JAMA Oncology

View Article ▸

Use of Bayesian Decision Analysis to Minimize Harm in Patient-Centered Randomized Clinical Trials in Oncology

Vahid Montazerhodjat, PhD,[1,2] Shomesh E. Chaudhuri, MS,[1,3] Daniel J. Sargent, PhD,[4] and Andrew W. Lo, PhD[X] [1,3,5,6]

[1]Laboratory for Financial Engineering, MIT Sloan School of Management, Cambridge, Massachusetts

[2]Department of Computer Science, Boston College, Chestnut Hill, Massachusetts

[3]Department of Electrical Engineering and Computer Science, MIT, Cambridge, Massachusetts

[4]Mayo Clinic, Rochester, Minnesota

[5]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts

[6]AlphaSimplex Group LLC, Cambridge, Massachusetts

[X]Corresponding author.

Article Information

Corresponding Author: Andrew W. Lo, PhD, Laboratory for Financial Engineering, MIT Sloan School of Management, 100 Main St, E62-618, Cambridge, MA 02142 (alo-admin@mit.edu).

Key Points

Question

How can patient preferences and burden of disease be explicitly incorporated into randomized clinical trials (RCTs) in oncology and what is the impact on statistical thresholds for drug approval?

Findings

In this analysis, Bayesian decision analysis (BDA) was applied to a data set of 10 clinical trials from the Alliance for Clinical Trials in Oncology. The BDA-optimal alphas were often much larger than 2.5% for terminal cancers with short survival times and no effective therapies (eg, pancreatic cancer) and smaller than 2.5% for less serious cancers with long survival times, several effective therapies, and high prevalence.

Meaning

Bayesian decision analysis can be applied to RCTs by choosing a sample size (n) and type 1 error rate (alpha) to minimize the overall expected harm to current and future patients, where expected harm is computed under both null and alternative hypotheses.

Abstract

Importance

Randomized clinical trials (RCTs) currently apply the same statistical threshold of alpha = 2.5% for controlling for false-positive results or type 1 error, regardless of the burden of disease or patient preferences. Is there an objective and systematic framework for designing RCTs that incorporates these considerations on a case-by-case basis?

Objective

To apply Bayesian decision analysis (BDA) to cancer therapeutics to choose an alpha and sample size that minimize the potential harm to current and future patients under both null and alternative hypotheses.

Data Sources

We used the National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) database and data from the 10 clinical trials of the Alliance for Clinical Trials in Oncology.

Study Selection

The NCI SEER database was used because it is the most comprehensive cancer database in the United States. The Alliance trial data was used owing to the quality and breadth of data, and because of the expertise in these trials of one of us (D.J.S.).

Data Extraction and Synthesis

The NCI SEER and Alliance data have already been thoroughly vetted. Computations were replicated independently by 2 coauthors and reviewed by all coauthors.

Main Outcomes and Measures

Our prior hypothesis was that an alpha of 2.5% would not minimize the overall expected harm to current and future patients for the most deadly cancers, and that a less conservative alpha may be necessary. Our primary study outcomes involve measuring the potential harm to patients under both null and alternative hypotheses using NCI and Alliance data, and then computing BDA-optimal type 1 error rates and sample sizes for oncology RCTs.

Results

We computed BDA-optimal parameters for the 23 most common cancer sites using NCI data, and for the 10 Alliance clinical trials. For RCTs involving therapies for cancers with short survival times, no existing treatments, and low prevalence, the BDA-optimal type 1 error rates were much higher than the traditional 2.5%. For cancers with longer survival times, existing treatments, and high prevalence, the corresponding BDA-optimal error rates were much lower, in some cases even lower than 2.5%.

Conclusions and Relevance

Bayesian decision analysis is a systematic, objective, transparent, and repeatable process for deciding the outcomes of RCTs that explicitly incorporates burden of disease and patient preferences.

Introduction

There is general agreement in the biomedical community that the development of therapies for certain diseases should take priority. This ethic has motivated legislative initiatives, such as the Orphan Drug Act of 1983, and underpins several important innovations in regulatory approval processes, such as the US Food and Drug Administration's (FDA) fast-track, breakthrough-therapy, accelerated-approval, and priority-review designations. However, none of these innovations directly address the critical issue of how to incorporate the patient's perspective in deciding whether a drug candidate should be approved or not.

