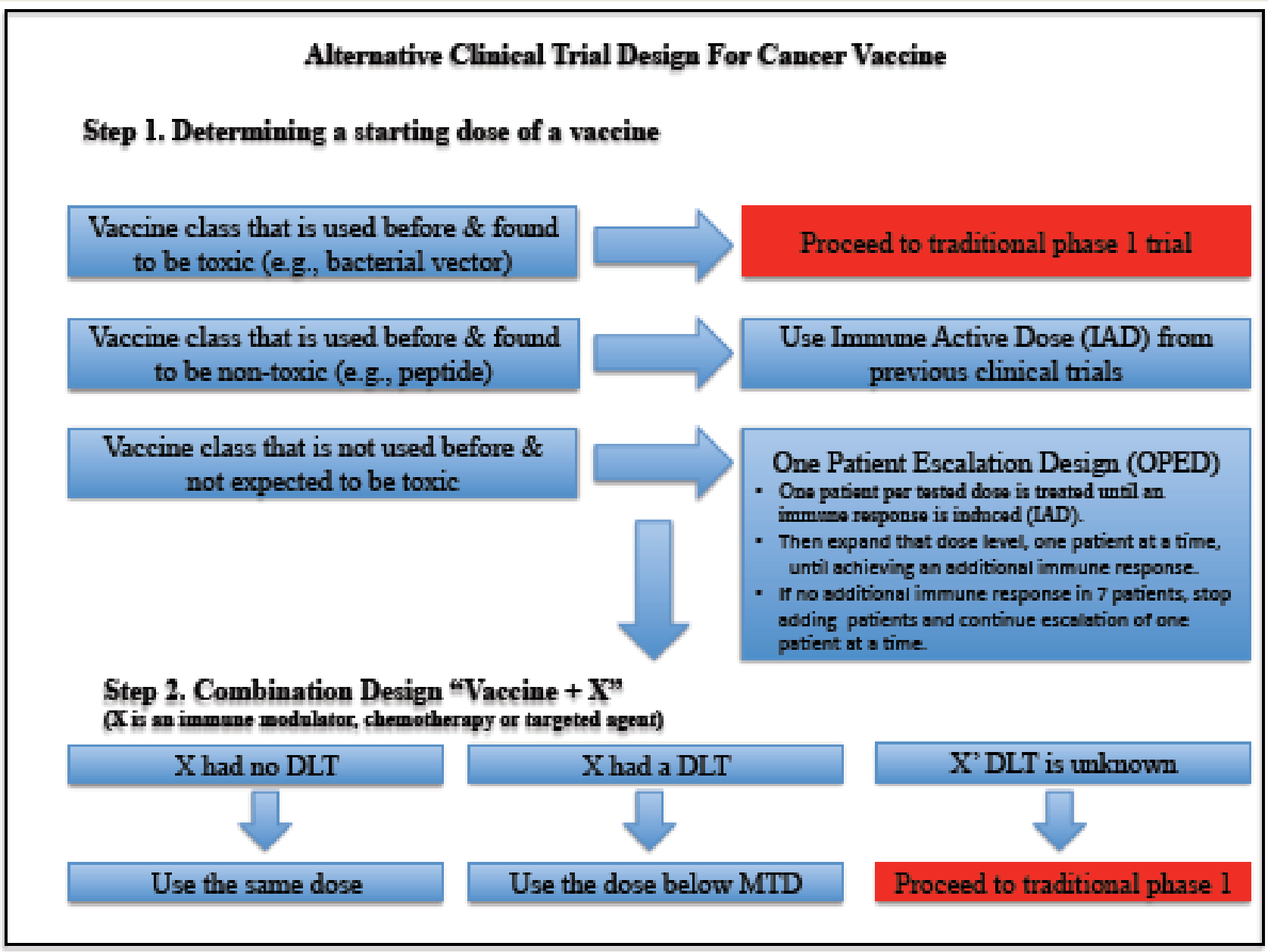
# Novel clinical trial designs for development of immunotherapy combinations

