This study J1S-MC-JV01 (NCT04145349) is a sub-study of Master Protocol J1S-MC-JAAA (NCT05999994)

Statistical Analysis Plan Version 3

A Randomized, Open-Label Phase 1/2 Study Evaluating Ramucirumab in Pediatric Patients and Young Adults With Relapsed, Recurrent, or Refractory Desmoplastic Small Round Cell Tumor

NCT04145349

Approval Date: 19-Jan-2022

1. Statistical Analysis Plan: J1S-MC-JV01: A Randomized, Open-Label Phase 2 Study Evaluating Ramucirumab in Pediatric Patients and Young Adults with Relapsed, Recurrent, or Refractory Desmoplastic Small Round Cell Tumor and J1S-MC-JV02: A Randomized, Open-Label Phase 2 Study Evaluating Ramucirumab in Pediatric Patients and Young Adults with Relapsed, Recurrent, or Refractory Synovial Sarcoma

Ramucirumab (LY3009806) Desmoplastic Small Round Cell Tumor and Synovial Sarcoma

J1S-MC-JV01: A Randomized, Open-Label Phase 2 Study Evaluating Ramucirumab in Pediatric Patients and Young Adults with Relapsed, Recurrent, or Refractory Desmoplastic Small Round Cell Tumor and J1S-MC-JV02: A Randomized, Open-Label Phase 2 Study Evaluating Ramucirumab in Pediatric Patients and Young Adults with Relapsed, Recurrent, or Refractory Synovial Sarcoma

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Statistical Analysis Plan Version1 electronically signed and approved by Lilly: 10-Dec-2018
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 13-Sep-2021
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on date provided below.

Approval Date: 19-Jan-2022 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

SAP Version 1 was approved prior to first patient visits for studies J1S-MC-JV01 and J1S-MC-JV02.

SAP Version 2 was approved prior to the interim futility analysis. The overall changes and ratioale for changes incorporated in Version 2 are as follows: Per communications with the FDA, an analysis plan was developed for the scenario when no or limited real-world evidence is available at interim. Per communications with the FDA, a sensitivity analysis using the complete-case population was added. Real-world evidence may be excluded from the analysis if both matching methods are unsuccessful in identifying a relevant cohort of historical patients.

SAP Version 3 was approved prior to database lock for interim futility analysis. Added "Missing" factor level to tumor size matching variable. Changed levels of lines of therapy matching variable to "1st line", "2nd line and above", and "Missing." Expanded eligible regimens in real-world evidence analysis. SAP Version 3 was approved prior to database lock for interim analysis.

4. Study Objectives

4.1. Primary Objective

The primary objective of Studies J1S-MC-JV01 (Study JV01) and J1S-MC-JV02 (Study JV02) is to evaluate efficacy on the basis of progression-free survival (PFS) per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) for:

- Desmoplastic small round cell tumor (DSRCT) (JV01): ramucirumab 12 mg/kg once every 2 weeks (Q2W) in combination with vinorelbine (25 mg/m^2 on Days 1, 8, 15 of a 28-day [4 weeks] cycle) and cyclophosphamide (25 mg/m^2 once daily [QD]) versus vinorelbine/cyclophosphamide alone
- Synovial sarcoma (SS) (Study JV02): ramucirumab 9 mg/kg on Days 1 and 8 every 3 weeks in combination with gemcitabine (900 mg/m^2 on Days 1 and 8 of a 3-week cycle) and docetaxel (75 mg/m^2 on Day 8 of a 3-week cycle) versus gemcitabine/docetaxel alone

4.2. Secondary Objectives

- Safety and tolerability (serious adverse events [SAEs], adverse events [AEs], safety laboratory assessments)
- Additional efficacy measures (overall response rate [ORR], duration of response [DoR], complete response [CR])
- Pharmacokinetics (PK) of the above ramucirumab combinations in children and young adults (maximum drug concentration and minimum drug concentration)
- Immunogenicity (incidence of immunogenicity)

4.3. Exploratory Objectives

- Overall survival (OS)
- Difference in proportion of patients who become eligible for surgical resection of lesions due to documented tumor response while on study therapy
- PFS2
- Biomarkers

5. Study Design

5.1. Study Design

5.1.1. Summary of Study Design

Studies JV01 and JV02 are randomized, multicenter, Phase 2 studies in pediatric and young adult patients with relapsed, recurrent, or progressive DSRCT (Study JV01) and SS (Study JV02) to evaluate the efficacy associated with ramucirumab in combination with chemotherapy versus chemotherapy alone. The primary endpoint of each study is PFS per RECIST v1.1, analyzed via a Bayesian hierarchical model that allows (1) adaptive borrowing on effect-size (log hazard ratio) between Studies JV01 and JV02 and (2) augmenting with historical control data via the use of informative prior distributions constructed from real-world (RW) control outcomes. See Section 6.6.1.2 for full mathematical detail regarding the Bayesian hierarchical model specification. In each study, a total of 30 patients will be randomized at a ratio of 2:1 to receive ramucirumab in combination with a tumor-specific chemotherapy backbone versus chemotherapy alone. A schematic diagram of the studies and their statistical/timing linkages is provided in Figure 5.1.

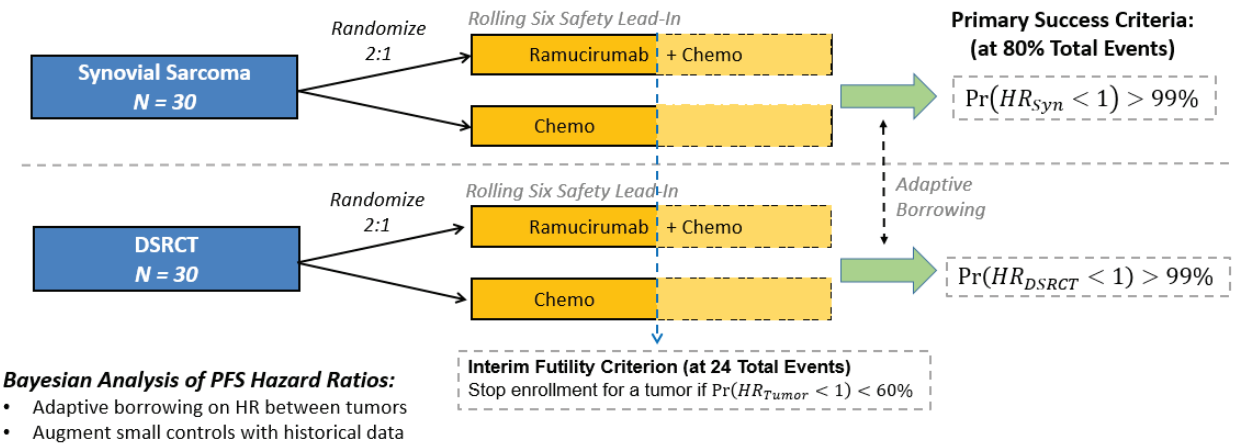


Figure 5.1. JVAA study design.

Safety Lead-in Period: To assess excessive toxicity associated with the experimental ramucirumab-based combinations, a safety lead-in period will be observed via the rolling 6 decision framework of Skolnik et al. (2008). The safety lead-in decisions will be evaluated individually for each of Study JV01 and Study JV02. Based on the first 2 to 6 dose-limiting toxicity (DLT) evaluable patients randomized to the ramucirumab arm at the planned dose for each study (12 mg/kg for Study JV01 or 9 mg/kg for Study JV02), the ramucirumab dose will be de-escalated (to 8 mg/kg for Study JV01 or 6 mg/kg for Study JV02) should any of the rolling-6 'de-escalate' criteria be met in a given study based on observed DLTs related to ramucirumab exposure. Otherwise, enrollment will continue as planned for the study. If the dose is de-

escalated, a study may be terminated for safety should any of the criteria for terminating the study be met due to DLTs observed at the low dose. Otherwise, enrollment will continue with ramucirumab dosing at the lower dose as planned. Enrollment in a study may be temporarily paused in certain circumstances in which 6 ramucirumab patients have enrolled at the current dose, but DLT data are pending in more than 1 patient. Patients who complete the lead-in period will continue until one of the discontinuation criteria are met.

