Rational Clinical Experiment: Assessing Prior Probability and Its Impact on the Success of Phase II Clinical Trials

Daniel M. Halperin, J. Jack Lee, Cecile Gonzales Dagohoy, and James C. Yao

All authors: The University of Texas MD Anderson Cancer Center.

Published online ahead of print at www.jco.org on August 10, 2015.

Supported in part by Grants No. N01CM-2011-00039 and CA016672 from the National Cancer Institute.

Presented at the 49th Annual Meeting of the American Society of Clinical Oncology, Chicago, IL, May 31-June 4, 2013, and the 2014 Gastrointestinal Cancers Symposium, San Francisco, CA, January 16-18, 2014.

Authors' disclosures of potential conflicts of interest are found in the article online at www.jco.org. Author contributions are found at the end of this article.

D.M.H. and J.C.Y. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Corresponding author: James C. Yao, MD, Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 426, Houston, TX 77030; e-mail: jyao@mdanderson.org.

© 2015 by American Society of Clinical Oncology

0732-183X/15/3326w-2914w/\$20.00

DOI: 10.1200/JCO.2015.61.4362

A B S T R A C T

Purpose

Despite a robust clinical trial enterprise and encouraging phase II results, the vast minority of oncologic drugs in development receive regulatory approval. In addition, clinicians occasionally make therapeutic decisions based on phase II data. Therefore, clinicians, investigators, and regulatory agencies require improved understanding of the implications of positive phase II studies. We hypothesized that prior probability of eventual drug approval was significantly different across GI cancers, with substantial ramifications for the predictive value of phase II studies.

Methods

We conducted a systematic search of phase II studies conducted between 1999 and 2004 and compared studies against US Food and Drug Administration and National Cancer Institute databases of approved indications for drugs tested in those studies.

Results

In all, 317 phase II trials were identified and followed for a median of 12.5 years. Following completion of phase III studies, eventual new drug application approval rates varied from 0% (zero of 45) in pancreatic adenocarcinoma to 34.8% (24 of 69) for colon adenocarcinoma. The proportion of drugs eventually approved was correlated with the disease under study ($P < .001$). The median type I error for all published trials was 0.05, and the median type II error was 0.1, with minimal variation. By using the observed median type I error for each disease, phase II studies have positive predictive values ranging from less than 1% to 90%, depending on primary site of the cancer.

Conclusion

Phase II trials in different GI malignancies have distinct prior probabilities of drug approval, yielding quantitatively and qualitatively different predictive values with similar statistical designs. Incorporation of prior probability into trial design may allow for more effective design and interpretation of phase II studies.

J Clin Oncol 33:2914-2919. © 2015 by American Society of Clinical Oncology

INTRODUCTION

Regulatory approval of a therapeutic compound is the culmination of efforts in basic science, medicinal chemistry, and clinical drug development. After careful selection and optimization of promising compounds, the medical community conducts clinical testing of only the most promising of those drugs, driven by patient volunteers who risk their personal health to benefit future patients. Clinical investigation therefore represents a solemn contract between investigators and patients who bear this risk to develop only compounds with the greatest chance of benefit.

Unfortunately, despite dramatic scientific advances, only a small proportion of drugs that undergo tests in humans gain approval. In one

analysis of 10 years of data from the world's 10 largest pharmaceutical companies, the rate of European Medicines Agency and US Food and Drug Administration (FDA) approval of new compounds tested in human clinical trials was 11%.¹ Only 5% of medications seeking oncologic indications received approval.

In the clinical development process, therapeutics are tested in progressive phases of clinical trials: phase I determines a recommended dose for future studies, phase II estimates activity, and phase III tests hypotheses about efficacy relative to the standard of care. Although oncologic therapies more frequently advance to phase II and phase III trials than do drugs for other indications, the proportion of cancer drugs gaining FDA approval after the completion of a phase III trial is lower,² highlighting that phase II

studies poorly predict eventual regulatory approval in oncology. Given that phase III trials account for 45% of clinical drug development costs³ and require more patients and time, such late-stage failures are quite costly to funding agencies, industry, and patients alike. Developing strategies to minimize phase III failures in trials of oncology drugs is therefore paramount.

