Supplementary Online Content

Montazerhodjat V, Chaudhuri SE, Sargent DJ, Lo AW. Use of bayesian decision analysis to minimize harm in patient-centered randomized clinical trials in oncology [published online April 13, 2017]. *JAMA Oncol*. doi:10.1001/jamaoncol.2017.0123

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 (∆μ_"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 ∆μ_"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 (∆μ_"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 ∆μ_"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.

This supplementary material has been provided by the authors to give readers additional information about their work.

Supplementary Online Content

Montazerhodjat V, Chaudhuri SE, Sargent DJ, Lo AW. Patient-Centered Randomized Clinical Trials in Oncology via Bayesian Decision Analysis.

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 (Δy_{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 $\Delta y_{\text{tox}} = [0\%, 3.15\%, 6.3\%, 9.45\%, 12.6\%]$, 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 ($\Delta\mu_{\text{tox}}$) 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 $\Delta\mu_{\text{tox}} = [4, 3, 2, 1, 0]$ 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 ($\Delta\mu_{\text{tox}}$) 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 $\Delta\mu_{\text{tox}} = [0, 1, 2, 3, 4]$ months. A lower limit of 40 is applied to the sample size to ensure the log-rank statistic is approximately standard normal.

1. Expected RCT Penalty

In this section, we formulate the expected penalty associated with a fixed-sample balanced two-arm randomized clinical trial (RCT) that involves cancer therapeutics. Let C be the cost for a given fixed-sample test with n patients in each arm and a critical value λ_n. Assuming prior probabilities p_0 and $p_1 = 1 - p_0$ for the null ($H_0 \equiv$ toxic drug) and alternative ($H_1 \equiv$ effective drug) hypotheses, respectively, the expected cost is given by

$$\mathrm{E}[C] = p_0\ \mathrm{E}[C \mid H_0] + (1 - p_0)\ \mathrm{E}[C \mid H_1]. \qquad (1)$$

If we further assume exponential distributions for the time to event (death) for each patient given a particular treatment, and for the time until an at least equally effective therapy is discovered, then the conditional expectations in (1) can be decomposed into two components: the expected cost per person given an inferior treatment, and the expected number of patients who receive an inferior treatment.

The cost factors can be formulated as two constants, c_1 and c_2, which represent the expected cost per person of being treated with the toxic drug given the null hypothesis, and the expected cost per person of not being treated with an effective drug given the alternative hypothesis, respectively,

$$c_1 = \mu_{\text{tox}}\ (y_{\text{tox}} - y) + (\mu - \mu_{\text{tox}})\ (1 - y)\ , \qquad (2)$$

$$c_2 = \mu\ (y - y_{\text{eff}}) + (\mu_{\text{eff}} - \mu)\ (1 - y_{\text{eff}})\ , \qquad (3)$$

where μ denotes the life expectancy of patients under the standard treatment, μ_{tox} is the life expectancy given a toxic drug, μ_{eff} is the life expectancy given an effective drug, y is the burden of disease, y_{tox} is the burden of disease given a toxic drug, and y_{eff} is the burden of disease given an effective drug. Intuitively, (2) states that the expected cost of being treated with the toxic drug relative to the standard treatment consists of a deteriorated condition over time μ_{tox}, and a complete loss of life over time $\mu - \mu_{\text{tox}}$. Similarly, (3) states that the expected cost of foregoing the effective drug for the standard treatment is the missed opportunity of an improved condition over time μ, and a disability weighted life extension over time $\mu_{\text{eff}} - \mu$. Note that in our paper we assumed an effective drug only increases life expectancy, and does not improve the patient's current conditions (i.e., $y = y_{\text{eff}}$).

