Supplemental Statistical Methods

Determination of Response of OMA to Treatment

There are five categories of OMA ratings: stance, gait, arm and hand function, opsoclonus, and mood/behavior. Each of the OMA categories was considered equally important. For each category, an individual patient's response was defined based on a comparison of the baseline evaluation to the "best" of three evaluations at the following time points: two-months, six-months, and one-year. Patients who completed the baseline evaluation of response of OMA to treatment and at least one of the evaluations at the two months, six months, or one-year time points were considered evaluable for response.

The response for each category was
either: CR = complete response;
PR = partial response;
NR = no response; or,
PD = progressively worse.

All five response assessments were taken together to define whether the patient was a "responder" or not, per Supplemental Table 1. Some patients did not have response assessment in all five categories. For however many response categories were reported, the number of CRs, PRs, NRs, and PDs were counted. All possible combinations of those totals are shown in Supplemental Table 1, which we have reproduced from the protocol. The level of detail shown in Supplemental Table 1 is necessary in order for the approach taken in this trial to be reproducible in other studies and for clinical utilization.

Here follows an example of the application of Supplemental Table 1. In the highlighted row, for a given patient with a CR in 3 categories, a PR in 1 category, and an NR in 1 category, the patient was categorized as a Responder. However, if a patient crossed over from NO-IVIG to IVIG+ or switched to ACTH at any time, then the patient was automatically categorized as a Non-Responder.

Statistical Monitoring Rule for Early Elimination of One or Both Treatment Arms for Insufficient OMA Response Rate

Accrue nine patients and randomize to PC vs. PC+IVIG. For the combined group (PC and PC+IVIG):

a) If ≤2/9 respond, then drop the PC only arm, as it is reasonable to conclude that PC is insufficiently active. Continue to accrue to PC+IVIG only. Accrue until there are a total of nine patients on PC+IVIG.
    i) If ≤2/9 respond, then conclude that PC+IVIG is also insufficiently active to warrant further study.
    ii) If ≥3/9 respond, then continue to accrue patients to PC+IVIG until there is a total of 26 PC+IVIG patients.
        1. If ≤7/26 respond, then conclude that PC+IVIG is also insufficiently active to warrant further study.
        2. If ≥8/26 respond, then conclude that PC+IVIG is active.

b) If ≥3/9 respond, then continue to accrue and randomize patients until there are nine on each arm.
    i) If either arm has ≤2/9 responders, then drop that arm and conclude that it is insufficiently active. Continue to accrue on the other "promising" arm until there is a total of 26 patients.

1. If ≤7/26 respond, then conclude that the “promising” arm is also insufficiently active to warrant further study.
2. If ≥8/26 respond, then conclude that “promising” arm is active.

ii) If neither arm has ≤2/9 responders, then continue to accrue and randomize until there are 26 patients in each arm. At the end of the study, a test of proportions at the 0.2 level of significance will be performed to compare the proportion of responders in each treatment arm.

iii) If both arms have ≤2/9 responders, then conclude both treatments are insufficiently active.

Accrual was continued at each step while patients were receiving treatment and the data for response were being reported and tallied. With a low expected accrual rate of 4-9 patients per year, it would have been infeasible to conduct this multi-stage design if the accrual was halted at each decision point.. With 26 patients on each arm, the multi-stage design has Type I and Type II error rates of approximately 0.1, although there was admittedly a slight loss of power for taking the extra look when there were nine patients per treatment arm. The test of proportions has 90% power to detect a 28% difference, and 80% power to detect a 23% difference in proportion of OMA responders for PC versus PC+IVIG.

Supplemental Table 1. Determination of Best Overall Response of OMA according to COG protocol ANBL00P3. The table provides the guide for interpretation of response based on how many of the 5 OMA severity categories described in Table 1 improved, remained the same or worsen and how these changes defined if a subject responded or not to the treatment with IVIG+ or NO-IVIG. The highlighted row is an example: If a patient had a CR in 3 categories, a PR in 1 category, and an NR in 1 category, then the patient was considered a Responder.

