9.0 Statistical design with exploratory endpoints –

The primary efficacy endpoint is IgG response, defined as having improvement in IgG level by at least 25% at 6 months, compared to baseline. We are targeting an IgG response rate of 50% and we would not be interested in pursuing further if the IgG response rate is 20% or lower. Using Simon's optimal two-stage design, we will enroll 12 patients in the first stage and the trial will be stopped if there are 3 or less IgG responders among the first 12 patients. If there are 4 or more responders, we will continue the enrollment until 31 patients are reached. The effect of lenalidomide is demonstrated and worth pursuing further if among the total 31 patients, at least 10 patients achieve IgG response. This design is based on a type I error rate of 0.025 and a 90% power. Under the null hypothesis, the probability of early termination is 79% and the expected sample size is 16 patients. Patient enrollment will be suspended if there are not at least 3 IgG responders in the first 12 patients and not all 12 patients have been evaluated at 6 months. To account for a potential 10% early drop-out rate, we will enroll 35 patients so as to have 31 patients with both baseline and 6-month evaluations on IgG. Potential reasons for early dropouts are patients' choice, inability to return to our center or changes in CLL status that may require a different intervention. Sensitivity

analysis will be performed to assess the impact of early dropouts. For example, we will analyze the data counting early dropouts either as "failures" (the most conservative approach) or as "missing" (thus, excluded from the analysis) and evaluate the response rate in both scenarios. Also, since the treatment requires at least 3 months of therapy before seeing an effect on immunoglobulin levels, we will also perform a secondary analysis by including only those patients who have been treated for at least 3 months, while counting those early dropouts between 3 and 6 months as "failures".

Lenalidomide-related toxicity (defined as any grade 3 or greater non-hematological toxicity) will be monitored using the method of Thall, Simon and Estey **[16]** and the study will be terminated early if toxicity occurs in more than 40% of patients. Specifically, denoting the probability of toxicity by p, the study will be stopped early if Prob (p>0.40 data) > .95. The above decision criterion will be applied in cohort size of 5. Assuming a beta (.4, .6) priori for p, the above decision criterion implies that we will stop the trial according to the table below. For example, the trial will be terminated if 4 or more patients experienced toxicities among the first 5 patients.

Number of patients	**Number of patients with toxicities is at least**
5	4
10	7
15	10
20	12
25	15
30	17

The operating characteristics of this study design based on 10,000 simulations are illustrated in Table 2.

Table 2. Operating Characteristics for Toxicity Monitoring Rule.

True toxicity rate	Prob (stop early)	Sample Size 25^{th}, 50^{th} and 75^{th} percentiles
0.20	0.006	35, 35, 35
0.30	0.039	35, 35, 35
0.40	0.152	35, 35, 35
0.50	0.447	10, 35, 35
0.60	0.808	5, 15, 30

For the primary analysis, we will estimate the rate of IgG response along with the 95% confidence interval. As secondary analyses, we will also estimate the rates of Pneumococcus and Influenza seroconversion response, as defined in the response criteria section, and the corresponding 95% confidence intervals. The study will be considered positive in regard to the secondary endpoints if seroconversion is observed in 60% or more of the subjects for at least one of the vaccinations given. Similar method will be used to estimate the rate of influenza or Pneumococcal pneumonia. The duration of IgG response will be estimated using the Kaplan-Meier method. The cytokine levels and serum properdin levels will be summarized using descriptive statistics and will be compared between IgG responders and non-responders using two-sample t-test or Wilcoxon rank-sum test, as appropriate.