The current approach in clinical trial design is to minimize the chance of ineffective treatment caused by a type 1 error, that is, a false-positive result. However, the arbitrary nature of the threshold for the probability of type 1 error, alpha, raises an ethical question about its justification. A 2.5% threshold may not be appropriate for terminal illnesses that have no effective therapies; such patients may prefer to take a bigger chance on a false-positive result, even if the likelihood of an effective therapy is small. To quote the noted biostatistician Donald Berry, "We should also focus on patient values, not just *P* values."

We propose to incorporate patient values and preferences into clinical trials in an objective, systematic, transparent, and repeatable manner using Bayesian decision analysis (BDA). This is a well-known quantitative framework for making the tradeoff between type 1 and type 2 errors, balancing the consequences of false-positive and false-negative errors on patients. While Bayesian methods have long been used in clinical trial design, they are less popular in practice, in part because of the research community's inexperience with unfamiliar methods. However, recently there has been renewed interest in the Bayesian approach, highlighted by the FDA's commitment to "facilitate the advancement and use of complex adaptive, Bayesian, and other novel clinical trial designs." Motivated by these developments, we previously proposed a novel framework to calculate the optimal values of the alpha and power for randomized clinical trials (RCTs) that minimize the expected harm to patients, given the parameters relevant to any specific disease.

Herein we apply this framework specifically to oncology therapeutics. The appropriate cost parameters and prior odds ratios were first estimated for the 23 most common cancer sites in the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) database, and used to construct hypothetically optimal balanced 2-arm fixed-sample RCTs to minimize the average impact of both types of errors on patients. We then applied this framework to actual clinical trial data from 10 current phase 3 studies sponsored by the Alliance for Clinical Trials in Oncology (Alliance), an NCI-funded group that

performs large national phase 2 and 3 clinical trials, and performed a similar analysis using various patient-appropriate endpoints. We find that the BDA-optimal design is often starkly different in size, power, and sample size from the traditional approach.

Methods

We considered a hypothetical new therapy, with a given hazard ratio assuming it is effective, to be tested in a balanced 2-arm fixed-sample RCT, where the endpoint is overall survival. To specify a fixed-sample RCT, we required 2 parameters: the number of participants in each arm of the study, n, and the probability of type 1 error, alpha, where the null hypothesis is the case where the drug is ineffective and possibly toxic (the power can be calculated using the sample size of the RCT, ie, n, and its alpha). The RCT search space for the optimal trial consists of all possible combinations of n and alpha with each pair of values defining a particular fixed-sample RCT.

To define the potential harm or cost associated with a given RCT, we considered the 2 possible outcomes for the therapy: effective or ineffective. If the therapy is effective, the 2 costs associated with an RCT are: (1) the duration of the trial, when patients outside of the treatment arm are not receiving the therapy; and (2) the loss to all patients who could have benefited if this effective therapy is incorrectly rejected in the trial. If the therapy is ineffective and possibly harmful, the costs are: (1) the adverse effects of the therapy on patients in the treatment arm during the trial; and (2) the adverse effects on all patients who use this therapy if it is incorrectly approved. These costs depend on a number of auxiliary parameters—the degree and duration of health benefits for an effective therapy and the severity of adverse effects for an ineffective therapy—that can be estimated using epidemiological and clinical-trial data.

Once these costs have been estimated for each scenario, they were multiplied by the probability of each scenario and summed to yield an overall expected cost of the RCT—not to be confused with the financial costs associated with the RCT—which is often called "Bayes risk" in decision theory. The objective of BDA is to compute the optimal sample size (n*) and type 1 error (alpha*) that jointly minimize the expected cost of the trial. In other words, we sought to conduct a trial that minimizes the average cost to patients—both in the trial and in the general population—where the average is taken over both possibilities of effective and ineffective therapies.

BDA-optimal trials can also be interpreted as trials that minimize the expected harm to patients, where harm is either: type 1 harm—an extra burden on patients owing to the adverse effects of the treatment in the case of a toxic and ineffective drug, caused by a false-positive result; or type 2 harm—a missed opportunity to reduce the burden of disease on patients owing to the length of the RCT (even if the drug is approved) and/or a rejection of an effective treatment in the RCT, caused by a false-negative result.

Type 2 harm is rarely discussed in medical and lay communities because it is difficult to quantify the number of missed opportunities, especially compared with the highly visible backlash created by incorrectly approving a toxic drug. However, missed opportunities to reduce the burden of disease on current and future patients, ie, type 2 harm, have real and quantifiable social costs, just as type 1 harm does. Unless these types of harm are properly balanced against each other, highly conservative drug approval processes may not be protecting all patients from harm. The primary objective of this article is to propose an objective method for balancing these harms explicitly.