Richard Simon, D.Sc.
Biometric Research Branch
National Cancer Institute
rsimon@nih.gov
http://brb.nci.nih.gov

# Differences Between Therapeutic Vaccines and Cytotoxics

- Many vaccines are incapable of causing immediate serious or life threatening toxicity at doses feasible to manufacture
  - Phase I dose escalation starting from low dose may not be necessary
- Effective vaccination regimens may require combining multiple components (adjuvants, cytokines, costimulatory molecules)




Osama Rahma, Emily Gammoh, Samir Khleif

# Principle

- To detect a large treatment effect does not take many patients or fancy designs

# Optimal single arm two-stage phase II design using tumor shrinkage

- To distinguish 10% ($p_0$) response rate from 40% ($p_1$) response rate with 10% false positive and false negative error rates:
  - Accrue 5 patients. Stop if no responses
  - If at least 1 response, continue accrual to 18 patients total
    - "Accept" treatment if at least 4/18 responses
- For regimens with 10% true response rate, the probability of stopping after 5 patients is 59%

- To distinguish 5% ($p_0$) response rate from 25% ($p_1$) response rate with 10% false positive and false negative error rates:
  - Accrue 9 patients. Stop if no responses
  - If at least 1 response, continue accrual to 24 patients total
    - "Accept" treatment if at least 3/24 responses
- For regimens with 5% true response rate, the probability of stopping after 9 patients is 63%

# Screen 5 treatment regimens

- Accrue (randomize) 9 patients to each treatment (45 patients total)
- Accrue 15 more patients for the treatment regimens for which the number of first stage responses is 1 or more

- If none of the treatments are any good, the expected total sample size is

  $45+5x(1-.63)x15=74$

# Phase II RCT with PFS endpoint

- 1 regimen with randomized control group
- $\alpha$ = 0.10 type 1 error rate
- Detect relatively large treatment effect
- E.g. power 0.8 for detecting 40% reduction in 12 month median PFS requires 70 total events
  - 67% increase in median; eg 6 mos -> 10 months
  - 67% increase in median; eg 3 mos -> 5 months
- Interim analysis can terminate accrual early for futility

# Phase II RCT with PFS endpoint

- Randomized control group
- $\alpha$ = 0.10 type 1 error rate
- Detect relatively large treatment effect
- E.g. power 0.8 for detecting 33% reduction in 12 month median time to recurrence requires 112 total events
  - 50% increase in median; eg 6 mos -> 9 months
- Interim analysis can terminate accrual early for futility

# Improving the efficiency of randomized phase II trials with PFS endpoint

- Multiple vaccine regimens can share one control group in 3 arm trial
- Two stage design:
  - First stage randomize between K vaccine regimens and control
  - Select one vaccine regimen for second stage of accrual for continued randomization against control
  - First stage selection may be based on immunolical response endpoint with final analysis based on PFS

# $2^K$ factorial design

- Basic vaccine V with K possible additional components; e.g. A, B, C
- Randomize patients among the 8 regimens
  - V
  - V+A
  - V+A+B
  - V+A+C
  - V+A+B+C
  - V+B
  - V+B+C
  - V+C

# $2^K$ factorial design

- To evaluate whether A contributes to outcome, compare outcomes for the two composite groups containing and not containing A respectively
  - V
  - **V+A**
  - **V+A+B**
  - **V+A+C**
  - **V+A+B+C**
  - V+B
  - V+B+C
  - V+C

# $2^K$ factorial design

- To evaluate whether B contributes to outcome, compare outcomes for the two composite groups containing and not containing B respectively
  - V
  - V+A
  - **V+A+B**
  - V+A+C
  - **V+A+B+C**
  - **V+B**
  - **V+B+C**
  - V+C

# $2^K$ factorial design

- Compute sample size as for a single 2-arm trial but use a reduced significance level $\alpha$ because 3 comparisons will be performed.
- Assumes that components are additive or synergistic, but not antagonistic
- This can be used as a phase II design to optimize the regimen that will be used in phase III or to screen for synergistic combinations
  - If apparent synergism detected, it can be validated in a subsequent more conventional phase II design