In addition, much as clinicians must often make decisions without a gold standard clinical test to confirm a diagnosis, we must occasionally make treatment decisions without a gold standard randomized controlled phase III clinical trial. In such scenarios, a phase II study may represent the best available evidence, but the best available evidence is not always sufficient. Therefore, improved understanding of the meaning of phase II studies could aid clinical decision making.

GI malignancies comprise a wide spectrum of common and rare diseases with varied biology and management such as colorectal adenocarcinoma, the fourth most common malignancy in the United States, with an estimated incidence of more than 140,000 cases and more than 50,000 deaths in 2013, and pancreatic adenocarcinoma, with more than 45,000 cases and 38,000 deaths in 2013.⁴ Others include esophageal, gastric, hepatobiliary, and neuroendocrine tumors. These cancers comprise a broad, clinically relevant sample for attempting to gain a deeper understanding of clinical drug development.

We hypothesized that one reason phase II studies in oncology are unable to predict eventual regulatory approval could be that the statistical designs of phase II trials do not adequately incorporate the prior probability of success, which generates systematically misleading data on which subsequent phase III studies are based. We therefore evaluated the statistical elements of phase II trials and the prevalence of studies that are testing active compounds to estimate their positive predictive value (PPV) for approval, specifically testing whether the prior probability of drug approval was different across different GI cancers.

Conceptually, the analysis is similar to evaluation of a clinical test, such as computed tomography screening for lung cancer, in which the PPV and negative predictive value (NPV) are functions of prior probability in the population, together with the false-positive and false-negative rates inherent to the test (equations 1 and 2 in the Appendix [online only]). As demonstrated in the evolution of that technique, consideration of prior probability and PPV are critical to effective use of tests.⁵⁻⁹

We therefore assessed the type I and type II error rates used in phase II studies throughout GI oncology, along with rates of eventual drug approval to approximate prior probability, which allowed us to estimate the predictive value of phase II trials.

METHODS

Trial Selection

A systematic search of the clinicaltrials.gov database for interventional phase II and phase III trials of antineoplastic therapies in adults with esophageal cancer, gastric cancer, colon adenocarcinoma, rectal adenocarcinoma, hepatocellular carcinoma, biliary cancers, pancreatic adenocarcinoma, and gastroenteropancreatic neuroendocrine tumors was initiated in January 1, 1999, and continued until December 31, 2004. This 6-year period was selected to allow time for all involved studies to mature.

Table 1. Clinical Trials Published for Disease Types

Disease	No. of Studies	No. of Studies Published	Publication Rate (%)*
Neuroendocrine	17	8	47.1
Pancreatic	70	42	60.0
Biliary	21	12	57.1
Hepatocellular	49	20	40.8
Colon	100	49	49.0
Rectal	103	48	46.6
Gastric	54	31	57.4
Esophageal	59	15	25.4

NOTE. Some studies included multiple disease types, so the sum is greater than 317 (total No. of identified studies).

*Percentage of studies eventually published represented as the publication rate.

Trial Analysis

A review of each trial's database entry and associated publications was conducted to identify the statistical design of that trial and the compounds evaluated. Results were reviewed to determine whether a trial met its pre-defined primary end point.

End Points

The FDA drug approval database, National Cancer Institute drug database, and centerwatch.org were searched to find dates of approval and label changes for each compound tested. The first end point considered was new drug application (NDA), defined as any of the studied compounds with no prior FDA approval receiving an indication in the studied disease more than 12 months after the date when the trial was begun. Given a minimum of 6 months from NDA submission to approval,¹⁰ this 12-month cutoff was used to exclude studies that did not affect the approval process. The second end point considered was a supplemental NDA (sNDA), defined as any of the studied compounds with prior FDA approval in any disease other than the studied disease that received an indication in the studied disease more than 12 months after the date the trial was begun. Approval of NDA or sNDA was considered "success" for that end point.

