therapeutic Cancer Vaccine Development? A Recommendation for Alternative Design

Osama E. Rahma1,2, Emily Garnham1, Richard M. Simon3, and Samir N. Khalef1,4

Abstract

Purpose: Phase I clinical trials are generally conducted to identify the maximum tolerated dose (MTD) or the biologically active dose (BAD) using a traditional dose-escalation design. This design may not be applied to cancer vaccines, given their unique mechanisms of action. The FDA recently published "Guidance for Industry: Clinical Considerations for Therapeutic Cancer Vaccines." However, many questions about the design of cancer vaccine studies remain unanswered.

Experimental Design: We analyzed the toxicity profile in 239 phase I therapeutic cancer vaccine trials. We addressed the ability of dose escalation to determine the MTD or the BAD in trials that used a dose-escalation design.

Results: The rate of grade 3/4 vaccine-related systemic toxicities was 1.25 adverse events per 100 patients and 2 per 1,000 vaccines. Only two of the 127 dose-escalation trials reported vaccine-related dose-limiting toxicities, both of which used bacterial vector vaccines. Out of the 116 trials analyzed for the dose-immune response relationship, we found a statistically significant dose-immune response correlation only when the immune response was measured by antibodies (P < 0.001) or type II hyperglycemia (P = 0.05). However, the increase in cellular immune response did not appear further sustainable with the continued increase in dose.

Conclusions: Our analysis suggests that the risks of serious toxicities with therapeutic cancer vaccines are extremely low and that toxicities do not correlate with dose levels. Accordingly, the conventional dose-escalation design is not suitable for cancer vaccines with few exceptions. Here, we propose an alternative design for therapeutic cancer vaccine development. Clin Cancer Res. 20:458-67. ©2014 AACR.
Alternative clinical trial design for cancer vaccine

Step 1. Determining a starting dose of a vaccine

| Vaccine class that is used before and found to be toxic (e.g., bacterial vector) | Proceed to traditional phase I trial |
| Vaccine class that is used before and found to be non-toxic (e.g., peptide) | Use IAD from previous clinical trials |
| Vaccine class that is not used before and not expected to be toxic | OPED |
| | • One patient per tested dose is treated until an immune response is induced (IAD). |
| | • Then, expand that dose level, one patient at a time, until achieving an additional immune response. |
| | • If no additional immune response in 7 patients, stop adding patients and continue escalation of one patient at a time. |

Step 2. Combination design “Vaccine + X” (X is an immune modulator, chemotherapy, or targeted agent)

| X had no DLT | X had a DLT | X’ DLT is unknown |
| Use the same dose | Use the dose below MTD | Proceed to traditional phase I |
- Hypothesis testing phase II
  - Type I and type II errors
  - Clinical endpoint
- Discovery phase II
  - Type III error
  - Clinical endpoint and
  - Intensive immune surveillance with single cell technologies
Type III Error

The error you make by not studying an intervention

Large treatment effects are observable with small sample sizes if you do the study
Need for companies to make immune modulating agents available to academics for discovery studies, particularly with novel combinations
Phase II designs of combinations

Sequential design: \{\text{anti-PD-1}\} \rightarrow \{\text{anti-PD-1 + X}\}
Phase II designs of combinations

Randomized selection design in patients resistant to \{anti-PD-1\}:

\{anti-PD-1 + X\} vs \{anti-PD-1 + Y\} vs \{anti-PD-1 + Z\}
Randomized factorial selection design:

\{\text{antiPD1} + X\} \text{ vs } \{\text{antiPD1} + Y\} \text{ vs } \{\text{antiPD1} + X+Y\} \text{ vs } \{\text{antiPD1}\}

Analyzed as separate two group comparisons:

\{\text{anti PD1} +- Y\} \text{ vs } \{\text{anti PD1} + X +- Y\}
Phase II design for discovery should not be driven by regulatory considerations.
Endpoints for phase IIb studies

Best response over 6 months of treatment
Endpoints for phase III studies

OS or Durable response over 6 months of treatment
Figure 1. Kaplan–Meier Curves for Overall Survival and Progression-free Survival in the Intention-to-Treat Population.