Now consider the expected number of patients who receive the toxic drug under the null hypothesis. We denote a_1, a_2, and a_3 as the expected number of current patients, patients diagnosed during the trial period, and patients diagnosed after the trial period who receive the toxic drug given a type I error, respectively. Defining s, f, and η to be the start-up time before patient enrollment, the follow-up period after enrolling the last patient, and the patient accrual period (i.e., the inverse of the accrual rate), respectively, and assuming uniform (evenly spaced) patient enrollment, the length of the trial period can be calculated as $l = s + f + 2\eta n$. Given a type I error, a current patient will receive the toxic drug if the patient lives beyond the trial period (probability $\exp(-l/\mu)$, where $\exp(\cdot)$ is the exponential function), and a new treatment is not discovered before the end of the trial (probability $\exp(-l/\tau)$, where τ is the expected time until an alternative effective therapy is discovered). Assuming these are independent events,

$$a_1 = N \exp\left(-\frac{l}{\mu}\right) \exp\left(-\frac{l}{\tau}\right), \qquad (4)$$

where N is the disease stage-specific prevalence. Similarly, a patient diagnosed with the disease during the trial phase at time t will receive the toxic drug if the patient lives beyond the trial period (probability $\exp(-(l - t)/\mu)$), and a new therapy is not discovered before the end of the trial. Given an incidence rate I,

$$a_2 = \int_0^l I \exp\left(-\frac{l-t}{\mu}\right) \exp\left(-\frac{l}{\tau}\right) \mathrm{d}t \qquad (5)$$

$$= I\mu\left(1 - \exp\left(-\frac{l}{\mu}\right)\right) \exp\left(-\frac{l}{\tau}\right). \qquad (6)$$

Finally, a patient diagnosed with the disease after the trial phase at time t will receive the toxic drug if a new treatment has not yet been discovered, and the drug has not yet been taken off the markets. Assuming the time until the adverse effects of the toxic drug are discovered after it is mistakenly approved is given by T,

$$a_3 = \int_0^T \frac{It}{\tau} \exp\left(-\frac{t}{\tau}\right) \exp\left(-\frac{l}{\tau}\right) \mathrm{d}t \qquad (7)$$

$$= I\tau\left(1 - \left(1 + \frac{T}{\tau}\right) \exp\left(-\frac{T}{\tau}\right)\right) \exp\left(-\frac{l}{\tau}\right). \qquad (8)$$

Since the sample size is fixed, we find that n in-trial patients always receive the toxic drug, and therefore the expected cost under the null hypothesis is given by

$$\mathrm{E}[C \mid H_0] = c_1 \cdot [\Phi(-\lambda_n)(a_1 + a_2 + a_3) + n], \qquad (9)$$

where $\Phi(-\lambda_n)$ is the probability the drug is approved (Φ is the cumulative distribution function of a standard normal random variable).

In a similar fashion, we now consider the alternative hypothesis, and the expected number of patients who do not receive the effective drug and die before an alternative therapy is discovered. We denote b_1 as the expected number of patients who satisfy this criterion given a type II error. In the case where the effective drug is mistakenly rejected, any patient diagnosed with the disease before an alternative therapy is discovered will miss the opportunity to take the effective drug, and therefore

$$b_1 = \frac{\tau}{\tau + \mu}\left[N + \int_0^\infty I \exp\left(-\frac{t}{\tau}\right) \mathrm{d}t\right] \qquad (10)$$

$$= \frac{\tau}{\tau + \mu}(N + I\tau), \qquad (11)$$

where $\tau/(\tau + \mu)$ is the probability that the patient dies before an alternative effective therapy is

discovered. Even when the drug is correctly approved, a number of patients will not receive the effective drug and die before an alternative therapy is found because of the trial's delay. Specifically, a current patient would be negatively affected by the delay if either an alternative therapy was developed before the end of the trial period and they died before this alternative therapy was released, or an alternative therapy was not developed by the end of the trial period, and they died before the trial's effective drug was released. The expected number of current patients affected by the delay given no type II error, b_2, is then

$$b_2 = \int_0^l \int_v^\infty N \cdot \frac{1}{\tau} \exp\left(-\frac{\mu}{\tau}\right) \cdot \frac{1}{\mu} \exp\left(-\frac{v}{\mu}\right) \mathrm{d}u \, \mathrm{d}v \tag{12}$$

$$= N \cdot \frac{\tau}{\tau + \mu} \left(1 - \exp\left(-\frac{l}{\tau}\right) \exp\left(-\frac{l}{\mu}\right)\right). \tag{13}$$