Statistical Considerations

Descriptive statistics such as the success rate were computed to summarize the findings. FDA approval rates were compared by using Fisher's exact test. Median follow-up was determined by using the Kaplan-Meier method with a cutoff date of December 1, 2013. All statistical analyses were performed by using SPSS v21.0 (SPSS, Chicago, IL).

RESULTS

Clinical Trial Characteristics

We identified a total of 317 phase II trials, with 159 including more than one disease. The number of studies in each disease, together with the number of studies subsequently published, is summarized in Table 1. Median type I (Fig 1A) and type II (Fig 1B) error rates were largely similar among diseases, ranging from 0.05 to 0.1 and from 0.1 to 0.2, respectively. Five studies specified a two-sided type I error rate, and 85 studies used a one-sided type I error rate. Median follow-up time was 12.5 years for NDA-eligible compounds and 12.8 years for sNDA-eligible compounds.

Outcomes

Aggregate success rates, by disease site, are given in Figure 1C. Several diseases yielded no NDAs or sNDAs from the phase II trials

Fig 1. Statistical elements of phase II studies. Median (A) type I and (B) type II error rates are shown by disease. Horizontal lines represent the median, and asterisks and circles represent outliers. (C) Aggregate success rates, by disease, for phase II studies leading to eventual regulatory approval of a compound with (supplemental new drug application [sNDA], blue) or without (new drug application [NDA], gold) approval in another disease. The absolute number (x) of eligible studies (y) approved is also expressed as the fraction (x/y) over each bar expressing the percentage.

Table 2. PPV and NPV for Successful NDA Based on the Prior Probability and Median Type I and Type II Errors From Figure 1

Disease	Type I Error (median)	Type II Error (median)	Prior Probability (%)	PPV (%)	NPV (%)
Neuroendocrine	0.075	0.1	12.50	63.2	98.5
Pancreas	0.095	0.1	0.00	0	100
Biliary	0.05	0.1	0.00	0	100
HCC	0.09	0.1	3.85	28.6	99.6
Colon	0.05	0.116	34.78	90.4	93.9
Rectal	0.05	0.101	34.29	90.4	94.7
Gastric	0.05	0.1	0.00	0	100
Esophageal	0.05	0.13	0.00	0	100

Abbreviations: HCC, hepatocellular carcinoma; NDA, new drug application; NPV, negative predictive value; PPV, positive predictive value.

during this time period, whereas several new drugs in colon and rectal cancers eventually gained approval. Disease site was strongly associated with both NDA ($P \leq .001$) and sNDA ($P = .002$) by Fisher's exact test.

Predictive Probability of Typical Phase II Trials

By using the observed median type I and type II error rates together with the prior probability approximated by the NDA success rate in Figure 1, we calculated the PPV and NPV of a representative phase II trial in each disease (Table 2). PPV and NPV can also be represented as a function of prior probability and type I or type II error (Fig 2), demonstrating that PPV is primarily a function of type I error and prior probability, and NPV is primarily a function of type II error and prior probability. The relationship between PPV, NPV, and prior probability is shown in Appendix Figure A1 (online only). These results derive directly from Bayes' theorem (Equation 3 in the Appendix).

Phase III Success Rates

To test the hypothesis that the proportion of phase III studies leading to FDA approval correlated with the PPV of phase II studies, we evaluated the approval of compounds tested in phase III studies during the same time period. These studies are not necessarily directly linked to the phase II studies, but they are an illustrative sample. These results are shown in Table 3 and demonstrate an NDA success rate in colon cancer of 58.8% (10 of 17) compared with 12.5% (two of 16) in pancreatic adenocarcinoma. Rates of success for NDAs ($P = .016$) and sNDAs ($P = .018$) differed by disease site. Although quantitative comparison has substantial inherent variability given the small numbers of phase III studies, estimates of PPV from Table 2 are included for qualitative comparison.

