Supplemental Response Definitions

Response and Event Definitions for Solid Tumor Xenograft Models

Response: For individual mice, progressive disease (PD) was defined as < 50% regression from initial volume during the study period and > 25% increase in initial volume at the end of study period. Stable disease (SD) was defined as < 50% regression from initial volume during the study period and ≤ 25% increase in initial volume at the end of the study. Partial response (PR) was defined as a tumor volume regression ≥50% for at least one time point but with measurable tumor (≥ 0.10 cm^3). Complete response (CR) was defined as a disappearance of measurable tumor mass (< 0.10 cm^3) for at least one time point. A complete response was considered maintained (MCR) if the tumor volume was <0.10 cm^3 at the end of the study period. For treatment groups only, if the tumor response was PD, then PD was further classified into PD1 or PD2 based on the tumor growth delay (TGD) value. TGD values were calculated based on the numbers of days to event. For each individual mouse that had PD and had an event in the treatment groups, a TGD value was calculated by dividing the time to event for that mouse by the median time to event in the respective control group. Median times to event were estimated based on the Kaplan-Meier event-free survival distribution. If a mouse had a TGD value ≤ 1.5, that mouse was considered PD1. If the TGD value was > 1.5, the mouse was considered PD2. Mice that had PD but did not have an event at the end of the study were coded as PD2.

Event-free survival: An event in the solid tumor xenograft models was defined as a quadrupling of tumor volume from the initial tumor volume. Event-free survival was defined as the time interval from initiation of study to the first event or to the end of the study period for tumors that did not quadruple in volume. The time to event was determined using interpolation based on the formula:

$$t_x = t_1 + (t_2 - t_1)\ln(V_e / V_1) / \ln(V_2 / V_1),$$

where t_x is the interpolated day to event, t_1 is the lower observation day bracketing the event, t_2 is the upper observation day bracketing the event, V_1 is the tumor volume on day t_1, V_2 is the tumor volume on day t_2 and V_e is the event threshold (4 times initial tumor volume for solid tumor xenografts).

Response and Event Definitions for Acute Lymphoblastic Leukemia (ALL) Xenograft Models

Individual mice were categorized as PD if their percentage of hCD45 cells never dropped below 1% and they had an event before the end of the study period. An event is defined as hCD45 cells above 25% in the peripheral blood with times to event calculated as above. Individual mice were classified as SD if their percentage of hCD45 cells never dropped below 1% and no event occurred before the end of the study. PR was assigned if the percentage of cells dropped below 1% for any one time point regardless of whether the percentage eventually reached 25%. A CR was assigned if the percentage of hCD45 cells dropped below 1% for 2 consecutive weeks of the study and regardless of whether the percentage reached 25% or not. A CR was considered maintained if the percentage of

hCD45 was less than 1% for the last three measurements of the study. For treatment groups, PD was further classified into PD1 and PD2 according to the TGD value.

The time to event was determined using interpolation based on the formula:

$$t_x = t_1 + (t_2 - t_1)\ln(V_e / V_1) / \ln(V_2 / V_1),$$

where t_x is the interpolated day to event, t_1 is the lower observation day bracketing the event, t_2 is the upper observation day bracketing the event, V_1 is the hCD45 percentage on day t_1, V_2 is the tumor volume (or hCD45 percentage) on day t_2 and V_e is the event threshold (25% for ALL xenografts).

Summary statistics and analysis methods

Overall Group Response: Each individual mouse was assigned a score from 0 to 10 based on their response: PD1=0, PD2=2, SD=4, PR=6, CR=8, and MCR=10, and the median for the group determined the overall response. Studies in which toxicity was greater than 25% or in which the control group was not at least SD, were considered inevaluable and were excluded from analysis. Treatment groups with PR, CR, or MCR are considered to have had an objective response. Agents inducing objective responses are considered highly active against the tested line, while agents inducing stable disease or PD2 are considered to have intermediate activity, and agents producing PD1 are considered to have a low level of activity against the tested line.

Tumor Volume T/C value: Relative tumor volumes (RTV) for control (C) and treatment (T) mice were calculated at day 21 or when all mice in the control and treated groups still had measurable tumor volumes (if less than 21 days). The mean relative tumor volumes for control and treatment mice for each study were then calculated and the T/C value was the mean RTV for the treatment group divided by the mean RTV for the control group. For the tumor volume T/C response measure, agents producing a T/C of $\leq$ 15% are considered highly active, those with a mean tumor volume T/C of $\leq$ 45% but > 15% are considered to have intermediate activity, and those with mean T/C values > 45% are considered to have low levels of activity [1].

EFS T/C value: An EFS T/C value was defined by the ratio of the median time to event of the treatment group and the median time to event of the respective control group. If the treatment group did not have a median time to event, then EFS T/C was defined as greater than the ratio of the last day of the study for the treatment group divided by the median time to event for the control group. For the EFS T/C measure, agents are considered highly active if they meet three criteria: a) an EFS T/C > 2; b) a significant difference in EFS distributions ($p \leq 0.050$), and c) a net reduction in median tumor volume for animals in the treated group at the end of treatment as compared to at treatment initiation. Agents meeting the first two criteria, but not having a net reduction in median tumor volume for treated animals at the end of the study are considered to have intermediate activity. Agents with an EFS T/C < 2 are considered to have low levels of activity. Xenografts in which the median EFS for the control line

was greater than one-half of the study period or in which the median EFS for the control line did not exist are considered not evaluable for the EFS T/C measure of activity.

1. Plowman J CR, Alley M, Sausville E, Schepartz S. US-NCI testing procedures. In: Feibig HH BA, editor. Relevance of tumor models for anticancer drug development. Basel: Karger.; 1999. p 121-135.