Interim Futility Analysis: For Study JV01, an interim futility analysis will be triggered when approximately 24 total PFS events have been observed across Study JV01 and Study JV02 with a minimum of 8 events in Study JV01. At the interim futility look, the Bayesian analysis must provide a minimum of 60% posterior probability of treatment superiority (PFS hazard ratio [HR] <1 for DSRCT patients) for enrollment on Study JV01 to continue. Otherwise, enrollment on Study JV01 will be stopped. Likewise for Study JV02, an interim futility analysis will be triggered when approximately 24 total PFS events have been observed across Studies JV01 and JV02 with a minimum of 8 events in Study JV02. At the interim futility look for Study JV02, the Bayesian analysis must provide a minimum of 60% posterior probability of treatment superiority (PFS HR <1 for SS patients) for enrollment on Study JV02 to continue. Otherwise, enrollment on Study JV02 will be stopped. Details of the Bayesian model of PFS to be applied at the interim futility analysis are available in Section 6.11.

Primary Analysis: If a given study passes the futility analysis (and thus continues enrollment), the primary analysis will be triggered when PFS events have occurred for approximately 80% of the total enrolled patients (across both Studies JV01 and JV02, regardless of whether or not one stopped early for futility). To conclude success for the investigation in Study JV01, the Bayesian analysis must yield a minimum of 99% posterior probability of treatment superiority (PFS HR <1) for the DSRCT population. Likewise, to conclude success for the investigation in Study JV02, the Bayesian analysis must yield a minimum of 99% posterior probability of treatment superiority (PFS HR <1) for the SS population.

5.1.2. Determination of Sample Size

Traditional operating characteristics associated with the proposed Bayesian design were evaluated via trial simulation. Trial simulations were implemented using the statistical software package R. Simulation results were independently replicated. Note that due to the adaptive borrowing on PFS effect-size between Studies JV01 and JV02, joint scenarios of truth in both DSRCT and SS must be considered when evaluating operating characteristics for each study individually. Under the proposed analysis framework, the sample size is considered adequate to support the primary objective:

- **Type I error:** Type I error here refers to the event the 99% success criterion is reached for a given tumor when in reality the underlying PFS HR (in that tumor) is equal to 1. Given the stringency of the Bayesian success criterion (i.e., 99% posterior probability of superiority), false positives are unlikely in Studies JV01 and JV02. In particular, under the joint null scenario in which neither tumor cohort truly benefits from ramucirumab-based therapy (i.e., $HR_{SS} = HR_{DSRCT} = 1$), the Type I error rate for each study

individually is approximately .003. Importantly, the Type I error rate remains low even under scenarios of strong heterogeneity in effect-size between Studies JV01 and JV02. In particular, if, for example, $HR_{DSRCT} = 1$ but $H_{SS} = .5$, the probability of Type I error in DSRCT is still <2%. This is due to both the adaptive nature of the hierarchical borrowing and the stringent primary success criterion.

- **Power:** Given the large magnitude of PFS benefit targeted in the young adult/pediatric setting, Studies JV01 and JV02 are unlikely to miss truly standard of care-changing improvements due to Type II error. Under the target scenario in which both tumors benefit substantially from ramucirumab-based therapy (i.e., $HR_{SS} = HR_{DSRCT} = .33$), the Bayesian analysis of PFS yields statistical power of approximately 82% to conclude success for each study individually (note, the power for success in *at least 1* study is >82%). For reference, a traditional log rank analysis of an individual study at 24 PFS events (80% of 30 patients enrolled) at $\alpha = .003$ (1-sided) would carry approximately 43% power at $HR = .33$. Note, this calculation did not include a futility analysis such as that proposed in Studies JV01 and JV02, so a fairer assessment of power under the traditional approach would actually be <43%.
- Simulation results over additional joint null/alternative and control scenarios (including scenarios of strong heterogeneity in effect-size between the 2 addenda and mismatch of historical/prospective controls) are tabulated and reviewed in Section 6.6.1.6.

The stringent primary success criteria for each tumor, $\Pr(HR < 1) > 99\%$, was calibrated to ensure that meeting the primary endpoint should imply both statistical significance and large estimated magnitude of patient benefit (additional PFS) for the pediatric/young adult populations of interest. Based on a large simulation study (joint results of all scenarios considered in Section 6.6.1.6), when the 99% posterior probability threshold is reached, the associated estimate of the PFS HR is no larger than approximately .51. Under an example assumption of 3 months for control median PFS (and a further assumption of exponentially distributed PFS), the minimal effect size of $HR = .51$ would correspond to approximately 3 months of additional PFS in this population with high unmet medical need.

5.2. Method of Treatment Assignment

Patients who meet all criteria for enrollment will be randomly assigned to receive study treatment. Before each patient's enrollment into the study, an eligibility check must be conducted between the investigational site and the Eli Lilly and Company (Lilly) clinical research personnel to confirm that each patient meets all enrollment criteria. Upon confirmation of eligibility, the site will register the patient by assigning the patient a unique study identification number via the Interactive Web Response System (IWRS), which is accessible 24 hours a day. Study drug will be allocated to patients using the IWRS.

Approximately 30 patients for each of Studies JV01 and JV02 will be randomized in a 2:1 ratio to receive ramucirumab in combination with tumor-specific chemotherapy versus chemotherapy alone, respectively. Randomization will be conducted separately between Studies JV01 and JV02.

For Studies JV01 and JV02, randomization will be stratified according to staging at relapse (metastatic versus locally advanced).

6. A Priori Statistical Methods

6.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. The interpretation of the study results will be the responsibility of the Lilly Clinical Research Physician/Clinical Research Scientist (CRP/CRS) and statistician. The CRP/CRS and statistician will also be responsible for the appropriate conduct of internal reviews for both the final study report and any study-related material to be authorized by Lilly for publication.

Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol and the justification for making the change will be described in the clinical study report.

Unless indicated below, all analyses (including all tables, figures, and listings) will be conducted separately for Studies JV01 and JV02. A list of analyses will be generated for each study, with tumor-specific modifications made to the associated analysis specifications to accommodate any differences in reporting (e.g., baseline/demographic information).

For Bayesian analyses, posterior medians and 80% (equal tailed) Bayesian credible intervals will be provided for relevant quantities unless otherwise stated below. Full detail regarding specification of prior distributions for prespecified Bayesian analyses are provided in Section 6.6.1.2.

All frequentist tests of treatment effect will be conducted at a 1-sided alpha level of .1, unless otherwise stated, and all confidence intervals (CIs) will have 2-sided coverage equal to 80% .

6.1.1. Populations for Analysis

The following analysis sets will be defined for this study:

Intention-to-treat (ITT) analysis set: will include all randomized patients. Should the ramucirumab dose be de-escalated during the safety lead-in period, all randomized patients will still be included in the ITT analysis set regardless of assigned dose. The ITT analysis of efficacy data will consider allocation of patients to treatment groups as randomized and not by actual treatment received. This analysis set will be used for all baseline and efficacy assessments.

Safety analysis set: will include all randomized patients who received any quantity of study treatment, regardless of their eligibility for the study. The safety evaluation will be performed based on the first dose of study treatment a patient actually received, regardless of the arm to which he or she was randomized. The safety analysis set will be used for all dosing/exposure and safety analyses.

Pharmacokinetic analysis set: will include all randomized patients who received at least 1 full dose of study treatment and have baseline and at least 1 postbaseline evaluable PK sample.

Biomarker analysis set: will include the subset of patients from the ITT analysis set from whom a valid assay result has been obtained.

6.1.2. Handling of Dropouts and Missing Data

Baseline will refer to the last nonmissing observation before first administration of any treatment unless stated otherwise. Missing data, except for dates, will not be imputed. When dates are used in calculations, missing days will be replaced with 15th of the month and missing day/month with 01 JULY. Where windows are allowed for data collection and there is more than 1 reading in any window, appropriate consideration will be given as to whether only 1 value from the window should be used, and if so how it should be chosen. This could either be the mean (geometric mean) or the value closest to the mid-point of the window or the value closest to the data collection time of another variable if the analysis involves time-matched analyses.

Detailed censoring rules for evaluation of the primary PFS endpoint are provided in Table 6.1.

In analyses of tumor response, patients who are inevaluable for tumor response per RECIST v1.1 will be included in the denominators.

6.1.3. Multiple Comparisons/Multiplicity

The primary (efficacy) endpoint (PFS) is to be evaluated via the prespecified Bayesian hierarchical analysis. The Bayesian model and success criteria do not make formal adjustments for multiplicity as Studies JV01 and JV02 are 2 separate investigations; thus control of the familywise error rate between both tumors is not planned. While borrowing between Studies JV01 and JV02 is a component of the analysis, individual conclusions will be drawn for SS and DSRCT. The success criterion may be reached in 1, both, or neither of the 2 tumor types. Different estimates of PFS and PFS HR (with associated posterior credible sets) will be obtained for each tumor.

No adjustments for multiplicity are planned for any secondary or exploratory endpoints.