DISCUSSION

We present a sample of trials in GI oncology and quantify their aggregate success rate by estimating prior probability of success for future studies. We also estimate the PPV and NPV for typical phase II trials in each disease by using the median type I and type II error rates reported. Phase III success rates seem to be similar to these predictive values. Such data demonstrate that prior probability and predictive value are

Fig 2. Relationships between statistical parameters and predictive value. Positive predictive value as a function of prior probability for various (A) type I and (B) type II error rates. Negative predictive value as a function of prior probability for various (C) type I and (D) type II error rates. In (A) and (C), $\beta = 0.1$, and in (B) and (D), $\alpha = 0.1$.

estimable parameters. Furthermore, PPV is highly sensitive to the variations in prior probability in GI oncology. As we hypothesized, the type I and type II error rates in phase II studies do not seem to be related to prior probability of success, indicating a failure to incorporate prior probability into phase II study design.

Table 3. Rates of Eventual Approval Among Drugs Evaluated in Eligible Phase III Studies During the Same Time Period

Disease	sNDA Approved of Those Eligible	sNDA Rate (%)	NDA Approved of Those Eligible	NDA Rate	NDA PPV*
Pancreas	0 of 3	0	2 of 16	12.5	0
Biliary	0 of 0	NA	0 of 1	0	0
HCC	0 of 4	0	0 of 3	0	28.6
Colon	0 of 4	0	10 of 17	58.8	90.4
Rectal	0 of 2	0	8 of 14	57.1	90.4
Gastric	2 of 2	100	1 of 2	50	0
Esophageal	0 of 3	0	0 of 3	0	0

NOTE. Neuroendocrine was not included in this table because there were no phase III trials initiated during the study period.
Abbreviations: HCC, hepatocellular carcinoma; NA, not available; NDA, new drug application; PPV, positive predictive value; sNDA, supplemental new drug application.
*The anticipated PPV from Table 2 is included for reference.

These observations empower clinical investigators to design better studies and enable clinicians, pharmaceutical companies, funding agencies, and health authorities to interpret phase II results more informatively; we submit that phase II results can easily be misinterpreted if prior probability is not appreciated. For study design, the dependence of the PPV on prior probability and type I error suggests that tailoring type I error to the prior probability would improve the ability of phase II studies to predict patient benefit and regulatory approval. Furthermore, reporting PPV would allow more standardized interpretation of phase II trial results, moving beyond promising results to predictive ones. This distinction between regulatory decisions following phase III studies, which necessarily focus on the former, and optimizing drug development, which should focus on the latter. Considering the PPV of phase II studies is not entirely new,¹¹⁻¹⁴ although the concept has not, to the best of our knowledge, been taken to the conclusion presented herein, which represents a fundamental step in operationalizing previous proposals in the broad context of clinical medicine.

We hope that these data will improve clinical practice by quantitatively contextualizing phase II results, allowing clinicians to make more informed decisions when incorporating phase II data into patient care. Although phase II trial results can be encouraging, our data

help quantify their reliability before subsequent confirmation in a manner distinct from qualitative checklists.

For diseases with lower prior probability, strategies must be used to improve the PPV of phase II studies. Multiple potential strategies are possible, including use of our prior probability estimates as a starting point for Bayesian trial designs, selecting type I and type II error rates tailored to the prior probability and desired predictive value in standard frequentist designs, or requiring additional positive phase II studies to reduce program-wide type I error. Although this study focuses mainly on frequentist hypothesis testing, it bears mention that a Bayesian design would be able to estimate the probability of clinical benefit only by incorporating the prior probability of such a benefit, as estimated in this study. Additional strategies to contain phase II size include seeking larger effect sizes, targeting a patient subpopulation more likely to benefit, or incorporating randomized designs, all of which can reduce the likelihood of false positives. It is likely that different solutions are required in different scenarios; using the same design for all trials would likely be as ill-advised as using the same type I and type II error rates for diseases with vastly different prior probabilities.