# Screening treatments

- Type I error – a false positive conclusion
- Type II error – a false negative conclusion
- Type III error – failing to study an effective treatment

# Randomized Selection Design With Binary Endpoint

- Large set of candidate treatments
- $\Theta$ = proportion of the candidates that are effective
- $P_{bad}$=true response prob for ineffective regimen
- $P_{bad}$=true response prob for effective regimen

- N total patients available for study
- Perform randomized phase II trial and select the arm with the highest observed response rate for further study
- If the trial has K arms, it will have N/K patients per arm
- With N total patients, determine K and n to maximize probability of selecting an effective regimen for further study

Probability of Selecting a good regimen $p_{bad}$=0.1, $p_{good}$=0.5, $\theta$ =0.1, N=100

| n | K | Probability |
|---|---|---|
| 5 | 20 | 0.626 |
| 10 | 10 | 0.590 |
| 15 | 7 | 0.511 |
| 20 | 5 | 0.414 |
| 25 | 4 | 0.344 |

Probability of Selecting a good regimen $p_{bad}$=0.1, $p_{good}$=0.3, $\theta$ =0.1, N=100

| n | K | Probability |
|---|---|---|
| 5 | 20 | 0.319 |
| 10 | 10 | 0.375 |
| 15 | 7 | 0.383 |
| 20 | 5 | 0.341 |
| 25 | 4 | 0.309 |

Probability of Selecting a good regimen $p_{bad}$=0.1, $p_{good}$=0.3, $\theta$ =0.25, N=100

| n | K | Probability |
|---|---|---|
| 5 | 20 | 0.615 |
| 10 | 10 | 0.708 |
| 15 | 7 | 0.717 |
| 20 | 5 | 0.673 |
| 25 | 4 | 0.642 |