6.2. Patient Disposition

A detailed description of patient disposition will be provided for Studies JV01 and JV02 separately. It will include a summary of the number and percentage of screened patients randomized in the studies, reasons for discontinuation from study treatment, and reasons for discontinuation from the study. Reason for discontinuation from both study treatment and the study will be listed by the predetermined categories. For treatment discontinuation, these include progressive disease, AE, death, withdrawal by subject, physician decision, noncompliance with study drug, protocol deviation, study terminated by the institutional review board/ethical review board (IRB/ERB)/sponsor, lost to follow-up; for study discontinuation, these include study terminated by sponsor, withdrawal by subject, lost to follow-up, death. If the reason for discontinuation is AE or death, the associated AE or cause of death will be reported. The disposition will also be listed. All patients randomized in the study will be included in the summaries and listings.

6.3. Patient Characteristics

Patient characteristics will be summarized and listed for all patients randomized by treatment, including:

- Patient demographics (including age, sex, race and ethnicity, screening height and weight, and screening derived body surface area)
- Baseline disease characteristics (including disease staging at baseline and initial diagnosis, performance status at baseline, time-to-relapse after front-line therapy for advanced disease, number of prior lines of systemic therapy for advanced disease, prior radiotherapy, prior surgery and associated outcome)

6.4. Treatment Compliance

All intravenous (IV) study drugs will be administered only at the investigational sites by the authorized study site personnel. As a result, treatment compliance is ensured for IV drugs. All study drugs in Study JV02 are to be administered intravenously; thus summary of treatment compliance will not be applicable in Study JV02.

In Study JV01, cyclophosphamide will be administered orally. Patient compliance for cyclophosphamide will be assessed at each visit by direct questioning and counting of returned tablets. Dose compliance will be calculated as (total amount of drug taken (units)/total amount of drug prescribed (units))*100%. "Total amount of drug taken" will be derived from the difference between the total number of tablets dispensed and returned over the course of the patient's treatment across all visits before treatment discontinuation. "Total amount of drug prescribed" is the sum of products of dosing intervals and the expected dose for each interval. Total dose prescribed should take into consideration any dose adjustment(s) before treatment discontinuation date.

6.5. Concomitant and Post-Discontinuation Therapy

Prior and concomitant medications and therapies (e.g., transfusions) will be listed and summarized by treatment arm for the safety population in each study.

The numbers and percentages of patients receiving postdiscontinuation anticancer therapies will be provided by type of therapy (surgery, radiotherapy, or systemic therapy), and by drug class and/or name, overall and by line of therapy.

6.6. Efficacy Analyses

6.6.1. Primary Outcome and Methodology

The primary efficacy endpoint, PFS, will be analyzed via a joint hierarchical Bayesian model (see Section 6.6.1.1). While individual inferences/conclusions will be made for each Study JV01 and Study JV02, a single joint Bayesian hierarchical model will be fit to the PFS data from both Study JV01 and Study JV02 as described below. The Bayesian analysis involves (1) augmenting the control arms with matched (tumor-specific) historical RW data via a power prior approach

and (2) adaptive borrowing of PFS effect-size (log HR for treatment versus control) between Study JV01 and Study JV02 via a random-effects specification within the Bayesian hierarchy.

While PFS from Studies JV01 and JV02 will be modeled within the same Bayesian hierarchical model for the primary analysis, the model is constructed such that respective success criteria can and will be evaluated separately (and likewise estimates of HR with associated credible intervals will differ between Studies JV01 and JV02). One, both, or neither of Studies JV01 and JV02 may meet the criterion for success on the basis of PFS. The primary success criteria at the primary analysis (to occur when approximately 80% of patients enrolled on Studies JV01 and JV02 have had documented PFS events) are:

- For Study JV01, the Bayesian model must yield posterior $\Pr(HR_{DSRCT} < 1 \mid Data) > 99\%$ in order to call success in DSRCT
- For Study JV02, the Bayesian model must yield posterior $\Pr(HR_{SS} < 1 \mid Data) > 99\%$ in order to call success in SS
- Note that if Studies JV01 or JV02 stop early for futility (at the futility interim analysis) that study cannot claim success at future analyses.

To correspond with the 99% superiority criterion, 98% equal-tailed posterior credible intervals will be provided along with estimates (posterior medians) of PFS HR (treatment versus control) in each tumor (i.e., separately for Studies JV01 and JV02). Plots of estimated (Weibull) survival curves will be provided for each study separately. Estimates of landmark PFS at months 3, 6, 9, and 12 (along with 80% equal-tailed posterior credible intervals) will be provided.

6.6.1.1. Primary PFS Endpoint Definition and Censoring Rules

Progression-free survival (PFS) is defined as the time from randomization until the first investigator-determined objective progression as defined by RECIST v1.1, or death from any cause in the absence of progressive disease. Patients known to be alive and without disease progression will be censored at the time of the last adequate tumor assessment (a detailed PFS event/censoring scheme is provided in Table 6.1).

Table 6.1. Progression-Free Survival Event/Censoring Scheme

Situation[a]	Event/Censor	Date of Event or Censor
Investigator assessed tumor progression or death	Event	Earliest date of PD or death
No tumor progression and no death	Censored	Date of last adequate radiological assessment or date of randomization (whichever is later)[b]
Unless		
No baseline radiological tumor assessment available	Censored	Date of randomization
No adequate postbaseline radiological tumor assessment available and death reported after 2 scan intervals following randomization[b,c]	Censored	Date of randomization

New anticancer treatment started and no tumor progression or death within 14 days	Censored	Date of adequate radiological assessment before (start of new therapy +14 days) or date of randomization (whichever is later)
Tumor progression or death documented immediately after 2 or more scan intervals following last adequate radiological tumor assessment or randomization (whichever is later)[b,c]	Censored	Date of last adequate radiological assessment or date of randomization (whichever is later)[b]

Abbreviations: CR = complete response; PD = progressive disease; PFS = progression-free survival; PR = partial response; SD = stable disease.

a Symptomatic deterioration (i.e., symptomatic progression that is not radiologically confirmed) will not be considered as tumor progression.

b Adequate radiological tumor assessment refers to an assessment with one of the following responses: CR, PR, SD, or PD.

c Radiologic imaging for tumor assessment will be performed every 6 weeks (±7 days) after randomization (regardless of treatment delays) during the Study Treatment Period, until disease progression OR study completion or 14 months after randomization, whichever occurs first. Refer to footnote b for any patient whose disease has not progressed by 14 months after randomization. Any patient whose disease has not progressed by 14 months after randomization will be evaluated for response every 12 weeks (±7 days) from 14 months after randomization, until disease progression OR study completion OR 3 years after randomization, whichever occurs first; then as per standard clinical practice after that.

6.6.1.2. Bayesian Hierarchical Model for Primary Progression-Free Survival Endpoint

Let t_i and e_i denote the observed PFS event/censoring time and event indicator (i.e., $e_i = 1$ if PFS event and $e_i = 0$ if censored), respectively, for each prospective patient i. Let $C(i) = 1,2$ indicate the tumor cohort to which patient i belongs. We consider a parametric (in this case Weibull) survival model to yield appropriate balance of flexibility/stability given small sample sizes. Each patient i thus contributes the following to the joint likelihood:

$$f(t_i, e_i|\boldsymbol{\theta}) = \exp\{-H(t_i|\boldsymbol{\theta})\} \cdot h(t_i|\boldsymbol{\theta})^{e_i},$$

where $h(t_i|\boldsymbol{\theta})$ denotes the Weibull hazard:

$$h(t_i|\boldsymbol{\theta}) = \exp\{\beta_{C(i)} \cdot Trt(i)\} \cdot \nu_{C(i)} \cdot \lambda_{C(i)} \cdot t_i^{\nu_{C(i)}-1}$$

and $H(t_i|\boldsymbol{\theta})$ denotes the corresponding cumulative hazard:

$$H(t_i|\boldsymbol{\theta}) = \exp\{\beta_{C(i)} \cdot Trt(i)\} \cdot \lambda_{C(i)} \cdot t_i^{\nu_{C(i)}}.$$

In the above notation, we note that $HR_{SS} = \exp(\beta_1)$ and $HR_{DSRCT} = \exp(\beta_2)$. The quantities ν_1, ν_2, λ_1, and λ_2 are the usual shape and rate parameters characterizing the Weibull family of likelihood distributions (subscripts indicating tumor cohort).