The most provocative implication of our results is that patients, funds, and investigator effort can be allocated differently when tailored phase II trials better predict phase III success. If we perform a smaller number of phase III trials that are more likely to succeed *a priori*, substantial resources would be freed up to pursue additional promising agents in phase II testing. These results should not deter us from testing compounds in difficult diseases but should refocus our efforts on screening more potentially active compounds before advancing to phase III trials.

To illustrate the quantitative impact of a tailored trial design, we can model the impact of adjusting the design of a group of studies performed in a population of patients with a particular disease. More than 25,000 patients with cancer enroll onto clinical trials annually.¹⁵ For a given disease (eg, hepatocellular carcinoma), we may assume that a total of 2,000 patients may enroll onto phase II clinical trials over several years. Although the exact estimate is hypothetical, we can use this theoretical patient population to model different scenarios.

Whereas phase II studies could use any of a number of designs, we considered studies that used an average of 35 patients each, which would be the maximum sample size in a study that used a Simon

optimal two-stage design¹⁶ with the response rate under H_0 of 0.1 and that under H_1 of 0.3 and a type I error of 0.1 and type II error of 0.1. On the basis of these parameters, an average of 57.14 studies could be conducted with 2,000 patients. Given our estimate of a true-positive rate of 3.85% for compounds tested in this disease, the PPV of these studies would be 26.5%, and the NPV would be 99.6%. We would therefore expect, on average, 1.98 of the 57.14 studies to be truly positive, and 5.49 to be falsely positive, meaning 7.47 compounds would be eligible for advancing to a phase III study, with 73.5% proving negative.

However, if we tailor the statistical design of the studies to the prior probability of success, the results could be dramatically different. By using a similar example of a Simon two-stage design with an H_0 of 0.1 and H_1 of 0.3 with type I and type II errors of 0.02 and 0.2, respectively, the maximum sample size would still be 35 patients. The expected PPV of such studies would be 61.5%, with an NPV of 99.2%. The same 57.14 studies could be completed with the same 2,000 patient volunteers, with 1.76 being truly positive and 1.10 being falsely positive. Thus, a similar number of truly active compounds would be advanced into phase III studies. However, fewer compounds would be advanced to phase III studies overall, and the predicted phase III failure rate would be reduced from 73.5% to 38.4%. This latter design would save resources and prevent patients from receiving inactive therapy, with the exact number depending on the size of the phase III studies. Table 4 presents potential scenarios. Of note, even if phase III studies used only 200 patients, the 922 patient volunteers saved by not pursuing false-positive phase II results could participate in 26 additional phase II studies, identifying one additional truly active compound. The number of patients in this example is hypothetical; the proportions hold true regardless of the size of the population enrolling onto these studies.

Our study has inherent limitations. First, we present a historical sample of clinical trials in an era of increasingly rapid developments in our understanding and therapy of these diseases. Consistent with a Bayesian approach, we present these point estimates as estimates to be refined iteratively as our understanding of the molecular aspects of cancer and of therapeutic approaches evolves. Similarly, estimates of a prior probability of zero should improve with more data. Second, our analysis of study design parameters included only published trials,

Table 4. Hypothetical Impact of Tailored Phase II Trial Design on Patient Use in Phase III Studies

No. of Phase II Studies	Type I Error	Type II Error	No. of Positive Phase II Studies	No. of True-Positive Phase II Studies	No. of Patients per Phase III Study	Total No. of Patients in Phase III Studies
57.14	0.1	0.1	7.47	1.98	200	1,494
57.14	0.02	0.2	2.86	1.76	200	572
57.14	0.1	0.1	7.47	1.98	400	2,988
57.14	0.02	0.2	2.86	1.76	400	1,144
57.14	0.1	0.1	7.47	1.98	600	4,482
57.14	0.02	0.2	2.86	1.76	600	1,716
57.14	0.1	0.1	7.47	1.98	800	5,976
57.14	0.02	0.2	2.86	1.76	800	2,288