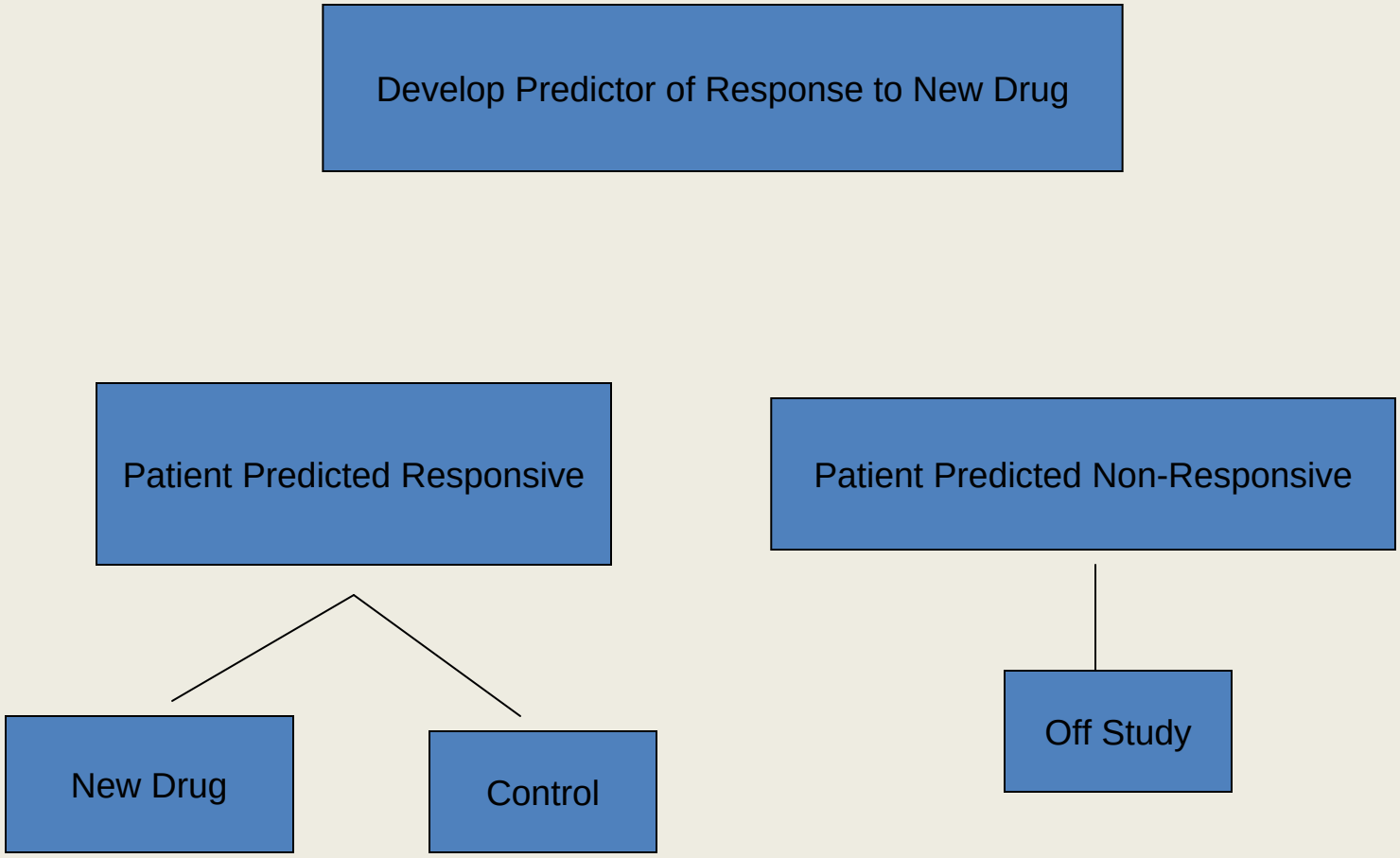
# Phase III designs

- Cancers of  a primary site often represent a heterogeneous group of diseases that differ with regard the oncogenesis and response to treatment
- Current approaches for the design and analysis of phase III clinical trials
  - lack power for identifying treatment effects for subsets of patients
  - Lead to adoption of treatments to which most patients do not benefit
- Current approaches to post-hoc subset analysis are not adequate as a reliable basis for predictive oncology

- How can we develop new treatments in a manner more consistent with modern tumor biology and obtain reliable information about what regimens work for what kinds of patients?

# When the Biology is Clear

- Develop a classifier that identifies the patients most likely to benefit from the new drug
- Develop an analytically validated test
- Design a focused clinical trial to evaluate the effectiveness of the new treatment in test + patients