The mechanism with which we facilitate the adaptive borrowing between tumors comprises essentially a random-effects meta-analysis of the log HRs:

$$\beta_1, \beta_2 \sim N(\mu, \tau^2),$$

with $\tau \sim U(0,1)$ and $\mu \sim N(0,1)$.

The prior distribution for the between-tumor standard deviation τ is an important driver of the degree of borrowing between tumors. This prior distribution maintains conservativism regarding variability (of ramucirumab PFS effect size) between tumors based on historical data across the ramucirumab program. A Bayesian random-effects meta-analysis of the log PFS HR observed in all completed randomized phase 3 trials of ramucirumab was conducted. Note that these studies together comprise a very heterogeneous pool of tumor types relative to Studies JV01 and JV02 (2 subtypes of soft-tissue sarcoma). The resulting posterior median of the standard deviation parameter (governing variability between true log PFS HRs) was equal to approximately .2, which is substantially below the median and upper limit of the $U(0,1)$ prior chosen here. Importantly, we note that (under the proposed prior for τ) even in the event of very strong heterogeneity in true effect size ($HR_{SS} = 1$ and $HR_{DSRCT} = .5$ or vice versa) the Type I error for the primary decision rule in the negative tumor is still controlled below 2% (1-sided) (see Table 6.5.).

The hyperprior distribution for the across-tumor mean parameter μ has been set to a Normal distribution with mean equal to zero and standard deviation equal to 1, representing a somewhat conservative prior (yet quite diffuse on the log HR scale).

6.6.1.3. Prior Distributions for Control Progression-Free Survival

To complete the above Bayesian hierarchy, we must specify the prior distributions for parameters governing the Weibull survival curves for patients receiving the control regimens, namely the parameters ν_1, ν_2, λ_1, and λ_2. The prior distributions for these parameters will be derived from RW PFS data obtained from patients with DSRCT and SS treated with chemotherapy in the relapsed setting, matched to the prospective patients enrolling on the ramucirumab arms of Studies JV01 and JV02. As control outcomes must be allowed to differ between the 2 tumors, the matching algorithm and subsequent prior specification will be performed separately within each tumor.

At the primary analysis, a total of 20 historical RW patients per tumor will be matched to the corresponding 20 prospective patients (enrolled on the ramucirumab arm) according to the prespecified algorithm provided below. A simple power prior approach (Ibrahim et al. 2015) will be adopted to formally incorporate the matched RW historical control subjects into the Bayesian analysis of prospective PFS in a manner that appropriately downweights the influence of the historical data. Precisely as above (for the prospective data), let t_i^{hist} and e_i^{hist} denote the observed PFS event/censoring time and event for each historical patient i. Let $C^{hist}(i) = 1,2$ indicate the tumor cohort to which historical control patient i belongs. Each historical patient i thus contributes the following to the reweighted joint likelihood:

$$f(t_i, e_i|\boldsymbol{\theta}) = \left[\exp\left\{-H\left(t_i^{hist}\middle|\boldsymbol{\theta}\right)\right\} \cdot h\left(t_i^{hist}\middle|\boldsymbol{\theta}\right)^{e_i^{hist}}\right]^{a_0},$$

where $h(t_i|\boldsymbol{\theta})$ denotes the Weibull hazard:

$$h(t_i|\boldsymbol{\theta}) = \nu_{C(i)} \cdot \lambda_{C(i)} \cdot t_i^{\nu_{C(i)}-1}$$

and $H(t_i|\boldsymbol{\theta})$ denotes the corresponding cumulative hazard:

$$H(t_i|\boldsymbol{\theta}) = \lambda_{C(i)} \cdot t_i^{\nu_{C(i)}}.$$

Note that here the parameters ν_1, ν_2, λ_1, and λ_2 are exactly those which appear in the joint likelihood for the prospective data above, so that the reweighted likelihood for the historical data (together with the hyperpriors discussed in the following paragraph) functions as an informative prior for those quantities. Importantly, the power parameter a_0 will be set to a prespecified value of .5 to substantially downweight the influence of the historical control data relative to the prospective control data, corresponding to a prior effective sample size in each tumor of approximately 10 events (and fewer if a number of the matched historical patients were censored for PFS).

Finally, we specify hyperpriors for ν_1, ν_2, λ_1, and λ_2. For ν_1 and ν_2 , we assume independent truncated normal priors, $Normal(1,1) \cdot I(\nu_j > 0)$ (which centers the curves weakly at the more parsimonious exponential survival model), and for λ_1 and λ_2 , we assume diffuse/noninformative independent gamma priors, $Gamma(.1, .1)$.

If adequate real-world outcomes are not available at interim (see Section 6.11), the futility analysis will use the Bayesian hierarchical model in Section 6.6.1.2 using only prospective patients and the noninformative hyperpriors for ν_1, ν_2, λ_1, and λ_2 described in the previous paragraph.

6.6.1.4. Retrospective (Real World) Data to Inform Control Priors

The prior distributions for PFS on the control arms in Study JV01 and Study JV02 will be constructed (as described above) from PFS outcomes for patients with relapsed DSRCT or SS treated with chemotherapy or pazopanib in the RW setting. These data will be obtained from chart reviews conducted at institutions including US sites. Eligible patients will have been initially diagnosed with DSRCT or SS before age 40 years. This age limit is higher than the eligibility criterion for Studies JV01 and JV02, as pathology and outcomes are not expected to vary based on age in these diseases; therefore, the age range was expanded to facilitate improved matching on other important prognostic factors. Eligible patients must have progressed on at least 1 course of chemotherapy for relapsed, recurrent, or progressive disease in order to estimate PFS for at least 1 line of therapy. Variables to be extracted include patient and baseline/disease characteristics, as well as treatments (surgery, radiotherapy, chemotherapy) and outcomes since initial DSRCT or SS diagnosis. Progression dates will be based primarily on physician notes and will not require re-interpretation of radiologic scans. The target sample size is 100 charts each for DSRCT and SS at the time of the final analysis.

6.6.1.5. Prespecified Historical Control Matching Algorithm

A propensity-based matching algorithm will be utilized to match each of the 20 prospective ramucirumab patients with a historical control patient (conducted separately within each tumor). The independent (baseline/demographic) variables to be used for purposes of matching are provided in Table 6.2. Historical patients with chart reviews containing multiple eligible lines of therapy will be used as distinct observations for matching, with PFS calculated from the relevant line of therapy. The dependent variable is the binary cohort indictor (prospectively randomized

ramucirumab subject versus historical control subject). The propensity score models (constructed separately for Study JV01 and Study JV02) will be developed using main-effects logistic regression (e.g., PROC LOGISTIC in SAS).

A nearest-neighbor propensity-score matching algorithm will be implemented on the logit of propensity score against a caliper of width equal to 0.2 of the pooled standard deviation of the logit of the propensity scores. Matching *with replacement* will be used so that a unique cohort of matched historical patients is obtained regardless of any arbitrary ordering of the prospective patients. In the case of tied propensities, all tied nearest neighbors will be incorporated into the power prior, with the power parameter for each corresponding likelihood term simply adjusted downward from the original $a_0 = .5$ to the quantity $.5/K_i$ where K_i is the number of tied matches obtained for prospective patient i.

If for any reason (e.g., small sample size) the model-based calculation of propensities is not feasible, the metric-based Mahalanobis nearest-neighbor (with replacement) approach will be used for matching. If Mahalanobis matching is also not feasible (e.g. sample covariance matrix cannot be estimated or inverted), real-world outcomes will be omitted from the analysis.

The propensity model and matching algorithm will be executed at both the interim futility analysis and the final analysis of PFS. The propensity model will be updated at the final analysis based on the full population of prospectively randomized (ramucirumab) patients and historical control patients. The subset of patients who were included in the interim analysis may therefore be matched to different historical control patients at the final analysis than at the interim analysis.

Table 6.2 Variables for Matching Real-World Data (Historical Control) and Prospective Patients

Variable Name	Description	Categories/Units	Levels for Matching
Age	Patient age on C1D1 of line of systemic therapy for advanced disease in the relapsed/recurrent/progressive setting.	Years	• <18 • ≥18 • Missing
Line of Therapy	Line of systemic therapy for advanced disease	Integer	• 1nd line • 2rd line and above • Missing
Time-to-Relapse	Time between the last date of administration of initial course of therapy (systemic, radiotherapy, surgery) and date of diagnosis for relapsed/recurrent disease. If disease progressed during the initial treatment, time-to-relapse is 0 (months).	Years	DSRCT: • <1 year • ≥1 year • Missing SS: • <2 years • ≥2 years • Missing
Metastatic Disease at Relapse	Indictor of the presence of metastases as of C1D1 of line of systemic therapy for advanced disease in the relapsed/recurrent/progressive setting.	Years	• Yes • No • Missing
Initial Tumor Size	Size of tumor at initial diagnosis of SS or DSRCT	Cm	DSRCT: • <10 cm • ≥10 cm • Missing SS: • <5 cm • ≥5 cm • Missing

Abbreviations: C1D1 =Cycle 1 Day 1; Cm = centimeter; DSRCT = desmoplastic small round cell tumor; SS = synovial sarcoma.