NOTE. Assuming that phase II studies are conducted using two-stage Simon optimal design with H_0 of 10% and H_1 of 30%, given a disease in which prior probability of success is 3.85%, a study using observed type I and type II error parameters will generate more false positives than true positives. In the scenarios presented here, a tailored phase II program accounting for the prior probability will reduce the No. of patients required for phase III studies by 62%. As the No. of patients required for a phase III study increases, the benefit of tailored trial design and reduction in false-positive phase II studies becomes larger, ranging from 922 if phase III studies have an average of 200 patients to 3,688 if phase III studies have an average of 800 patients.

which may be different from unpublished trials, especially if a publication bias against negative studies exists. Some may raise the concern that our data may not be sufficiently mature, but a contemporaneous analysis revealed a median time from initial clinical trial submission to eventual FDA approval of 6.3 years,¹⁷ so a median follow-up time of 12.7 years should allay such concerns. sNDA rates may be particularly sensitive to financial decisions to stop development as drugs near the end of their patents. Randomized studies account for only a small fraction of our sample set, so although the concept of prior probability should hold with such designs, more data will be required to understand the impact of randomization. In addition, we do not directly account for expansion cohorts in phase I studies but would submit that our framework applies to any study that estimates efficacy with a frequentist or Bayesian approach. Finally, it is notable that the studies we considered may not be entirely independent because it was not possible to completely distinguish the drug classes and clinical populations under study.

It remains likely that multiple other strategies will also help improve drug development. Using molecularly targeted therapies in defined patient populations is expected to improve results. We suggest that incorporating prior probability into studies testing these compounds would be better than not incorporating prior probability at all. It is possible that primary tumor histology will not be the most important factor in determining prior probability, and multivariable models to estimate prior probability would be a significant refinement of the single-variable model used in this proof-of-concept study.

We acknowledge the limitations of intermediate end points used in phase II studies and welcome the incorporation of more predictive intermediate end points. As more predictive end points are used in phase II studies, the probability of approval should increase.

Clear future directions would include analysis of additional parameters beyond primary tumor site and expansion to more recent trials. As the granularity of our data improves, we may be able to more

specifically define the situation being studied and therefore more accurately estimate the prior probability of success. We defer the question of what would be an ideal phase II PPV; it is certainly in need of improvement.

Our modeling of the impact of tailored design should assist researchers in designing future clinical trials. By understanding the prior probability of success under defined circumstances, our trial designs can be targeted with as much nuance as our molecular therapeutics. Our explicit goal is optimizing the odds of regulatory approval and patient benefit. We believe this goal can be achieved through phase II trials that advance compounds with higher probabilities of success into a smaller number of higher-yield phase III trials. Although such a concept may initially seem drastic, the application of Bayesian principles could bring the concepts of evidence-based medicine to the generation of the evidence, thereby enhancing the success rate of drug development.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Daniel M. Halperin, J. Jack Lee, James C. Yao