Although the effectiveness and possible adverse effects of a drug are not precisely known at the time of the RCT design, it is still possible to list scenarios—both positive and negative—that the drug might face, along with their implications for patients. It is also possible to construct plausible estimates of the likelihood of each scenario using the information that the trial investigators and sponsors have at their disposal from previous clinical phases at the time of the RCT design. Therefore, not only is it practical to design a quantitative framework where the risks of a treatment are balanced against its benefits, it is also ethically necessary to ensure that both types of harm are accounted for when deciding whether a drug should be approved.

Results

The utility of BDA-optimal RCTs can be illustrated by applying the methodology to each of the 23 most common cancer sites based on estimated prevalence counts (prevalence proportions times US population estimates) listed in the NCI's SEER database. For each cancer site, we determined the optimal balanced 2-arm fixed-sample RCT for testing a therapy that targets the late stage of the cancer, where the endpoint is overall survival. A complete list of assumptions on the RCT setting is provided in Table 1. These are clearly hypothetical examples, because treatment for each cancer site is highly dependent on the stage and the patient (see the Supplement for the specific assumptions underlying the cost estimates and probabilities for types 1 and 2 errors). To allow the reader to verify the impact of specific assumptions, we have provided an easy-to-use interactive tool in the Supplement that calculates the BDA-optimal RCT design for various input parameter values. The results are contained in Table 2.

The entries in this table show that cancers with the worst prognoses, eg, cancers of the brain and pancreas, have relatively large BDA-optimal type 1 error rates (alpha) of 47.9% and 26.6%, respectively. Patients with terminal disease simply cannot afford to miss any effective drugs that can extend their lives by 11 months for brain cancer, and by 5 months for pancreatic cancer. These values differ greatly from the BDA-optimal type 1 error rates of breast cancer, colorectal cancer, and lymphomas—17.6%, 13.1%, and 12.2 to 12.8%, respectively. The prognosis for this set of cancers is considerably more optimistic than that of the former set, even for patients with late-stage disease. It is worth noting, however, that in all cases the type 1 error rates recommended by the BDA far exceed the traditional standard of 1-sided alpha, namely, 2.5%. Finally, although there is, in general, little variation in optimal type 2 error rates, in cancers with the best prognosis, Hodgkin lymphoma and cancer of the testis, the recommended power is well below 90%, owing to the need to keep the trial duration short to avoid exposing too many patients to inferior medications in the treatment arms of these trials.

A sensitivity analysis is provided in the Supplement to investigate the robustness of these results to perturbations in our model's key parameters. We found that cancers with poor prognoses consistently had relatively large BDA-optimal type 1 error rates and small optimal RCT sample sizes. Our observation that a patient with a poor prognosis cannot afford to miss any effective drugs—even in the face of greater risk of false-positive results—is robust over a wide range of parameters. Moreover, all the type 1 error rates recommended by the BDA analysis remain far in excess of the traditional 2.5% 1-sided alpha. However, the specific critical value and sample size of each optimal RCT is sensitive to the underlying assumptions. For example, a 15% increase in the a priori probability of an ineffective therapy from 65% to 80% leads to a more conservative trial design, reducing the optimal alpha for brain cancer RCTs from 48% to 19% and increasing the optimal sample size from 152 to 268. Conversely, decreasing either the patient accrual rate or the toxic effects of an ineffective therapy leads to less conservative (ie, larger alpha and smaller sample size) RCT designs. Intuitively, decreasing the patient accrual rate increases the trial length, and for patients with short life expectancies, the optimal tradeoff involves maintaining a relatively short trial length.

Similarly, decreasing the toxic effects of an ineffective drug under the null hypothesis reduces the cost of a more aggressive RCT design. When taken to the limit of no toxic effects—clearly an unrealistic assumption—the optimal RCT design becomes extremely aggressive and the protocol approves the majority of investigational drugs after minimal clinical trial study. In this case, there are few benefits gained by rejecting an ineffective drug, mitigating the tradeoff central to the expected cost optimization. Note that a nontoxic therapy in this model is one that is equally as effective as the standard treatment, and therefore should be considered a limiting case. This example highlights the need for carefully considered assumptions and accurately calibrated cost models when implementing the BDA-framework (Supplement).