6.6.1.6. Simulations of Primary Progression-Free Survival Endpoint

Traditional operating characteristics (e.g., Type I and Type II error rates) for the proposed Bayesian analysis have been evaluated via trial simulation. For simplicity, in all scenarios, PFS outcomes were simulated from exponential distributions (noting that the actual Bayesian model for the PFS analysis allows nonconstant hazard). Throughout, a fixed historical control prior (the same for each tumor) for the rate and shape parameters of the control Weibull curves was constructed with prior effective sample size of approximately 10 events to mimic the reality of the above prior construction algorithm. To embed realistic control assumptions, the marginal prior for the rate parameter was set to a Gamma (10, 42.78) based on a digitized Kaplan-Meier curve for PFS outcomes in patients with SS who received gemcitabine in combination with

docetaxel in the relapsed setting (Abouharb et al. 2014). The median of this prior is .226. Except for the scenarios in Table 6.3 (in which the impact of biased historical controls is explicitly studied), the prospective PFS data on control was simulated from the exponential distribution with hazard equal to .226. The marginal prior for the shape parameter was set to a Uniform (.5,1.5), so that its center was consistent with the exponential assumption from which the patient data were simulated.

Table 6.3 provides the simulated probabilities of passing the interim futility criterion (continuing enrollment) and passing final (primary) success criterion for each of 7 scenarios in which the underlying HR is the same in both tumors. Note that to pass the final success criterion, the interim criterion must also be passed (otherwise the final analysis would never have been conducted) so the last 2 columns actually provide the probability of passing *both* the interim and final criteria. We see that under the joint null (Scenario 1), the stringent success criterion (99% Bayesian probability of superiority) yields very strong control of Type I error. Indeed, when *neither* tumor benefits from the addition of ramucirumab to standard chemotherapy, the probability of passing the primary success criterion is about .003 in each tumor. From Scenario 7, we see that when both tumors benefit notably from the addition of ramucirumab (HR=.33 in both tumors), under the proposed Bayesian design, the power for each tumor individually is 82%.

Table 6.3. Simulated Operating Characteristics

Scenario Number	True Hazard Ratio		Pr(Pass Interim)		Pr(Pass Final)	
	SS	DSRCT	SS	DSRCT	SS	DSRCT
1	1.00	1.00	0.44	0.44	0.003	0.003
2	0.80	0.80	0.67	0.67	0.03	0.03
3	0.70	0.70	0.79	0.79	0.07	0.07
4	0.60	0.60	0.89	0.89	0.18	0.18
5	0.50	0.50	0.96	0.96	0.37	0.37
6	0.40	0.40	0.99	0.99	0.64	0.64
7	0.33	0.33	0.99	0.99	0.82	0.82

Abbreviations: DSRCT = desmoplastic small round cell tumor; Pr=probability; SS = synovial sarcoma.

Additional simulation work was performed to evaluate operating characteristics of the design under scenarios of potential prior/data conflict. In Table 6.4, we evaluate the impact of potentially biased historical control outcomes on Type I error rates in the prospective study. By 'biased', we mean that the historical control data lead to priors that are located away from the true parameters governing PFS outcomes for the prospective control subjects. In Scenarios 8-11, the priors for control PFS hazard were centered at a value of .226 in each tumor with variability corresponding to approximately 10 events. However, in each Scenario 8-11, the true hazard on control was set to some value *smaller* than .226 (in both tumors simultaneously) which, of course, has the impact of inflating Type I error. From Scenario 11, we see that even under substantial prior/data conflict (i.e., when the prospective versus historical controls differ by a HR of .65 in *both* tumors), the Type I error rate remains small (<2%). This is the result of (1) a very stringent success criterion and (2) prior effective sample size of only 10 events per tumor so that the prospective data are not overwhelmed by the priors.

Table 6.4. Impact of Biased Historical Controls on Type I Error

Scenario Number	True CTRL Hazard		HR (Prosp. vs. Hist. CTRL)		Pr(Pass Final)	
	SS	DSRCT	SS	DSRCT	SS	DSRCT
8	.2147	.2147	0.95	.95	0.006	0.006
9	.1921	.1921	0.85	0.85	0.008	0.008
10	.1695	.1695	0.75	0.75	0.014	0.014
11	.1469	.1469	.65	.65	0.018	0.018

Abbreviations: CTRL = Control; DSRCT = desmoplastic small round cell tumor; hist. = historical; HR = hazard ratio; Pr=probability; prosp = prospective; SS = synovial sarcoma.

All scenarios above assume treatment versus control HR = 1 (in both tumors) so that Pr(Final) represents Type I error (under joint null).

Table 6.5 provides operating characteristics under scenarios in which only 1 tumor benefits from the addition of ramucirumab to chemotherapy. For example, in each Scenario 12-16, the true HR was set equal to 1 in SS but some value <1 for DSRCT. Most notably, under very strong heterogeneity in effect-size (Scenario 16) the Type I Error is still controlled below 2% for the negative tumor (SS) despite the Bayesian borrowing from a positive tumor. On the other hand, the power to conclude success in the positive tumor (with HR=.5) is of course lower than when both tumors benefit by that margin (see Scenario 5). Viewed from a Bayesian perspective, this is an expected (and desirable) attribute of the model. Observing negative results in 1 subtype should somewhat discredit positive observations in the other subtype while positive results in *both* subtypes should be reinforcing.

Table 6.5. Impact of Heterogeneous Effect-Size Between Tumors

Scenario Number	True Hazard Ratio		Pr(Pass Interim)		Pr(Pass Final)	
	SS	DSRCT	SS	DSRCT	SS	DSRCT
12	1	0.80	0.50	0.61	0.008	0.018
13	1	0.70	0.52	0.71	0.007	0.033
14	1	0.60	0.57	0.80	0.011	0.078
15	1	0.55	0.58	0.84	0.012	0.110
16	1	0.50	0.60	0.88	0.015	0.164

Abbreviations: DSRCT = desmoplastic small round cell tumor; Pr=Probability; SS = synovial sarcoma.

Table 6.6 provides operating characteristics under the scenario where real-world controls are not incorporated in the interim analysis. Rather than using a Gamma(10, 42.78) prior for the rate parameters, a noninformative Gamma(0.1, 0.1) prior is used. The marginal prior for the shape parameter is kept as a Uniform (.5,1.5) as in previous simulations. The prior/data conflict scenarios (Table 6.4) are omitted from this simulation.

Table 6.6. Impact on Interim Analysis when No RW data is Included

Scenario Number	True Hazard Ratio		Include RW Outcomes		No RW Outcomes	
			Pr(Pass Interim)		Pr(Pass Interim)	
	SS	DSRCT	SS	DSRCT	SS	DSRCT
1	1	1	0.44	0.44	0.39	0.39
2	0.8	0.8	0.67	0.67	0.57	0.58
3	0.7	0.7	0.79	0.79	0.69	0.69
4	0.6	0.6	0.89	0.89	0.8	0.79
5	0.5	0.5	0.96	0.96	0.89	0.89
6	0.4	0.4	0.99	0.99	0.96	0.95
7	0.33	0.33	0.99	0.99	0.98	0.98
12	1	0.8	0.5	0.61	0.45	0.52
13	1	0.7	0.52	0.71	0.48	0.6
14	1	0.6	0.57	0.8	0.5	0.67
15	1	0.55	0.58	0.84	0.53	0.72
16	1	0.5	0.6	0.88	0.54	0.76

Conducting the interim analysis without incorporation of historical control information decreases the probability of passing the interim futility criteria (i.e. continue enrollment in a given tumor). Section 5.1.2 lists HR = 0.33 in both tumor types as the target scenario. This is where the Bayesian analysis reaches 82% power for a given tumor type at the final analysis. At HR = 0.33 for SS and DSRCT, the probability of passing the interim is decreased by only 1%. When the HRs are equal (Scenarios 1-7), there is a modest decrease in power with the largest difference of 10% occurring at HR = 0.70. A larger decrease in power at interim can be seen in Scenarios 12-16 where there is significant heterogeneity between tumor types. When both HR = 1, the probability of incorrectly continuing the trial decreases from 0.44 to 0.39. Overall, the interim analysis retains sufficient power if RWE cannot be included in the analysis while not increasing the probability of passing the interim under the null case of HR = 1.