Similarly, a patient diagnosed with the disease during the trial phase at time t would be negatively affected by the delay if they died before an alternative therapy was discovered, and either this alternative therapy was developed after time t and before the end of the trial, or they died before the end of the trial period. The expected number of future patients affected by the delay given no type II error, b_3, can then be formulated as

$$b_3 = \int_0^l \int_0^{l-t} \int_{v+t}^\infty I \cdot \frac{1}{\tau} \exp\left(-\frac{\mu}{\tau}\right) \cdot \frac{1}{\mu} \exp\left(-\frac{v}{\mu}\right) \mathrm{d}u \, \mathrm{d}v \, \mathrm{d}t \tag{14}$$

$$= I \cdot \frac{\tau}{\tau + \mu} \left(\tau \left[1 - \exp\left(-\frac{l}{\tau}\right)\right] - \mu \left[\exp\left(-\frac{l}{\tau}\right)\left(1 - \exp\left(-\frac{l}{\mu}\right)\right)\right]\right). \tag{15}$$

Collecting terms, the expected cost under the alternative hypothesis is given by

$$\mathrm{E}[C \mid H_1] = c_2 \cdot [\Phi(\lambda_n - \delta_n) \cdot b_1 + [1 - \Phi(\lambda_n - \delta_n)] \cdot (b_2 + b_3)], \tag{16}$$

where $\Phi(\lambda_n - \delta_n)$ is the probability of a type II error, and δ_n is the mean of the log-rank statistic in the Cox Proportional Hazard regression under the alternative hypothesis. We have the following expression for δ_n:

$$\delta_n = -\frac{1}{2} \log(r) \sqrt{\sum_{k=0}^{1} \sum_{i=0}^{n-1} d_{i,k}}, \tag{17}$$

where r denotes the hazard ratio, μ/μ_{eff}, and $d_{i,k}$ is the probability that a subject in trial arm k will suffer an event (death) during the observation period. Subjects in the control arm ($k = 0$) have a shorter life expectancy than subjects in the experimental arm ($k = 1$) who receive the effective drug. Therefore $d_{i,1} = 1 - \exp\left(-\frac{o_{i,1}}{\mu}\right)$, and $d_{i,2} = 1 - \exp\left(-\frac{o_{i,2}}{\mu_{\text{eff}}}\right)$, where $o_{i,k} = f + \frac{2i+k}{2n-1} \cdot 2\eta n$ is the observation period for subject i in trial arm k.

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

In the control arm of the study, patients are administered the standard of care for the late stage of the specific cancer, whereas an investigational drug is administered to the patients in the treatment arm. We assume that the enrolled patients are randomized to either arm with equal probability, and the patients' enrollments in the RCT are uniformly spread across an enrollment period. The duration of the enrollment period is determined by the patient accrual rate for each cancer type, which increases linearly with the U.S. prevalence of the distant stage of the cancer, and is bounded between 100 and 800 patients per year for all the studied cancer sites. We further assume that the time required for the study setup before the first patient enrollment is 1 year, and that the time between the enrollment of the last patient and the final data analysis is set to either the expected survival time of the patients in the control arm or 3 years, whichever is lower.

Hence, the follow-up time for patients who enroll in the RCT earlier is longer compared to the patients who enroll at a later time in the enrollment period. We assume that the Cox proportional hazards model is used to analyze the time-to-event data at the time of analysis.

Assume there is a 35/65 chance for the investigational drug to be effective or ineffective and toxic, respectively (this prior distribution is based on historical averages[14,15]). Note that baseline statistics and pre-trial information for a given cancer site and therapy can be used to further calibrate these parameters. We analyze the robustness of our results to this and other key model parameters in the next section.

If the drug is effective, it is assumed to extend the patient's life by 30% of the expected survival time for patients with the distant stage of the cancer—or equivalently, a hazard ratio of 76.9% denoting the alternative hypothesis. To avoid extremely long survival extensions, we limit the gain in survival due to the effective drug to 2.5 years (30 months). We also assume that patients in this extended time before their death will experience the average disease burden that patients currently experience due to each cancer in the U.S. In other words, if the drug is effective, it only increases a patient's life by a certain number of months, and it does not improve the state of health of patients before death.

If the drug is ineffective, it is assumed to shorten each patient's life by 2 months on average, and to have side effects with a burden of 6.3% per year, the estimated average burden of disease associated with the adverse effects of medical treatments in the U.S. Burden of Disease Study 2010.[16] This level of 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. We assume the worst-case scenario for the side effects; i.e., the adverse effects are persistent until the patient's death, and there will be no treatment to alleviate these side effects.