A practical illustration of the BDA methodology can be obtained using actual clinical-trial data from the Alliance portfolio to compute BDA-optimal RCTs for 10 of the phase 3 clinical trials currently actively enrolling or following patients, and comparing the results with the current designs of the Alliance trials.

The results are presented in Table 3, where the last 3 columns characterize the BDA-optimal RCT for each cancer site, arranged by rows. The features of BDA-optimal RCTs are summarized in Figure 1 and Figure 2, which show substantial departures from the comparable parameters of the Alliance trials, especially for high-mortality and low-prevalence cancers.

The differences between traditional and BDA-optimal RCTs are especially striking in 4 rows of Table 3: glioblastoma (row 1); castration-resistant metastatic prostate cancer (row 4); stage III colon cancer (row 8); and early-stage prostate cancer (clinical stage ≤T2a, row 10).

For glioblastoma (GBM), there was a stark contrast between the conventionally designed current RCT and the BDA-optimal RCT. The sample size for the conventional RCT was 400 patients, while the BDA-optimal sample size was 104, a 74% reduction. Moreover, the type 1 error rate for the BDA-optimal trial was 47.5%, much larger than the standard 2.5% 1-sided type 1 error rate set in the traditional RCT (in fact, the Alliance trial used twice the standard 2.5% type 1 error in recognition of the limited population and poor prognosis of GBM patients).

The smaller number of patients and larger alpha in the BDA-optimal trial were more permissive than the comparable values for traditional RCTs so as to reduce type 2 harm. The decrease in type 2 harm was large enough to offset the excess risk resulting from the extra permissiveness in the trial, and the overall penalty —the expected harm to current and future patients—was minimized under the BDA-optimal RCT.

For castration-resistant metastatic prostate cancer, we also observed a clear difference between the traditional and BDA-optimal RCTs. The sample size of the BDA-optimal RCT was only 55% of the sample size for the traditional RCT, 676 vs 1224 patients, and the type 1 error rate for the BDA-optimal trial was almost 8 times higher than that of the traditional RCT, 20.4% vs 2.5%. This was not surprising, since patients with late-stage prostate cancer have a median overall survival time as low as 35 months.

For stage III colon cancer, these patients have a 79% 5-year survival rate, and the traditional and BDA-optimal RCTs were almost equivalent, with sample sizes of 2500 vs 2232, and type 1 error rates of 2.5% vs 2.3%, respectively.

Finally, for early-stage prostate cancer (clinical stage ≤T2a) therapies, the BDA-optimal RCT was more conservative than the current Alliance RCT. The BDA-optimal RCT was slightly smaller than the traditional RCT, 418 vs 464 patients, while allowing a much smaller chance for false-positive results— 0.9% vs 2.5% in the conventional RCT. In this case, the harm from approving an ineffective therapy was considerably more serious than rejecting an effective one because the burden of disease was relatively less severe while the adverse effects of an ineffective therapy would impact a large number of patients, hence the more conservative BDA-optimal parameters.

Limitations

Our findings must be qualified in several respects. First, we have considered only traditional fixed-sample RCTs; in practice, adaptive trial designs may include an interim analysis for early signals of efficacy, futility, or toxic effects, or may be adaptive in other ways. Any of these possible adaptations in any given trial may alter the optimal type 1 and 2 error rates and appropriate modifications to our calculations are required to determine the optimal designs for these settings.

Second, the trials considered here use the overall survival endpoint, which is clear and of unambiguous importance. However, for a variety of reasons, many trials use alternative endpoints, such as progression-free survival, the clinical relevance of which is less clear. Study-specific definitions of type 1 and 2 harm would require greater subtlety in trials with endpoints other than overall survival.

Third, owing to recent advances in cancer biology and a better understanding of cancer molecular profiles, it is clear that cancer—even within a single site—refers to a collection of heterogeneous diseases with different molecular and genetic profiles. Our framework can be readily adapted to subdiseases within each of these cancers, provided that relatively accurate information on the burden of these subdiseases and their survival statistics, prevalence, incidence, and death rates are available.

Fourth, even though type 1 errors like 47.5% for GBM may be optimal for terminal illnesses with no existing treatments, they could inadvertently encourage the development of marginal therapies. This adverse incentive can be addressed by asking the FDA to create a new class of experimental therapeutics that have fixed terms of contingent approval, contingent on stringent postapproval monitoring where more data will be collected and analyzed. If the new data confirm the therapy's efficacy, the contingent approval status can be converted to unconditional approval, otherwise the contingent approval expires.