6.6.1.7. Sensitivity Analyses for Primary Progression-Free Survival Endpoint

Besides the primary analysis of PFS per Bayesian hierarchical analysis described above, traditional (frequentist) assessments will be performed to provide additional context. Non-parametric estimates of survival curves will be obtained via the method of Kaplan and Meier. The (unstratified) log rank test will be used to compare treatment arms for Studies JV01 and JV02 separately and without integration of historical control data.

Four sensitivity analyes are planned within the Bayesian framework: (1) historical control priors based on formal expert elicitation will be used in lieu of the power priors constructed from RW data (2) borrowing between tumors will be removed (3) weakly informative priors will be used

for control PFS and (4) the between-tumor borrowing and historical control borrowing will be removed so that separate Weibull models with weakly informative priors will be used.

Sensitivity to using "missing" as a level for matching variables (Table 6.2) will also be evaluated. The matching algorithm will be rerun using only the complete case population of historical patients. Any prospectively randomized patient with missing values in one or more of matching variables will be excluded from the matching algorithm but will be still included in the Bayesian analysis of PFS. The Bayesian Weibull model will be refit and the associated impact on statistical inference will be studied.

Sensitivity to the choice of historical control subject matching algorithm will also be evaluated. Besides the propensity-based method corresponding to the primary analysis, a nearest-neighbor (Mahalanobis) metric-based matching approach will be implemented for matching historical control patients for inclusion into the prior distributions. The associated impact on statistical inference will be studied.

6.6.1.8. Expert Elicitation of Control Progression-Free Survival

Formal prior elicitation of leading experts in DSRCT and SS will be conducted to obtain characterization of expected PFS outcomes on chemotherapy in the relapsed setting. The elicitation interviews (approximately 1 hour) are to be conducted with each expert individually per a common protocol (see template in Appendix 2). The results will be synthesized across the experts to obtain prior distributions to be used in sensitivity analysis of PFS (see Section 6.6.1.7).

6.6.2. Secondary Efficacy Analyses

Overall response rate (ORR) is defined as the number of patients who achieve a best overall response of CR or PR divided by the total number of patients randomized to the corresponding treatment arm (ITT population). The confirmation of CR and PR is required. The ORR for each study arm, with 80% CI (per the method of Clopper and Pearson 1934), will be summarized separately for Studies JV01 and JV02. Comparison between experimental and control arms will be performed separately between Studies JV01 and JV02 using Fisher's exact test with 1-sided level .1.

The CR rate is defined as the number of patients who achieve a best overall response of CR divided by the total number of patients randomized to the corresponding treatment arm (ITT population). The confirmation of CR is required. The CR rate for each study arm, with 80% frequentist CI, will be summarized separately for Studies JV01 and JV02.

Duration of response is defined as the time from the date measurement criteria for CR or PR (whichever is first recorded) are first met until the first date that disease is recurrent or objective progression is observed, per RECIST 1.1, or the date of death from any cause in the absence of objectively determined disease progression or recurrence. DoR will be calculated only for patients with confirmed PR or CR. DoR will be summarized for each treatment arm using descriptive statistics.

6.7. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Planned PK and PK/pharmacodynamic (PD) analyses are specified in a separate standalone PK/PD analysis plan.

6.8. Safety Analyses

Safety analyses will be based on the safety populations. Safety analyses will be conducted separately for Studies JV01 and JV02.

6.8.1. Extent of Exposure

The number of dose adjustments (dose omissions, reductions, delays, and increases), the number of cycles received, the duration of therapy, the cumulative dose, and dose intensity/relative dose intensity will be summarized by treatment arm for Studies JV01 and JV02 separately. Dose intensity is defined as actual cumulative amount of drug taken/duration of treatment. Relative dose intensity is calculated as (actual amount of drug taken/amount of drug prescribed)*100%.

6.8.2. Adverse Events

Adverse event (AE) verbatim terms will be provided by the investigators and then will be mapped by Lilly or its designee to corresponding terminology within the Medical Dictionary for Regulatory Activities (MedDRA) Lower Level term (LLT) dictionary. The investigator will use Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 (NCI 2009) to assign AE severity grades.

Pre-existing conditions are defined as AEs that begin but do not resolve before the first dose of study drug in each study period or events that occur after informed consent and resolve before the first dose of study drug. Pre-existing conditions will be presented by patient and can be combined with the listing of AEs, so that the history of the pre-existing conditions/AEs can be traced.

A treatment-emergent adverse event (TEAE) is defined as an event that first occurred or worsened in CTCAE grade after the first dose of study treatment. The MedDRA LLT will be used in the treatment-emergent computation.

Adverse events (AEs) will be summarized by maximum CTCAE grade (preferred term [PT] within system organ classification [SOC]). Summaries of AEs will include:

- Summary of all TEAEs
- Summary of all TEAEs related to study treatment
- Summary of all TEAEs leading to dose adjustment
- Summary of all AEs resulting in discontinuation from study treatment
- Summary of Adverse Events of Special Interest

6.8.3. Deaths, Serious Adverse Events, Other Significant Adverse Events

All deaths that occur during the study, within 30 days of the study treatment discontinuation, as well as the cause of death, will be summarized and listed.

Serious adverse events (SAEs) are those events that result in death, are life-threatening, require or prolong hospitalization, result in persistent or significant disability/incapacity, or cause congenital anomaly/birth defect. SAEs will be tabulated by SOC, PT terms, and relationship to study drug.

6.8.4. Clinical Laboratory Evaluation

All relevant hematology and chemistry laboratory values will be graded according to CTCAE Version 4. Shift tables showing the change from baseline to the worst grade on study will be presented.

6.8.5. Vital Signs and Other Physical Findings

Temperature, blood pressure, pulse rate, weight and Eastern Cooperative Oncology Group/Karnofsky/Lansky performance score (as applicable) will be summarized by cohort and overall.

6.8.6. Electrocardiograms

Local electrocardiograms (ECGs) will be summarized by cycle and overall. The summary by cycle will classify patients as having normal or abnormal. Additionally, QT, RR, and QRS intervals will also be collected and QTcF will be calculated based on these parameters. AEs that could be associated with abnormal ECGs will be presented, if appropriate.

6.9. Subgroup Analyses

Due to sample-size limitations, no formal statistical tests of treatment effect by subgroup will be prespecified. However, descriptive statistics for primary and secondary efficacy measures will be provided by treatment arm within the following subgroups for Studies JV01 and JV02 separately:

- Age (<18 vs. ≥18)
- Line of therapy (sescond line vs. third line and above)
- Time-to-Relapse
    - Study JV01: <1 yr vs. ≥1 yr
    - Study JV02: <2.5 yrs vs. ≥2.5 yrs
- Metastatic disease at relapse (yes vs. no)
- Tumor size at initial diagnosis
    - Study JV01: <10 cm vs. ≥10 cm

- Study JV02: <5 cm vs. ≥5 cm

6.10. Protocol Violations

Important protocol deviations that potentially compromise the data integrity and patients' safety will be listed. These deviations will include those defined by:

- Informed consent
- Inclusion/exclusion criteria
- Investigational product
- Study procedures
- Administrative/oversight
- Safety
- Other

On the basis of the discussion with the study team, the detailed description of each deviation within the above category and the method to identify each deviation will be listed in a separate document – Business Process Document: Important Protocol Deviations.

6.11. Interim Analyses

Studies JV01 and JV02 are unblinded. Safety data will be reviewed on an ongoing basis during the safety lead-in periods for Studies JV01 and JV02.

Due to the nature of the Bayesian hierarchical analysis of PFS (which includes a formal mechanism for adaptive borrowing on effect-size between Studies JV01 and JV02), timing of the preplanned interim analysis of PFS is based on the total number of PFS events observed across both Studies JV01 and JV02. An interim futility analysis will be triggered when approximately 24 total PFS events have been observed across Studies JV01 and JV02 with a minimum of 8 events in each study. While a single joint Bayesian hierarchical model will be fit to the PFS data from both Studies JV01 and JV02, study specific futility conclusions will be made based on tumor-specific output obtained from the Bayesian model. At the interim futility look, the Bayesian analysis must provide a minimum of 60% posterior probability of treatment superiority (PFS HR <1 for DSRCT patients) in order for enrollment on Study JV01 to continue. Otherwise, enrollment on Study JV01 will be stopped. Likewise, at the interim futility look, the Bayesian analysis must provide a minimum of 60% confidence in treatment superiority (PFS HR <1 for SS patients) in order for enrollment on Study JV02 to continue. Otherwise, enrollment on Study JV02 will be stopped.