We assume that within each cancer type, at each point in time, there is an expected 10-year period before a drug at least as effective as the assumed investigational agent is discovered. More precisely, we assume an exponential distribution with a mean of 10 years for the time until a new, equally effective or better drug is discovered for each cancer site. We also assume that if a toxic drug is incorrectly approved, its side effects will be discovered 10 years after the approval, and that the approved drug will then be taken off markets. This is a relatively conservative assumption, since dangerous side effects are often discovered significantly more quickly in practice. Finally, we assume that due to practical considerations, the power of the trial is always maintained at or below 90%. For a complete list of assumptions on the RCT setting, refer to Table 1. A full derivation of our expected cost model is provided in Section 1 of this Supplement.

3. Sensitivity Analysis

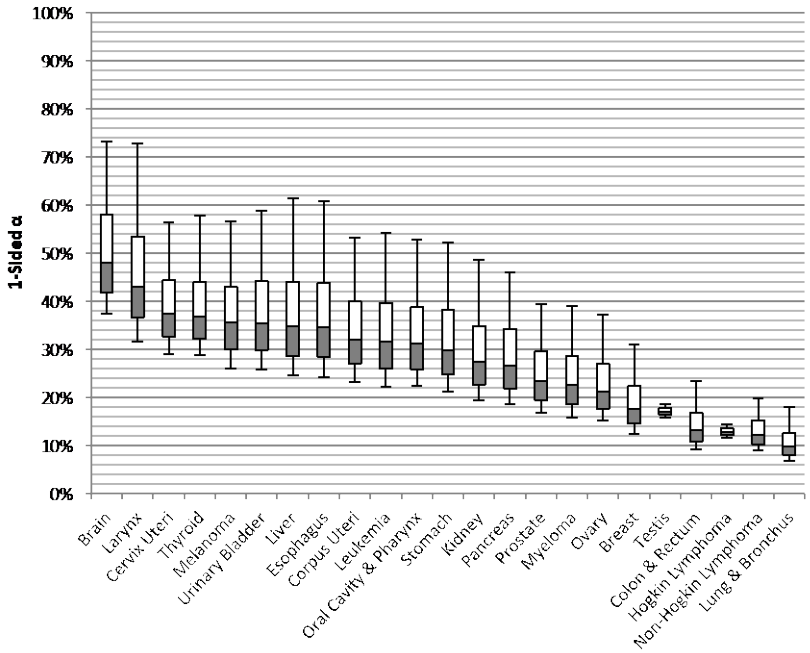
In this section, we investigate the robustness of our results to the underlying assumptions in our model. For each of the 23 cancer sites in our study, we determine the optimal balanced two-arm fixed-sample RCT for testing a therapy targeting the late stage of the cancer, where the endpoint is overall survival. We then vary the patient accrual rate, the a-priori probability of an effective therapy, and the toxicity of an ineffective drug about their proposed values, and obtain new optimal fixed-sample RCT designs for each value. The optimal α and sample size values associated with the perturbed parameters are shown in Figures 1–8.

In general, we find that cancers with poor prognoses consistently have relatively large BDA-optimal type I error rates (α), and small optimal RCT sample sizes. Our observation that a patient with a poor prognosis cannot afford to miss any effective drugs, even at the expense of assuming substantial risks, is therefore robust over a wide range of conditions. Moreover, all the type I error rates recommended by the BDA analysis remain far in excess of the traditional 2.5% one-sided α. However, the specific critical value and sample size of each optimal RCT is quite sensitive to the underlying assumptions. For example, a 15% decrease in the a-priori probability of an effective therapy from 35% to 20% reduces the optimal α value for brain cancer RCTs from 48% to 19%, and increases the optimal sample size 76% from 152 to 268 (see Figures 3–4). Conversely, decreasing either the patient accrual rate, or the toxicity of an ineffective therapy leads to less conservative—i.e., larger α, and smaller sample size—RCT designs (see Figures 1–2, and 5–8). Intuitively, decreasing the patient accrual rate increases the trial length, and for patients with short life expectancies, the optimal trade-off attempts to retain a relatively short trial length. Similarly, decreasing the toxicity of an ineffective drug under the null hypothesis reduces the cost of a more aggressive RCT design.