Financial support: James C. Yao

Administrative support: James C. Yao

Collection and assembly of data: Daniel M. Halperin, Cecile Gonzales Dagohoy

Data analysis and interpretation: Daniel M. Halperin, J. Jack Lee, James C. Yao

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

1. Kola I, Landis J: Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3:711-715, 2004
2. DiMasi JA, Grabowski HG: Economics of new oncology drug development. *J Clin Oncol* 25:209-216, 2007
3. DiMasi JA: The value of improving the productivity of the drug development process: Faster times and better decisions. *Pharmacoeconomics* 20:1-10, 2002
4. Siegel R, Naishadham D, Jemal A: Cancer statistics, 2013. *CA Cancer J Clin* 63:11-30, 2013
5. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al: Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 365:395-409, 2011
6. Bach PB, Mirkin JN, Oliver TK, et al: Benefits and harms of CT screening for lung cancer: A systematic review. *JAMA* 307:2418-2429, 2012
7. Boiselle PM: Computed tomography screening for lung cancer. *JAMA* 309:1163-1170, 2013
8. National Lung Screening Trial Research Team, Church TR, Black WC, et al: Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 368:1980-1991, 2013
9. Tammemägi MC, Katki HA, Hocking WG, et al: Selection criteria for lung-cancer screening. *N Engl J Med* 368:728-736, 2013
10. DiMasi JA, Hansen RW, Grabowski HG: The price of innovation: New estimates of drug development costs. *J Health Econ* 22:151-185, 2003
11. Ratain MJ: Phase II oncology trials: Let's be positive. *Clin Cancer Res* 11:5661-5662, 2005
12. Sharma MR, Karrison TG, Jin Y, et al: Resampling phase III data to assess phase II trial designs and endpoints. *Clin Cancer Res* 18:2309-2315, 2012
13. Yao JC, Meric-Bernstam F, Lee JJ, et al: Accelerated approval and breakthrough therapy designation: Oncology drug development on speed? *Clin Cancer Res* 19:4305-4308, 2013
14. Lee JJ, Liu DD: A predictive probability design for phase II cancer clinical trials. *Clin Trials* 5:93-106, 2008
15. Scoggins JF, Ramsey SD: A national cancer clinical trials system for the 21st century: Reinvigorating the NCI Cooperative Group Program. *J Natl Cancer Inst* 102:1371, 2010
16. Simon R: Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 10:1-10, 1989
17. Keyhani S, Diener-West M, Powe N: Are development times for pharmaceuticals increasing or decreasing? *Health Aff (Millwood)* 25:461-468, 2006

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Rational Clinical Experiment: Assessing Prior Probability and Its Impact on the Success of Phase II Clinical Trials

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Daniel M. Halperin

Other Relationship: Novartis

J. Jack Lee

No relationship to disclose

Cecile Gonzales Dagohoy

No relationship to disclose

James C. Yao

No relationship to disclose

Appendix

The following are equations used in our study for determining positive predictive value (PPV) and negative predictive value (NPV) along with Bayes Theorem. P represents probability, T represents test result, and D represents drug regulatory result.

$$PPV = \frac{P(T^+|D^+) \times P(D^+)}{P(T^+|D^+) \times P(D^+) + P(T^+|D^-) \times P(D^-)}$$

$P(D^+)$, prior probability of drug being effective/approved; $P(D^-)$, prior probability of drug NOT being effective/approved = $1 - P(D^+)$; $P(T^+|D^+)$, probability that positive trial-given drug will be effective/approved = $1 - \text{type II error rate}$; $P(T^+|D^-)$, probability that positive trial-given drug will NOT be effective/approved = type I error rate.

$$NPV = \frac{P(T^-|D^-) \times P(D^-)}{P(T^-|D^-) \times P(D^-) + P(T^-|D^+) \times P(D^+)}$$

$P(D^+)$, prior probability of drug being effective/approved; $P(D^-)$, prior probability of drug NOT being effective/approved = $1 - P(D^+)$; $P(T^-|D^-)$, probability that negative trial-given drug will NOT be effective/approved = $1 - \text{type I error rate}$; $P(T^-|D^+)$, probability that negative trial-given drug will be effective/approved = type II error rate.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

$P(A)$ is the probability of A being true; $P(B)$ is the probability of B being true; $P(A|B)$ is the probability of A being true, given that B is observed; $P(B|A)$ is the probability of B being true, given that A is observed.

Fig A1. Contour plots of positive and negative predictive value as a product of type I and type II error for various prior probabilities. (A) Prior probability of 0.05; (B) prior probability of 0.3; and (C) prior probability of 0.5. Blue lines, positive predictive value; gold lines, negative predictive value.