Targeted (Enrichment) Design

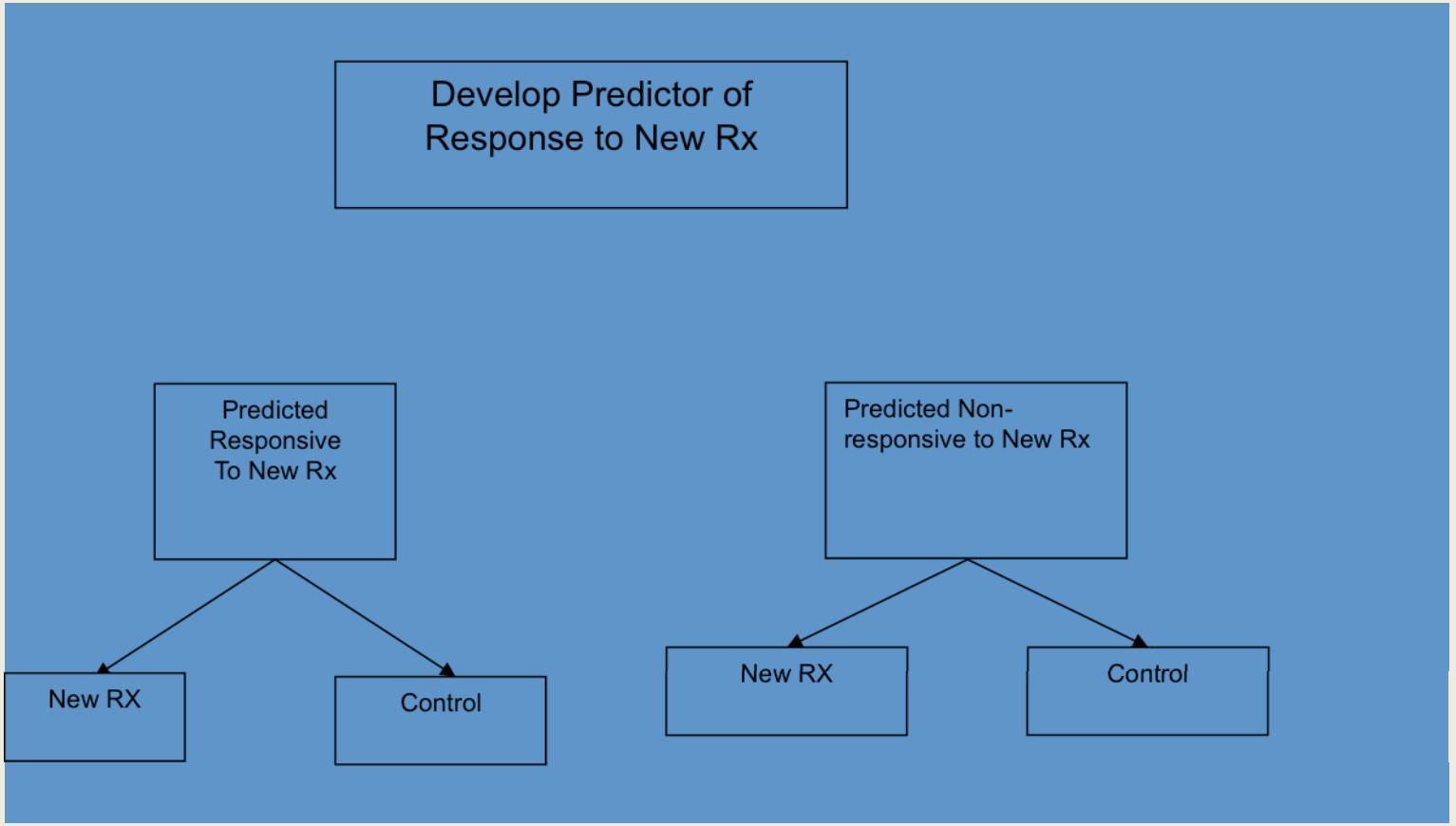
# Evaluating the Efficiency of Targeted Design


- Simon R and Maitnourim A. Evaluating the efficiency of targeted designs for randomized clinical trials. Clinical Cancer Research 10:6759-63, 2004; Correction and supplement 12:3229, 2006
- Maitnourim A and Simon R. On the efficiency of targeted clinical trials. Statistics in Medicine 24:329-339, 2005.


- http://brb.nci.nih.gov

- When less than half of patients are test positive and the drug has limited benefit for test negative patients, the targeted enrichment design requires dramatically fewer randomized patients than the standard design in which the marker is not used
- Website brb.nci.nih.gov provides computational tool for evaluating the efficiency of the targeted enrichment design for specific parameter settings of test accuracy and drug specificity

# Stratification Design for New Drug Development with Companion Diagnostic



# Key features

- The marker should be measured on all patients using an analytically validated test
- Trial should be sized to have adequate power for the comparison of treatments in test + patients at a reduced significance threshold (e.g. 0.02) and for comparison of treatments for overall ITT population at reduced significance threshold (e.g. 0.03)

# Phase III run-in design

Fangxin Hong & R Simon

- Start all eligible patients on a short run-in period on the new treatment
- Measure pharmacodynamic, immunologic or imaging biomarker on all patients at end of the run-in
- Randomize all patients to continue treatment on new treatment or to control regimen
- At final analysis, analyze separately the subset of patients who were marker responsive following the run-in period

# Acknowledgments


- Osama Rhama
- Emily Gammoh
- Samir Khleif
- Aboubakar Maitournam
- Boris Freidlin
- Fangxin Hong