Finally, we have confined our attention to patients' medical outcomes without considering the cost to patients and their families, to industry, or to society. New therapeutic agents often come at a very high financial cost, which, when taken into account, may raise the bar of success for new agents, thus lowering the acceptable type 1 error rate. On the other hand, the increased type 1 error rates that we have proposed may lower the cost of clinical trials and reduce the risk to sponsors, which may encourage drug development, lower drug costs, and further accelerate clinical research. To incorporate perspectives from the entire biomedical ecosystem, as well as the value of patient input to the drug development process, we have proposed that the FDA form a patient advisory board consisting of key stakeholder groups—patients, caregivers, physicians, biopharma executives, regulators, and policymakers—with the specific charge of formulating explicit cost estimates for type 1 and type 2 errors. These estimates can then be incorporated into the FDA decision-making process as additional inputs to their quantitative and qualitative deliberations.

Conclusions

Traditional RCTs do not necessarily minimize overall harm to current and future patients, especially for life-threatening cancers that currently have no effective therapies. In these cases, traditional RCTs are too lengthy, too conservative, and focused too much on rejecting ineffective drugs and avoiding false-positive results. This single-minded focus can result in missed opportunities to treat life-threatening conditions, which can sometimes harm more patients than mistakenly approving ineffective and possibly toxic drugs.

Conversely, for some less aggressive cancers, such as early-stage prostate cancer, the current thresholds of statistical significance are more permissive than the BDA-optimal thresholds. In these cases, traditional RCTs allow a larger chance of falsely approving ineffective and possibly toxic drugs, risking patients' health even though the potential benefits from these trials do not necessarily justify the risk.

The ability of the BDA framework to systematically weigh multifaceted tradeoffs that reflect a variety of perspectives combined with its flexibility and practicality make it a potentially valuable tool for optimal RCT design. While the framework is robust, we emphasize that careful consideration must be applied to the assumptions underlying the specific models in order to produce useful recommendations. If correctly implemented, the Bayesian perspective has the potential to benefit all stakeholders.

Notes

Supplement.

eAppendix 1. Expected RCT Penalty

eAppendix 2. Assumptions Underlying Hypothetical BDA-Optimal RCTs for 23 Cancer Sites

eFigure 1. Sensitivity of the BDA-optimal 1-sided α to the accrual rate for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to [150%, 125%, 100%, 75%, 50%] of the accrual rate proposed in our study for each cancer.

eFigure 2. Sensitivity of the BDA-optimal sample size to the accrual rate for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to [50%, 75%, 100%, 125%, 150%] of the accrual rate proposed in our study for each cancer. A lower limit of 40 is applied to the sample size to ensure the log-rank statistic is approximately standard normal

eFigure 3. Sensitivity of the BDA-optimal 1-sided α to the probability that the investigational drug is effective (p_1) for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to p_1= [20%, 27.5%, 35%, 42.5%, 50%].

eFigure 4. Sensitivity of the BDA-optimal sample size to the probability that the investigational drug is effective (p_1) for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to p_1= [50%, 42.5%, 35%, 27.5%,

20%]. A lower limit of 40 is applied to the sample size to ensure the log-rank statistic is approximately standard normal.

eFigure 5. Sensitivity of the BDA-optimal 1-sided α to the side-effect level of burden of an ineffective drug ($\Delta y\_"tox"$) under the null hypothesis for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $\Delta y\_"tox"$ = [12.6%, 9.45%, 6.3%, 3.15%, 0%], where a 6.3% burden means that each patient experiencing side-effects would be indifferent to living each year with the side effects, or to losing 6.3% of each year (about 23 days) if, for the rest of that year, they could live without the side effects.

eFigure 6. Sensitivity of the BDA-optimal sample size to the side-effect level of burden of an ineffective drug ($\Delta y\_"tox"$) under the null hypothesis for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $\Delta y\_"tox"$ = [0%, 3.15%, 6.3%, 9.45%, 12.6%], where a 6.3% burden means that each patient experiencing side effects would be indifferent to living each year with the side effects, or to losing 6.3% of each year (about 23 days) if, for the rest of that year, they could live without the side effects. A lower limit of 40 is applied to the sample size to ensure the log-rank statistic is approximately standard normal.

eFigure 7. Sensitivity of the BDA-optimal 1-sided α to the magnitude reduction in life expectancy due to the adverse effects of an ineffective drug ($\Delta \mu\_"tox"$) under the null hypothesis for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $\Delta \mu\_"tox"$ = [4, 3, 2, 1, 0] months.

eFigure 8. Sensitivity of the BDA-optimal sample size to the magnitude reduction in life expectancy due to the adverse effects of an ineffective drug ($\Delta \mu\_"tox"$) under the null hypothesis for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $\Delta \mu\_"tox"$ = [0, 1, 2, 3, 4] months. A lower limit of 40 is applied to the sample size to ensure the log-rank statistic is approximately standard normal.

