APPENDIX A

ADDITIONAL GUIDANCE FOR COMPLETING SECTION D, "BIOMARKER STATISTICAL PLAN"

The Biomarker Statistical Plan may require the following elements:

- Statistical methods for the primary analyses (e.g., Cox proportional hazards regression, conditional or unconditional logistic regression, etc.).
- Which biomarkers will be employed, and how they will be included in the analyses.
- If cutpoints will be used, specify the cutpoint(s). Provide the rationale for the cut-point(s) selected. What proportion of subjects is expected to have values above and below the proposed assay value cut-points? What magnitude of effect (e.g., treatment benefit) or outcome (e.g., prognosis) is expected for patients with assay results above and below the proposed cutpoint(s)?
- Criteria and metrics for defining significant changes (e.g., between timepoints, between responders and non-responders).
- Scoring system(s).
- Variable selection procedures.
- Will the analysis involve validation? (for example, of a prognostic signature)
- List of standard clinical variables to be incorporated into models or other analyses.
- Multiple-comparisons adjustment methods.
- Transformations applied to variables.

Note: If the trial objectives include an evaluation of the association of an integral marker with a new clinical endpoint or factor not previously studied, please explain how the magnitude of the association or effect will be measured and provide power calculations for any statistical tests that are planned.

Sample size(s) and rationale: Please provide a sample size and rationale. In general, power calculations are required for late-phase trials such as phase III and large phase II trials. Power calculations may not be needed for all early-phase trials, especially small trials; even so, it can still be useful to calculate the power to detect a given effect size with the known samples available.

> **Note:** The sample size rationale should include a clear explanation (or cited reference) for the method of sample size determination along with a statement of all assumptions required to perform that calculation so that an independent statistician would be able to reproduce the estimates from the information provided in the application.
>
> Typically, a sample size estimate will require assumptions about the following:
>
> - Anticipated distribution of marker values in the targeted population(s) (e.g., marker positivity rate if the marker is dichotomous)
> - Anticipated assay success rates.
> - Event rates or number of events anticipated.
> - Expected effect size(s).
>
> These assumptions and estimates need to be supported by preliminary data or previous studies that should be described either in this section or in the background section.

Example Statistical Section for a Small Study

- **Trial's primary analysis & study design:** Phase II trial with n=31 patients looking for an ORR of 10% vs 30% (one-sided alpha=0.05, power=86.5%).
- **Integrated biomarker:** PD-L1 singleplex IHC assessment. Specimens are collected at baseline.
- **Clinical endpoints to be used in the correlative analysis:** The primary endpoint will be proportion of patients experiencing an overall response (CR or PR) at 6 months. Response evaluation will be determined using the RECIST criteria.
- **Statistical analysis plan for biomarker analysis:**
    - All patients with available tumor samples will be included in this analysis.
    - PD-L1 will be described as percentage of cells showing any positive staining. Additional analyses will look at PD-L1 as a binary endpoint with cut points of 1% and 50%.
    - Analyses will be descriptive in nature. Box plots describing PD-L1 by response group and average PD-L1 % staining will be compared for patients experiencing a response and not experiencing a response. A logistic regression will be employed to correlate binary ORR with continuous PD-L1%. Additional exploratory analyses will assess proportion of patients experiencing an ORR in PD-L1 low and PD-L1 high groups, as defined by both a 1% and 50% cut point. A two-sided Fisher's Exact test will be used to compare ORR rate in the low versus high groups.

Example Statistical Section for a Large Study

- **Trial's primary analysis & study design:** Randomized (1:1) phase III trial of n=360 patients. Original study aimed to use a logrank test & has 82% power to detect a 2-sided test corresponding to 5 years median OS versus 8.33 years median OS (corresponding to an HR of 0.60). Accrual time was 2 years with 3 years additional follow-up.
- **Integrated biomarker:** PD-L1 singleplex IHC assessment. Specimen are collected at baseline.
- **Clinical endpoints to be used in the correlative analysis:** The primary endpoint will be PFS, where PFS is defined as time from randomization to time of first progression.
- **Statistical analysis plan for biomarker analysis:**
    - All patients with available FFPE specimen will be used in the analysis. It is assumed that 90% of patients will have specimen available/evaluable, for a total of n=324 specimen. PD-L1 positivity will be determined by the pre-specified cut point 1%.
    - The integrated analysis aims to assess whether PD-L1 is a predictive biomarker for experimental therapy. As such, the primary analysis will assess interaction between treatment and PD-L1 positivity. Descriptive analyses will display Kaplan-Meier curves comparing each treatment group in PD-L1 negative and positive patients, separately. Cox PH models will be used to assess the interaction between treatment and PD-L1 positivity, with PFS as the endpoint. Additional analyses will adjust for other known prognostic factors such as age and disease stage.
- **Power calculation** (method of Peterson and George): Assuming 2 years accrual, 3 years follow-up and exponential failure, a 1-sided 0.05 test has 90% power to detect an interaction in n=324 patients. This is based on previous information [citation] that there is a 70% rate of PD-L1 positivity by the cut point 1%, and assuming hazard rates of .2 and .05 on the experimental treatment in PD-L1 negative and positive subgroups, respectively. The HR for each stratum is assumed to be 1 in the PD-L1 negative group and .33 in the PD-L1 positive group.