3. Results

The PPTC has generally used sample sizes of 8 or 10 mice per treatment group. However, given the broad diversity of subtypes within most cancers and their potential responses to treatment, any systematic screen of new drugs must encompass a range of genotypic classes. This can put a significant burden on the pace of discovery. An alternative approach is to use fewer animals per tumor type, which allows testing a greater number of PDXs and therefore captures greater molecular and genetic heterogeneity. Murphy et al. (2016) retrospectively analyzed 67 agents evaluated by the PPTP to determine if the use of single-mouse experiments would reliably recapitulate the median objective response category of the group of 8 to 10 mice they found that a single tumor response accurately predicted the group median response 75% of the time.

In our current study, we update these findings with a systematic analysis of in vivo testing results from 2015-2018 within the PPTC. First, we estimate the power to detect an in-fold increase in time-to-event (defined below) across a range of tumor growth characteristics. Second, we evaluate these results from 2015-2016 data using the methodology of Murphy et al.

2. Methods

Data were compiled from experiments on 30 anti-cancer agents from 2015-2018, independent of the experiments reported by Murphy et al. Each agent was tested against 1 to 31 PDXs (median, 9) from a consortium-wide panel of 112 PDXs.

Time-to-Event

We evaluated time-to-event across control and treated mice, with event defined as quadrupling of tumor volume in solid tumors, mouse death, or loss of measurable tumor mass in liquid tumors.

Objective Response Categories

Objective response was assessed by the Investigational New Drug (IND) Study Committee of the Pediatric Preclinical Testing Program (PPTP). Each response is defined as:

- Complete Response (CR): disappearance of measurable tumor mass (both caliper measurements ≤ 0.1 mm) during study period.
- Partial Response (PR): measurable tumor mass reduction of 25% to 50%.
- Stable Disease (SD): measurable tumor mass reduction of ≤ 25% or growth ≥ 25%.
- Progressive Disease (PD1): mouse’s time-to-event is > twice of the Kaplan-Meier (KM) median time-to-event in control group.
- Progressive Disease (PD2): mouse’s time-to-event is > twice of the Kaplan-Meier (KM) median time-to-event in control group.

Table 1: Summary of Power Analysis for α = 0.05. Under both the log-normal and gamma models, a sample size of 5 mice per group yielded 87% power to detect a T/C of 2 assuming a CV of 0.3. Over 60% of tested PDXs have a CV lower than this. Our results suggest that for most PPTC PDXs, power for detecting major growth delay effects can be maintained while employing fewer mice than used previously.

![Bar chart showing the percentage of tested PDXs with different CV values and sample sizes.](chart)

Figure 1. Observed mean vs. standard deviation for time-to-event across 112 PDX models. Each mouse was untreated controls within each experiment. The slopes of different Coefficients of Variation (CV, standard deviation/mean) are also plotted for reference. Overall, the median CV was 0.257 with an interquartile range of 0.172 to 0.363.

Table 1: Summary of Power Analysis for α = 0.05. Under both the log-normal and gamma models, a sample size of 5 mice per group yielded 87% power to detect a T/C of 2 assuming a CV of 0.3. Over 60% of tested PDXs have a CV lower than this. Our results suggest that for most PPTC PDXs, power for detecting major growth delay effects can be maintained while employing fewer mice than used previously.

4. Discussion

Our results suggest that for most PPTC PDXs, the statistical power to detect major growth delay effects can be maintained while employing fewer mice than used previously. We have also replicated the results of Murphy et al. that when single-mouse experiments return the median objective response category in 75% or greater tested mice, and can therefore be effective experimental design when one wishes to screen an agent (or combinations of agents) across a large number of PDXs. The PPTC is committed to increasing sample sizes to 1, 2, and 3-mouse experimental designs where appropriate, and will also further investigate the appropriateness of lower sample sizes for their traditional experimental designs.

5. Reference

More Information

*Corresponding author: Stephen W. Erickson (serickson@rti.org)

Presented at: American Association for Cancer Research (AACR)

Supported by National Cancer Institute Grant CA199222

www.scpptc.org

[References cited in the document]


Figure 2. Observed skewness vs. coefficient of variation across 301 untreated, control groups, representing 112 unique PDX models (Panel A), and across 149 treated groups, representing 23 unique anti-cancer agents and 48 unique PDX models which had experienced a median T/C ratio between 2 and 5 inclusive (Panel B). Two lines show the theoretical relationship between skewness and CV in log-normal and gamma distributions; overall, both controls and treated experimental groups show a positive relationship between skewness and CV.

Figure 3. Histogram of individual mouse objective response (OR) deviations from their median group response. Median OR was calculated across treated mice within an experiment and was recorded as “0”. Individual mice within an experiment were scored according to how many OR categories away they were from the median. For example, a score of “-1” could represent an individual mouse with OR of Complete Response (CR) but the median for that treatment group was Maintained Complete Response (MCR). Single tumor response accurately predicts the group median response 82.0% of the time, and is within one category 93.5% of the time.