The median follow-up for overall survival (Panel A) in the ipilimumab (ipil)-plus-glycoprotein 100 (gp100) group was 21.0 months, and the median overall survival was 10.0 months (95% CI, 8.5 to 11.5); in the ipilimumab-alone group, the median follow-up was 27.8 months, and the median overall survival, 10.1 months (95% CI, 8.5 to 13.8); and in the gp100-alone group, the median follow-up was 17.2 months, and the median overall survival, 6.4 months (95% CI, 5.5 to 8.7). The median progression-free survival (Panel B) was 2.76 months (95% CI, 2.73 to 2.79) in the ipilimumab-plus-gp100 group, 2.86 months (95% CI, 2.76 to 3.02) in the ipilimumab-alone group, and 2.76 months (95% CI, 2.73 to 2.83) in the gp100-alone group. The rates of progression-free survival at week 12 were 49.3% (95% CI, 44.1 to 53.9) in the ipilimumab-plus-gp100 group, 57.7% (95% CI, 48.9 to 65.5) in the ipilimumab-alone group, and 48.5% (95% CI, 39.6 to 56.7) in the gp100-alone group.
Figure 1. Survival End Points.
Panel A shows the Kaplan–Meier curves for overall survival. The median follow-up for overall survival was 8.9 months in the nivolumab group and 6.8 months in the dacarbazine group. Panel B shows the Kaplan–Meier curves for progression-free survival.
**Run-In Phase III Trial Design With Pharmacodynamics Predictive Biomarkers**

Fangxin Hong, Richard Simon

**Background**

Developments in biotechnology have stimulated the use of predictive biomarkers to identify patients who are likely to benefit from a targeted therapy. Several randomized phase III designs have been introduced for development of a targeted therapy using a diagnostic test. Most such designs require biomarkers measured before treatment. In many cases, it has been very difficult to identify such biomarkers. Promising candidate biomarkers can sometimes be effectively measured after a short run-in period on the new treatment.

**Methods**

We introduce a new design for phase III trials with a candidate predictive pharmacodynamic biomarker measured after a short run-in period. Depending on the therapy and the biomarker performance, the trial would either randomize all patients but perform a separate analysis on the biomarker-positive patients or only randomize marker-positive patients after the run-in period. We evaluate the proposed design compared with the conventional phase III design and discuss how to design a run-in trial based on phase II studies.

**Results**

The proposed design achieves a major sample size reduction compared with the conventional randomized phase III design in many cases when the biomarker has good sensitivity (≥0.7) and specificity (≥0.7). This requires that the biomarker be measured accurately and be indicative of drug activity. However, the proposed design loses some of its advantage when the proportion of potential responders is large (>50%) or the effect on survival from biomarker positivity is small.

**Conclusions**

Incorporating a pharmacodynamic biomarker requires careful consideration but can expand the capacity of clinical trials to personalize treatment decisions and enhance therapeutics development.

**Correspondence to:** Richard Simon, PhD, Biometric Research Branch, National Cancer Institute, Bethesda, MD 20892; e-mail: rsimon@mail.nih.gov
Screen patients using broad eligibility

Run-in on test rx

Evaluate "response"

Off study

no response

response

Randomize

Stop test rx

Continue test rx
Screen patients using broad eligibility

Run-in on test rx

Evaluate "response"

response → Randomize

Off study → no response

Stop test rx  Continue test rx

Stop test rx  Continue test rx
Figure 1. The schema of run-in design with three motivating examples: 1) immunological biomarker in vaccine therapy, 2) imaging biomarkers for early response, and 3) mechanistic markers for drug resistance. PD = pharmacodynamics.