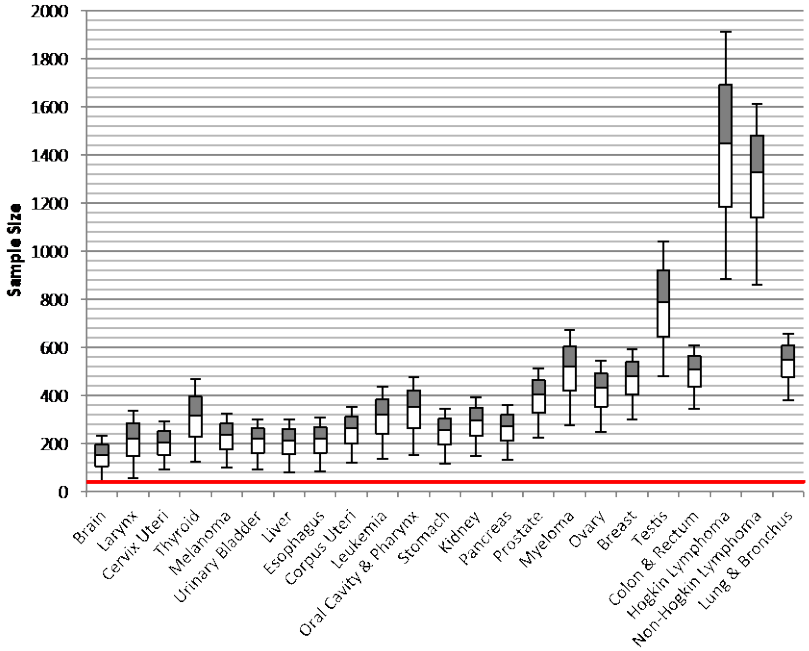
This final observation, when taken to its limit, affects the BDA-optimal RCT design substantially. Specifically, if either Δy_{tox} or $\Delta \mu_{\text{tox}}$ are set to 0% or 0 months, respectively, then the BDA-optimal RCT design becomes extremely aggressive and the protocol approves the majority of investigational drugs after minimal clinical trial study. In this case, there are few benefits gained by rejecting an ineffective drug,

mitigating the trade-off central to the expected cost optimization. Note that a non-toxic therapy in this model is one that is equally as effective as the standard treatment, and therefore should be considered a limiting case. This last example highlights the need for carefully considered assumptions and accurately calibrated cost models when implementing the BDA-framework.

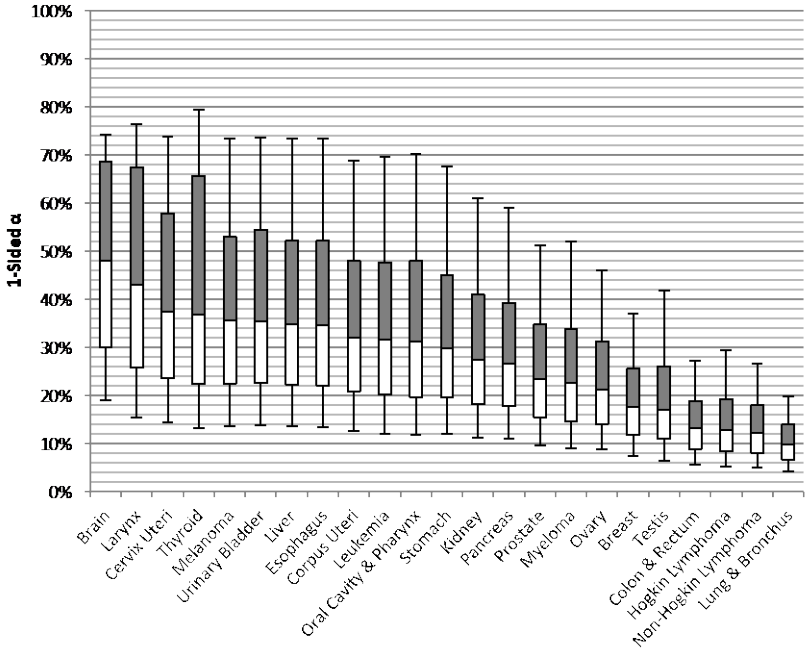
eFigure 1. Sensitivity of the BDA-optimal 1-sided α to the accrual rate for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to [150%, 125%, 100%, 75%, 50%] of the accrual rate proposed in our study for each cancer.