The interim futility analysis of PFS will utilize exactly the same Bayesian hierarchical model (including the historical control matching algorithm) as for the primary analysis. Regarding historical control matching at interim, the methodology is identical to that specified for the primary analysis in Section 6.6.1.5, except it is possible that <20 patients will have enrolled on each treatment arm of Studies JV01 and JV02 as of the interim analysis. Hence, <20 historical

control patients may be matched per tumor, but the power prior parameter will still be equal to .5.

If fewer than 10 real world historical control subjects are available within a given tumor type as of the database lock for the interim futility analysis, the control arm (for that study) will not be augmented with real-world data. The model to assess interim futility (Section 6.6.1.2) would therefore be fit using only prospectively randomized patients and the noninformative hyperpriors for ν_1, ν_2, λ_1, and λ_2 presented in Section 6.6.1.3.

6.12. Annual Report Analyses

The following reports are needed as requested for annual reporting purposes:

Development Safety Update Report:

- Cumulative Subject Exposure by Age Group and Sex
- Cumulative Subject Exposure by Racial Group
- Estimated Cumulative Subject Exposure
- Exposure Information
- Listing of Discontinuations Due to AE during the Reporting Period
- Listing of Subjects Who Died during the Reporting Period

Clinical Investigator's Brochure (IB):

- Listing and Summary of SAE
- Listing and Summary of Death
- Listing and Summary of TEAE (and by maximum CTCAE grade)
- Listing and Summary of Patient Disposition
- Listing and Summary of Study Drug Adjustment

Other reports may be requested if deemed appropriate for the IB.

6.13. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

- Summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AEs will be summarized by MedDRA PT within treatment group.
- An AE is considered 'Serious' whether or not it is a TEAE.

- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - the number of events experienced
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in <5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures, e.g., the clinical study report (CSR), manuscripts, and so forth.

In addition, a participant flow will be created that will describe how many enrolled patients completed the study, and for those who did not, the frequency of each reason for not completing. This analysis will be based on study discontinuation, not treatment discontinuation. A patient will be identified as having completed the study if the patient has received ≥1 dose of study drug and have ≥1 postbaseline tumor assessment at the time of the final analysis.

7. References

Abouharb S, et al. Synovial sarcoma: evaluation of response to treatment with gemcitabine and docetaxel. ASCO Annual Meeting poster. *J Clin Oncol*. 2014;32:5s. (suppl; abstr 10564).

Clopper, CJ, et al. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* 1934;26(4):404–413.

Ibrahim JG, et al. The power prior: theory and applications. *Stat Med*. 2015;34(28):3724-3749.

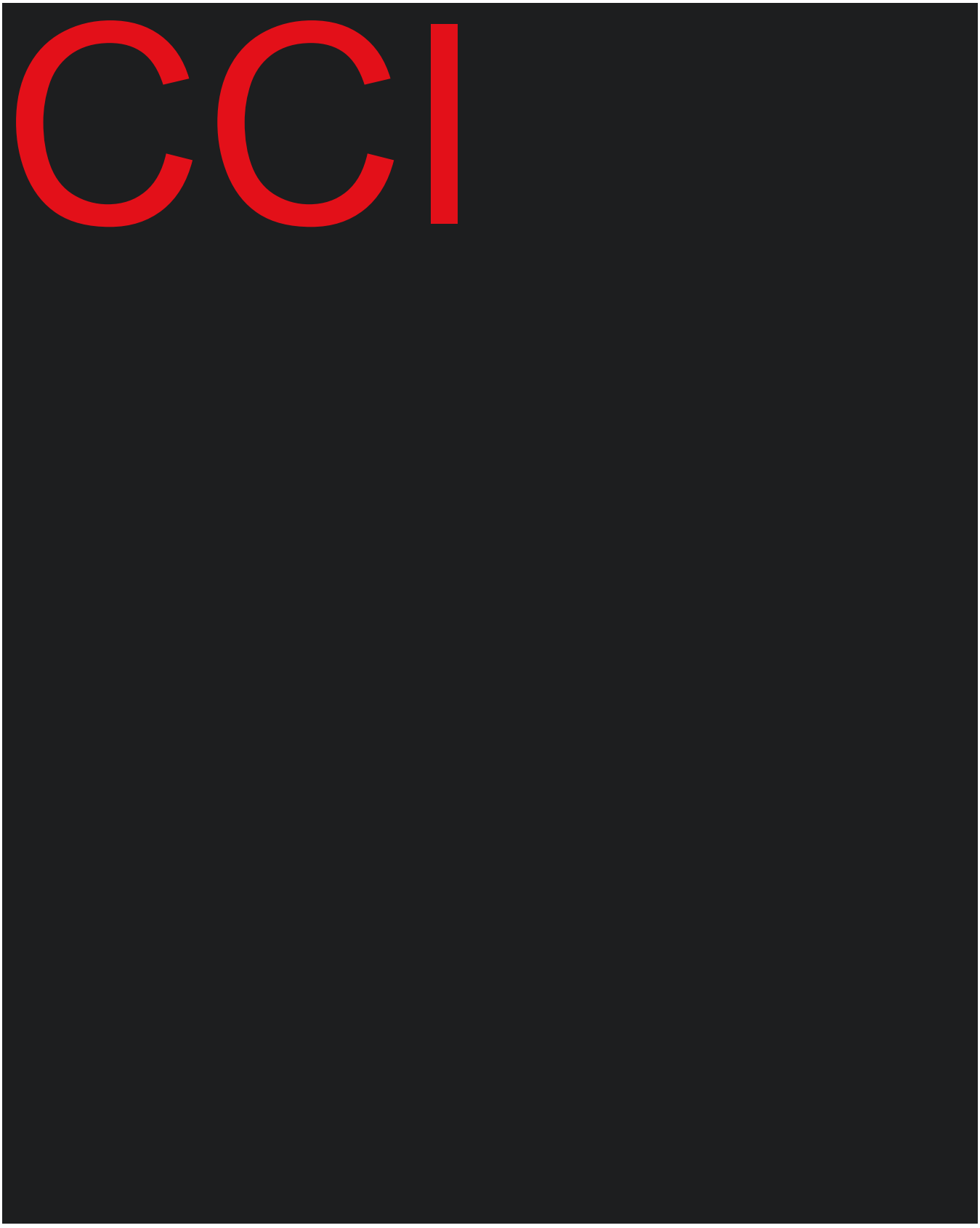
Lunn D, et al. The BUGS book: a practical introduction to Bayesian analysis. London: Chapman and Hall/CRC; 2012.

Plummer M. JAGS Version 4.3.0 user manual. 2017

Skolnik JM, et al. Shortening the timeline of pediatric phase I trials: the rolling six design. *J Clin Oncol.* 2008;26(2):190-195.

Appendix 1. JAGS Code for Bayesian Hierarchical Model of Primary PFS Endpoint

The approximate posterior distribution corresponding to the Bayesian hierarchical model of the primary PFS endpoint is to be obtained via Markov Chain Monte Carlo (MCMC) simulation using JAGS (Plummer 2017). The JAGS script for the Bayesian model is provided below.


CCI

Appendix 2. Prior Elicitation Protocol

Prior Information Elicitation Protocol

Date:

Project Identifiers

Compound: Ramucirumab
Indication: Synovial Sarcoma and Desmoplastic Small Round Cell Tumor (DSRCT)
Patient population: Pediatric patients and young adults (age ≤30 years) with relapsed, recurrent, or progressive synovial sarcoma or desmoplastic small round cell tumor

Therapeutic area: Oncology
Endpoints:

- Progression-Free Survival (PFS)

Expert:

Name:
Affiliation:
Email:
Phone:

Facilitator:

Name:
Affiliation:
Email:

Setting of Elicitation: Teleconference

1. Intention of the Prior Elicitation Exercise

Studies J1S-MC-JV01 and J1S-MC-JV02 are planned randomized Phase 2 studies evaluating ramucirumab in combination with chemotherapy versus chemotherapy alone in pediatric patients and young adults with relapsed, recurrent, or progressive synovial sarcoma (SS) or desmoplastic small round cell tumor (DSRCT). Due to modest sample sizes associated with the rarity of the 2 indications, a novel Bayesian design will be utilized whereby existing knowledge regarding expected patient outcome on the control regimens will be directly leveraged within the statistical analysis to augment information generated from the prospective control arms. The existing knowledge takes 2 forms: (1) historical patient data from retrospective chart reviews and (2) expert opinion from world leaders in the treatment of these conditions. *Prior elicitation* is the formal statistical framework by which we assess/document the later source of existing knowledge, expert opinion, in a manner that allows its conversion into mathematical objects, called *prior distributions* (see Figure 1 below), that can be formally leveraged within a statistical model. See Hampson et al. (2015) for an example of a clinical trial design in the rare/pediatric setting leveraging expert opinion directly within a prespecified primary analysis.