References

1. US Food and Drug Administration Guidance for industry: Expedited programs for serious conditions—drugs and biologics. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm358301.pdf. Published May 2014. Accessed June 20, 2016.

2. Berry DA. How to take clinical research to the next level. Fortune website. http://fortune.com/2015/10/26/cancer-clinical-trial-belmont-report/. Accessed October 28, 2015.

3. Berry DA. Trial design committee session. Presented at: GBM-AGILE Workshop; August 11–12, 2015; Phoenix, AZ.

4. Anscombe FJ. Sequential medical trials. J Am Stat Assoc. 1963;58(302):365-383.

5. Colton T. A model for selecting one of two medical treatments. J Am Stat Assoc. 1963;58(302):388-400.

6. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. Stat Med. 1995;14(3):231-246. [PubMed: 7724909]

7. Cheng Y, Su F, Berry DA. Choosing sample size for a clinical trial using decision analysis. Biometrika. 2003;90(4):923-936.

8. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. Stat Sci. 2004;19(1):175-187.

9. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006;5(1):27-36. [PubMed: 16485344]

10. Armitage P. Sequential medical trials: some comments on FJ Anscombe's paper. J Am Stat Assoc. 1963;58(302):384-387.

11. US Food and Drug Administration PDUFA reauthorization performance goals and procedures fiscal years 2018 through 2022. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM511438.pdf. Published July 2016. Accessed August 18, 2016.

12. Isakov L, Lo AW, Montazerhodjat V Is the FDA too conservative or too aggressive? a Bayesian decision analysis of clinical trial design. SSRN; 2015. https://ssrn.com/abstract=2641547. Accessed February 8, 2017.

13. Djulbegovic B, Kumar A, Soares HP, et al. . Treatment success in cancer: new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. Arch Intern Med. 2008;168(6):632-642. [PMCID: PMC2773511] [PubMed: 18362256]

14. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2012. Bethesda, MD: National Cancer Institute; 2014. http://seer.cancer.gov/csr/1975_2012/. Updated November 18, 2015. Accessed August 18, 2016.

15. Murray CJL, Atkinson C, Bhalla K, et al. ; U.S. Burden of Disease Collaborators . The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA. 2013;310(6):591-608. [PMCID: PMC5436627] [PubMed: 23842577]

16. Alberts SR, Sargent DJ, Nair S, et al. . Effect of oxaliplatin, fluorouracil, and leucovorin with or without cetuximab on survival among patients with resected stage III colon cancer: a randomized trial. JAMA. 2012;307(13):1383-1393. [PMCID: PMC3442260] [PubMed: 22474202]

Figures and Tables

Table 1.

Assumptions for RCTs

Parameter	Assumed Value	Comments
Probability that the drug is effective.	35%	This is estimated using historical numbers for oncology compounds and assuming 80% power for historical phase 3 RCTs.
Expected excess burden caused by toxic and ineffective drug for each patient.	6.3% years of life lost to disability per patient per year, the estimated average burden of disease associated with the adverse effects of medical treatments in the US Burden of Disease Study, 2010.	The condition caused by the toxic drug is such that each patient is indifferent between losing 6.3% of each year of healthy life and living with this condition each year. A percentage increase of burden means that each patient experiencing adverse effects would be indifferent to living each year with the adverse effects, or to losing 1% of each year if, for the rest of that year, they could live without the adverse effects.
Expected loss of life caused by toxic effects of the drug.	2 months per patient	We assume the toxic effects of treatment shorten each patient's life by 2 months on average. This can be because in 75% of instances it does not reduce a patient's life at all, and in the remaining 25%, it reduces a patient's life by 8 months.
Expected extended life resulting from effective treatment.	30% of end-stage patient's expected time to death.	If time to death for the distant stage of the cancer is 10 months, we assume effective treatment extends each patient's life by 3 months on average. If 30% of end-stage patient's expected time to death is more than 2.5 years, we set this parameter to 2.5 years.
Expected burden of disease in the extended months of life owing to taking effective treatment.	The same as the current average burden of disease (ie, its disability weight, which ranges from 0, no loss of health, to 1, complete loss of health or death).	We assume the effective treatment only extends life and does not improve the health state of patients compared with their current health state.
Time until adverse effects of a toxic drug are discovered after it is mistakenly approved.	10 years	We assume that if a toxic drug is falsely approved, its adverse effects will be discovered 10 years after the approval and the drug will be taken off the markets. This is a conservative estimate.
Start-up time before patient	1 year	Time before the RCT starts, needed for paperwork, etc. This time is not used for patient accrual.