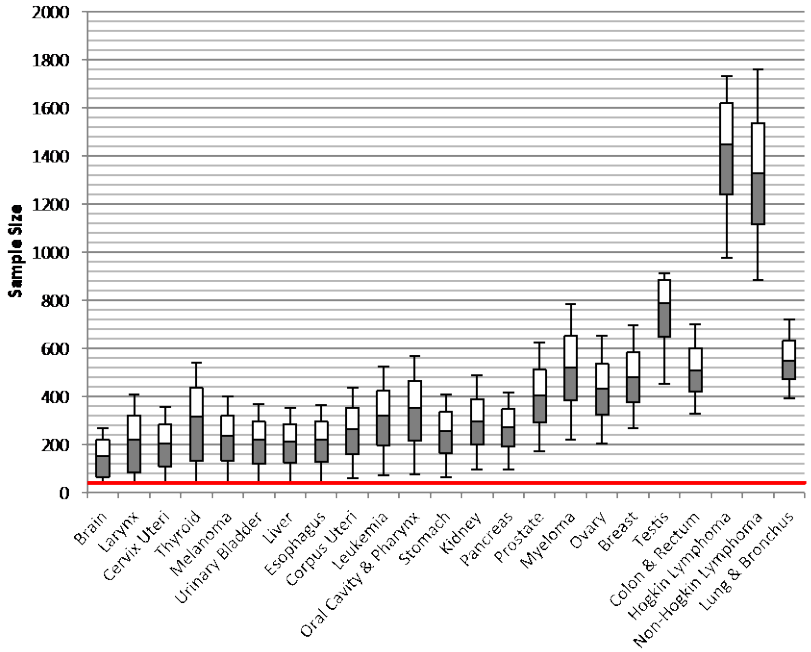
eFigure 2. Sensitivity of the BDA-optimal sample size to the accrual rate for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to [50%, 75%, 100%, 125%, 150%] of the accrual rate proposed in our study for each cancer. A lower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



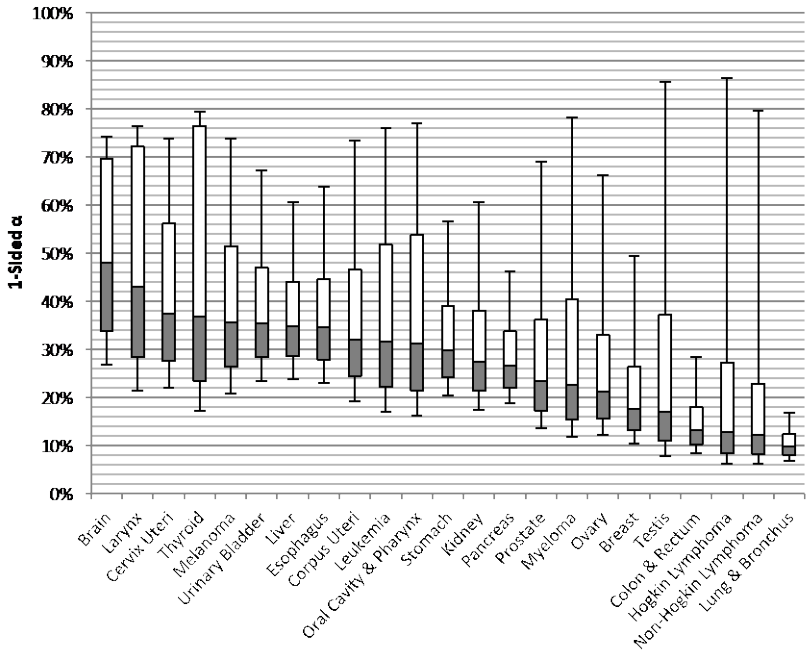
eFigure 3. Sensitivity of the BDA-optimal 1-sided α to the probability that the investigational drug is effective (p_1) for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $p_1 = [20\%, 27.5\%, 35\%, 42.5\%, 50\%]$.