2. Procedures

The prior elicitation exercise is an interview (conducted via teleconference in the present case) during which questions of specialized format are asked of the participating expert(s). The interview questions are directed at documentation of (1) the expert's judgement regarding the most plausible true value for a given endpoint of interest (e.g. median PFS, ORR, OS in some population of patients receiving some regimen of interest) and, importantly, (2) the expert's level of confidence/uncertainty regarding that opinion. Large uncertainty regarding expected patient outcome is common, and the statistical analysis will fully account for this uncertainty. Ultimately, the 2 pieces of information (1) and (2) will define a probability distribution which graphically describes the expert's beliefs.

Figure 1 shows examples of various prior distributions representing differing expert beliefs (and associated uncertainty) for median PFS in months. The peak (mode) of the distribution aligns with the expert's belief regarding the **most-likely** true value. Distributions that stretch over a wide range of values (along x-axis) and only reach lower density values (y-axis) indicate larger uncertainty. Very peaked distributions that rest over a more narrow range of values and attain high density indicate greater confidence in the underlying belief. For example, panels (a) and (b) of Figure 1 both show expert beliefs corresponding to a most-likely median PFS value of 3 months, but with substantially greater underlying uncertainty in that assessment for panel (b). Likewise, panels (c) and (d) both show expert beliefs corresponding to a most-likely median PFS value equal to 6 months, but with substantially greater uncertainty associated with that assessment for panel (d).

In an iterative fashion, the prior distribution implied by the experts' answers will be displayed graphically during the interview, with adjustments being made systematically to ensure the experts feel the shape and location of the distribution are consistent with their underlying beliefs.

Multiple experts will participate in the exercise via independent interviews. The range of opinions will then be collated into a single prior distribution that will ultimately support

statistical analysis of the prospective study results. The expert opinion will be de-identified in all reports/disclosures involving its use (e.g., communications to regulatory agencies, manuscripts, clinical study reports, etc.).

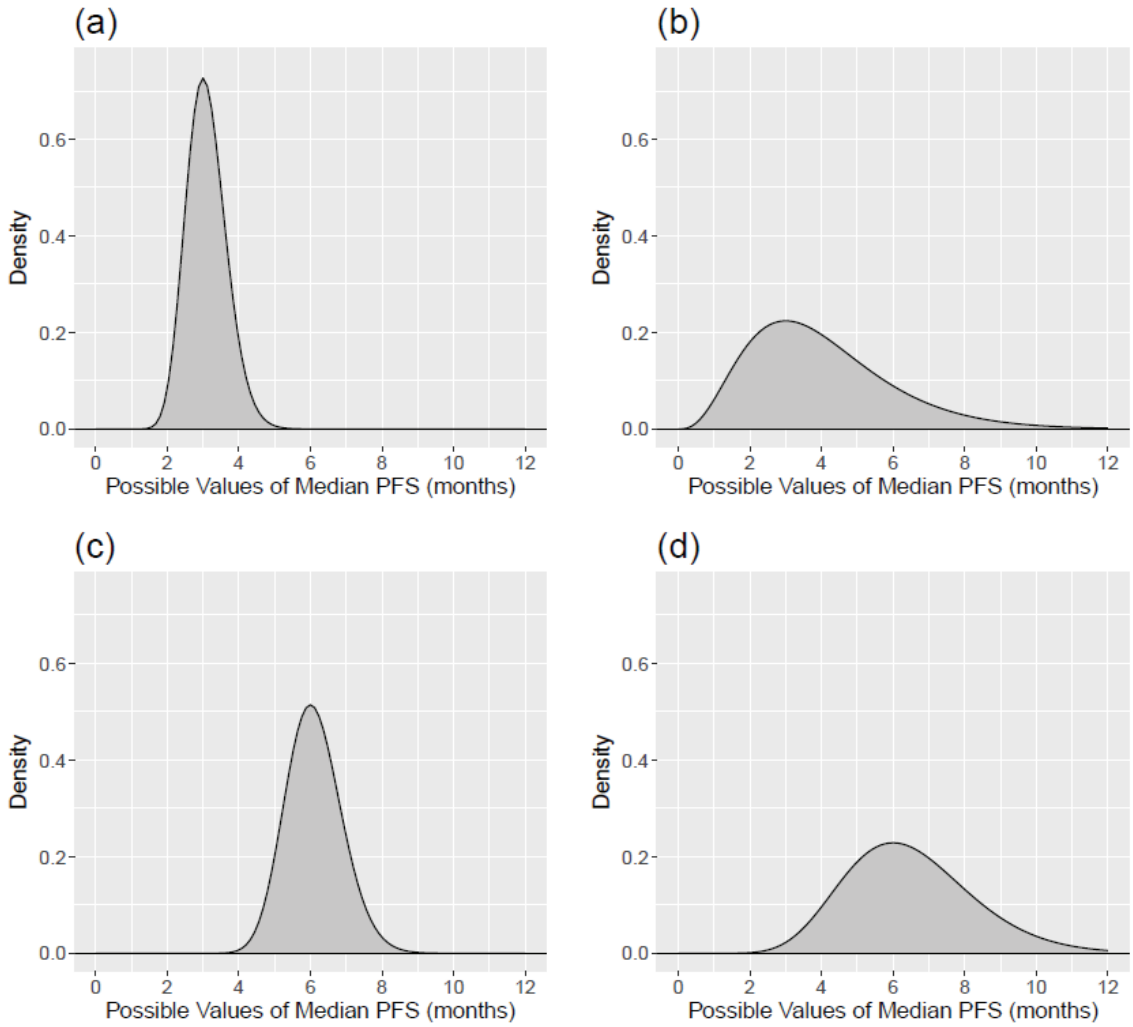


Figure 1. Examples of Prior Distributions Representing Various Expert Beliefs/Uncertainty

3. Interview Content

During the interview, the following questions will be asked of the experts (separately for SS and DSRCT and each chemotherapy backbone/control regimen of interest):

(1) Median PFS overall:

a. Imagine that 100 patients with relapsed, refractory, or progressive [SS or DSRCT] (age 30 years or less, ineligible for surgery, with measurable disease) enrolled on a clinical trial and were assigned to received [backbone/control]. For simplicity,

imagine that all patients enrolled on the same day. How long (months) would you expect it to take for **half** of those patients to have the PFS event?

b. In the scenario from a. above, apart from what you think the mostly likely time would be, what is the most optimistic (but still reasonable) guess for that time that someone could give?

(2) Additional PFS Time

a. Considering the scenario above in (1), for the subgroup of 50 patients that remained progression-free as of the time you gave, how much ***longer*** would you expect it to take for half of *them* to have the PFS event (relative to the date of enrollment)?

b. What is the most optimistic but still reasonable guess for that time that someone could give?

The following table (Table 1-1) will be filled for each patient population separately (synovial sarcoma and DSRCT) and for each regimen per the questions regarding PFS outlined above. Note that 'Chemo Options Overall' refers to all chemotherapy options that are used in the relapsed setting without regard for line of therapy or preference of individual experts or institutions.

Table 1-1

Indication	Control/Backbone Regimen	Median PFS in Initial 100		Additional Time in Latter 50	
		Most Likely	Optimistic	Most Likely	Optimistic
Synovial Sarcoma	Chemo Options Overall				
	Gem/Doc				
	Trabectedin				
DSRCT	Chemo Options Overall				
	Cyclophos (Oral)/ Vinorelbine (28d cycle)				
	Cyclophos (IV)/ Vinorelbine (21d cycle)				
	Cyclophos/Topotecan				

Abbreviations: Gem = gemcitabine; Doc = docetaxel; Cyclophos = cyclophosphamide

3. Papers and Supporting Information

Hampson, Lisa V, et al. "Elicitation of expert prior opinion: application to the MYPAN trial in childhood polyarteritis nodosa." *PLoS One* 10.3 (2015): e0120981.

4. Expert's Background

[To be filled per each expert contracted]

Leo Document ID = 34162b90-0090-410c-90f7-a953d1faf460

Approver: PPD

Approval Date & Time: 19-Jan-2022 14:24:00 GMT

Signature meaning: Approved