Figure 2. With sample sizes that give 80% power for the standard design, the trial-level power with the run-in design (solid lines) is shown when randomizing all patients, for a series of sensitivity and specificity of the biomarker, under 25%, 50%, and 75% prevalence of true responders, with no run-in effect ($\theta_0 = 1$).

![Graphs showing power in different conditions](image)

By guest on February 6, 2016
http://jnci.oxfordjournals.org/
Downloaded from
Tests of treatment effect in all patients and in the subset with biomarker greater than an optimized threshold can be performed using strategies previously introduced for the adaptive threshold design (19). Such a design enables the threshold to be optimized using phase III data while fully controlling the study-wise type I error, providing rigorous statistical testing of treatment effects overall and in a subset using an optimized cut point for marker positivity.

In some cases, there might be a minimal survival benefit from the run-in period on R+ patients randomized to control. The hazard ratio between control and experimental arms would be reduced, which leads to efficiency loss for the run-in design (Supplementary Figures 1 and 2, available online), particularly with high R+ prevalence. Therefore, we do not recommend a run-in design when a large run-in effect and large R+ proportion (>50%) might be expected. One might also be concerned about a detrimental effect of the run-in for R− patients because it might delay their starting the control regimen. When there is highly effective standard treatment, however, the test regimen is likely to involve the standard treatment plus new drug. When the control regimen is not highly effective, the delay is of less concern and the run-in might use the new drug alone if that provided the most informative biomarker assessment.

Although run-in designs have rarely been used for phase III oncology trials, the randomized discontinuation design is a phase II design that has been used (16). In randomized discontinuation design, all patients receive the new drug, and patients with early objective response or early progressive disease are excluded from randomization. The design can be more or less efficient than the standard design depending on the mechanism of action of the drug (20). The run-in design uses an unvalidated marker measurement and either randomizes all patients or only marker-positive patients; thus, it can potentially provide a much greater improvement in power than randomized discontinuation design. The proposed run-in designs use an intermediate (posttreatment) measurement as a predictive biomarker. However, the marker will not be used as a surrogate endpoint nor necessarily be a response marker; the biomarker is only being used to focus the treatment comparison on the subset of M+ patients. Although early response is one type of posttreatment predictive biomarker, it is not used as a surrogate for making claims about treatment effectiveness. For phase III run-in trials used for drug registration, if the trial demonstrates benefit of the new regimen only in M+ patients or if the M− patients are not randomized, the intended use of the drug would be for patients who were determined to be M+ based on the prespecified cut point after the defined run-in period. And the label should indicate that the estimated improvement in survival time is measured from the end of the short run-in period.

Our proposed design is not without limitations. It adds substantial complexity and cost to the clinical trial and requires substantial credentials for the marker. It depends on the development of an analytically validated assay for accurate measurement of the candidate marker before the initiation of the phase III clinical trial. It loses some efficacy when a large run-in effect and large R+ proportion (>50%) are expected. It would involve novel drug labeling when used for drug registration.

Advances in molecular and imaging technology provide substantial opportunities for development of biomarkers that measure early treatment effect. Early immunologic response, Figure 3. Average simulated number of events needed to have 80% trial-level power in the standard design (dotted line) and then run-in design when randomizing all patients (solid lines, top panels) or randomizing only M+ patients (solid lines, bottom panels), by sensitivity and specificity of the marker, under 25%, 50%, and 75% true responders, with no run-in effect (θ0 = 1). The standard design needs 1440, 420, and 210 events to have 80% power under 25%, 50%, and 75% true responders, respectively.
Need for public support for

Experimental/computational initiative for identifying new
antigenic targets for adoptive T cell therapy

Integrated experimental/computational systems biology
modeling of T cell effectors and their relationship to tumor
and immune system
Acknowledgements

Fangxin Hong, Dana-Farber Cancer Center

Samir Khlief, Georgia Regents University