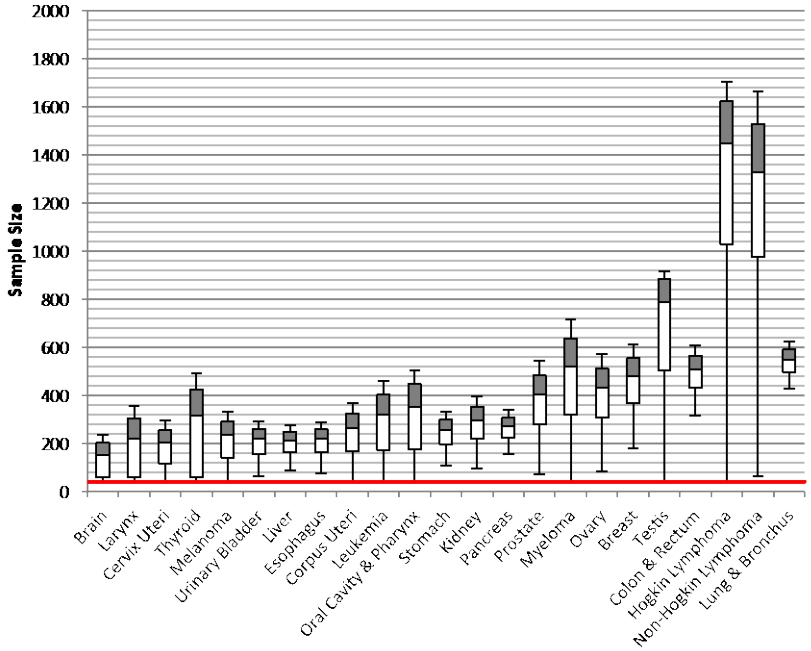
eFigure 4. Sensitivity of the BDA-optimal sample size to the probability that the investigational drug is effective (p_1) for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $p_1 = [50\%, 42.5\%, 35\%, 27.5\%, 20\%]$. A lower limit of 40 is applied to the sample size to ensure the log-rank statistic is approximately standard normal.



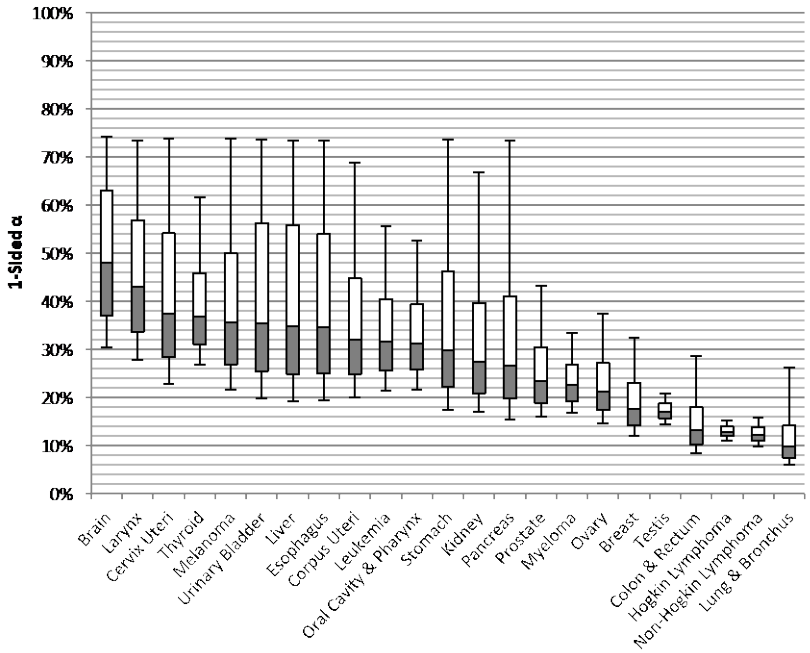
eFigure 5. Sensitivity of the BDA-optimal 1-sided α to the side effect level of burden of an ineffective drug (Δy_{tox}) under the null hypothesis for the 23 most common cancer sites in the U.S. From the lower to upper end of each box plot, the five-number summary corresponds to $\Delta y_{\text{tox}} = [12.6\%, 9.45\%, 6.3\%, 3.15\%, 0\%]$, 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 (Δy_{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 $\Delta y_{\text{tox}} = [0\%, 3.15\%, 6.3\%, 9.45\%, 12.6\%]$, 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 ($\Delta\mu_{\text{tox}}$) 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 $\Delta\mu_{\text{tox}} = [4, 3, 2, 1, 0]$ 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 ($\Delta\mu_{\text{tox}}$) 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.