Abbreviation: RCTs, randomized clinical trials.

Table 2.

Distant-Stage Statistics for the 23 Most Common Cancer Sites in the United States and the Characteristics of Their BDA-Optimal RCTs

| Cancer Site	% Burden of Disease	5-Year Survival	Stage Prevalence	Months Expected Control OS	Target OS Difference	Follow-up Period	Accrual Rate (Patients per Year)	Sample Size	% 1-Sided Alpha
Brain (and other nervous system)	13.4	20.6	2976	38	11	36	100	152	47.9
Breast (only female)	4.2	24.6	178 519	43	13	36	341	478	17.6
Cervix uteri (only female)	6.2	15.7	32 437	32	10	32	132	204	37.4
Colon and rectum	9.1	12.4	233 786	29	9	29	420	506	13.1
Corpus uteri (only female)	5.1	16.1	49 729	33	10	33	157	262	32.1
Esophagus	12.2	4.0	13 597	19	6	19	105	218	34.5
Hodgkin lymphoma	5.1	73.1	73 954	191	30	36	191	1448	12.8
Kidney and renal pelvis	5.6	11.2	60 148	27	8	27	172	296	27.4
Larynx	6.5	33.4	16 882	55	16	36	110	220	42.9
Leukemia	9.0	30.2	47 758	50	15	36	154	318	31.5
Liver and intrahepatic bile duct	9.9	2.9	9132	17	5	17	100	212	34.8
Lung and bronchus	15.6	4.0	233 021	19	6	19	419	548	9.7
Melanoma of the skin	4.5	15.8	39 863	32	10	32	143	234	35.6
Myeloma	13.3	43.1	85 175	71	21	36	207	520	22.5

Abbreviations: BDA, Bayesian decision analysis; OS, overall survival; RCTs, randomized clinical trials.

Table 3.

Comparison of Selected RCTs in the Portfolio of National Cancer Institute's Alliance for Clinical Trials in Oncology and Their Associated BDA-optimal RCTs

Cancer Site	Primary End Point	Control Group Outcome	Stage Prevalence	Target Hazard Ratio	Follow-up Time, Years	Target Accrual Rate	Months Survival Time	Additional Survival	Sample Size
Glioblastoma	OS	Median 21 mo	25 299	0.710	2.0	72	15	6	400
SCLC	OS	Median 23 mo	16 255	0.770	2.5	120	44	13	640
Bladder	OS	Median 13.8 mo	23 096	0.740	4.0	168	20	7	500
Prostate (CR met)	OS	Median 35 mo	111 824	0.770	1.5	400	46	14	1224
NSCLC	OS	Median 5 y	64 769	0.670	6.0	100	87	72	410
CLL	PFS	Median 34 mo	103 611	0.586	2.0	180	73	30	350
Lymphoma	EFS	Median 42 mo	164 888	0.650	3.0	100	115	30	430
Colon	DFS	3-y DFS rate of 72%	319 118	0.790	3.0	800	209	30	2500
Prostate (ES 3-y)	PFS	3-y PFS rate of 57.7%	2 236 474	0.670	3.0	156	240	30	750
Prostate (ES 2-y)	PFS	2-y PFS rate of 80%	2 236 474	0.472	2.0	180	240	30	464

Abbreviations: BDA, Bayesian decision analysis; CLL, chronic lymphocytic leukemia; CR met, castration-resistant metastatic prostate cancer; DFS, disease-free survival; EFS, event-free survival; ES 3-y and ES 2-y, early-stage prostate cancer with 3-year and 2-year follow-up periods; mo, months; NSCLC, non–small-cell lung cancer; OS, overall survival; PFS, progression-free survival; RCTs, randomized clinical trials; SCLC, small-cell lung cancer.

Figure 1.
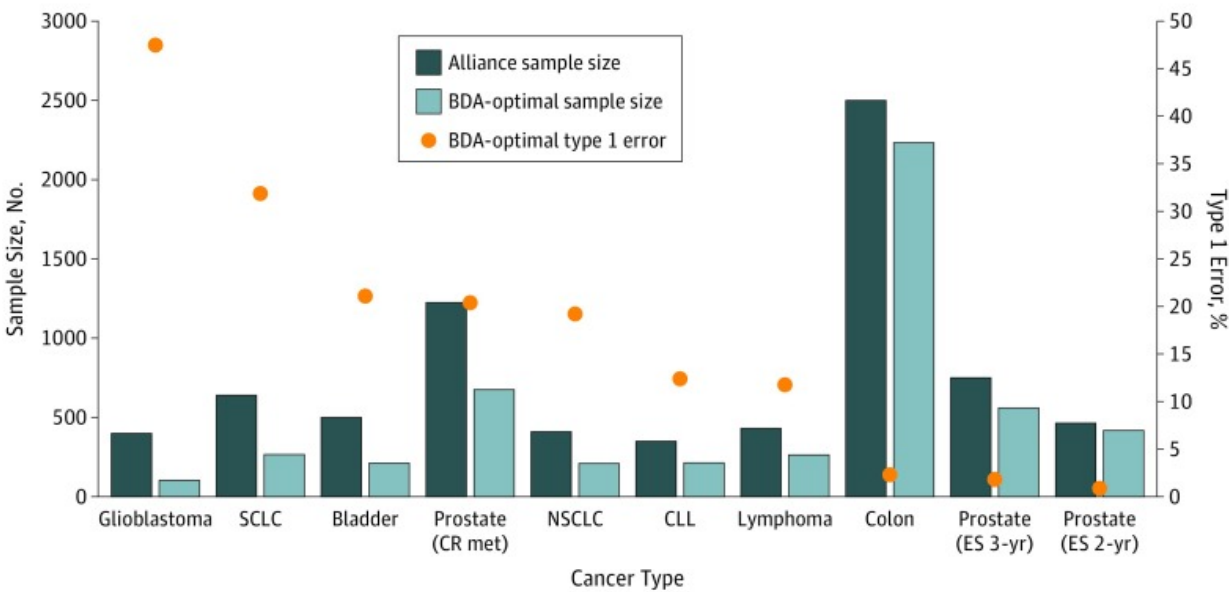
BDA-Optimal Type 1 Errors and Sample Sizes for Alliance Clinical Trials (Alliance Sample Sizes Also Displayed for Comparison)

BDA indicates Bayesian decision analysis; prostate (CR met), castration-resistant metastatic prostate cancer, and prostate (ES 3-yr) and prostate (ES 2-yr), early-stage prostate cancer with 3-year and 2-year follow-up periods, respectively. The BDA-optimal randomized clinical trials have larger type 1 errors for more deadly cancers with no effective therapies and smaller type 1 errors for less serious cancers.

Figure 2.
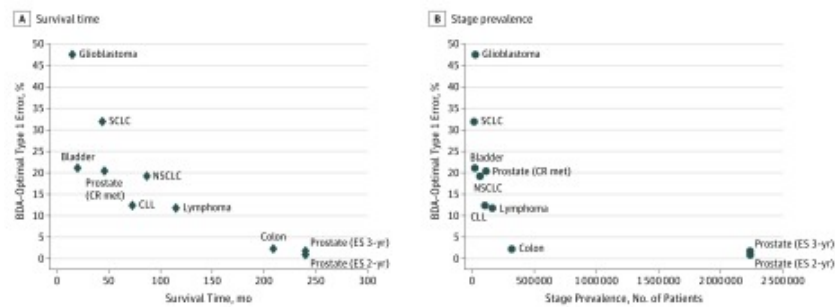
Scatterplot of Survival Time and Stage Prevalence Against BDA-Optimal Type 1 Errors for Alliance Clinical Trials

BDA indicates Bayesian decision analysis; CLL, chronic lymphocytic leukemia; NSCLC, non–small-cell lung cancer; prostate (CR met), castration-resistant metastatic prostate cancer; prostate (ES 3-yr) and prostate (ES 2-yr), early-stage prostate cancer with 3-year and 2-year follow-up periods, respectively; and SCLC, small-cell lung cancer. The BDA-optimal type 1 errors are larger for cancers with shorter survival times and lower prevalence, and smaller for less serious cancers with greater prevalence.