Number of OMA ratings categories with a given response assessment (not order dependent)				Overall Best OMA Response	
CR	PR	NR	PD	OMA Responder	OMA Non-Responder
5				√	
	5			√	
		5			√
			5		√
4	1			√	
4		1		√	
4			1	√	
	4	1		√	
	4		1	√	
1	4			√	
		4	1		√
	1	4			√
1		4			√
1			4		√
	1		4		√
		1	4		√
3	2			√	
3		2		√	
3			2	√	
	3	2		√	
	3		2		√
2	3			√	
		3	2		√
2		3			√
	2	3			√
2			3		√
	2		3		√
		2	3		√
3	1	1		√	
3	1		1	√	
3		1	1	√	
1	3	1		√	

Number of OMA ratings categories with a given response assessment (not order dependent)				Overall Best OMA Response	
CR	PR	NR	PD	OMA Responder	OMA Non-Responder
	3	1	1	√	
1	3		1	√	
1	1	3			√
	1	3	1		√
1		3	1		√
1	1		3		√
1		1	3		√
	1	1	3		√
2	1	1	1	√	
1	2	1	1	√	
1	1	2	1		√
1	1	1	2		√
2	2	1		√	
2	2		1	√	
2		2	1		√
2	1	2		√	
2	1		2	√	
2		1	2		√
	2	2	1		√
1	2	2		√	
	2	1	2		√
1	2		2	√	
1		2	2		√
	1	2	2		√
4				√	
	4			√	
		4			√
			4		√
3	1			√	
3		1		√	
3			1	√	
	3	1		√	
	3		1	√	
1	3			√	
		3	1		√
1		3			√
	1	3			√
1			3		√
	1		3		√

Number of OMA ratings categories with a given response assessment (not order dependent)				Overall Best OMA Response	
CR	PR	NR	PD	OMA Responder	OMA Non-Responder
		1	3		√
2	2			√	
2		2		√	
2			2		√
	2	2			√
	2		2		√
		2	2		√
2	1	1		√	
2	1		1	√	
2		1	1	√	
	2	1	1		√
1	2	1		√	
1	2		1	√	
1	1	2			√
1		2	1		√
	1	2	1		√
1	1		2		√
1		1	2		√
	1	1	2		√
1	1	1	1		√
3				√	
	3			√	
		3			√
			3		√
2	1			√	
2		1		√	
2			1		√
	2	1		√	
	2		1		√
1	2			√	
		2	1		√
1		2			√
	1	2			√
1			2		√
	1		2		√
		1	2		√
1	1	1		√	
1	1		1		√
1		1	1		√

Number of OMA ratings categories with a given response assessment (not order dependent)				Overall Best OMA Response	
CR	PR	NR	PD	OMA Responder	OMA Non-Responder
	1	1	1		√
2				√	
	2			√	
		2			√
			2		√
1	1			√	
1		1		√	
1			1		√
	1	1			√
	1		1		√
		1	1		√
1				√	
	1			√	
		1			√
			1		√

A complete response (CR) was defined as improvement from baseline opsoclonus-myoclonus ataxia to normal at any of the first three evaluations (two-months, six-months, or one-year without exacerbation); a partial response (PR) as improvement from baseline OMA to a lesser severity at any of the first three evaluations (two-months, six-months, or one-year without exacerbation); no response (NR) as no change from baseline at all of the first three evaluations (two-months, six-months, or one-year); and progressive disease (PD) as worsening such that none of the reported evaluation time points (two-months, six months, or one-year) are better than baseline and at least one of the time points, the evaluation is worse than baseline. For each of the 5 OMA categories, an individual patient's response was defined based on a comparison of the baseline evaluation to the "best" of three evaluations at two-months, six-months, and one-year. For however many OMA response categories were reported, the number of CRs, PRs, NRs, and PDs were counted.

Supplemental Table 2. Distribution of OMA scores by symptom and time point.

* Each symptom was scored as 0 (no symptoms); 1 (mild symptoms); 2 (moderate symptoms); or 3 (severe symptoms). The Total score at each time point is the summation of all 5 symptom scores. Patients missing 1 or more symptom scores were not assigned a Total score at that time point.

Time Point	Stance	Gait	Arm & Hand Function	Opsoclonus	Mood/Behavior	Total
	Mean (Median) [Range: Min, Max] {IQR: Q1, Q3} <N>					
Baseline	2.3 (2.0) [0, 3] {2.0, 3.0} <51>	2.3 (2.0) [0, 3] {2.0, 3.0} <52>	1.8 (2.0) [0, 3] {1.0, 2.0} <51>	1.7 (2.0) [0, 3] {1.0, 2.0} <52>	1.5 (2.0) [0, 3] {1.0, 2.0} <53>	9.6 (9.0) [2, 14] {8.0, 12.0} <50>
2 months	0.9 (1.0) [0, 3] {0.0, 1.0} <52>	1.2 (1.0) [0, 3] {0.0, 2.0} <52>	1.0 (1.0) [0, 3] {0.0, 2.0} <52>	0.6 (0.5) [0, 3] {0.0, 1.0} <52>	0.9 (1.0) [0, 2] {0.0, 1.0} <52>	4.6 (4.0) [0, 14] {2.5, 6.0} <52>
6 months	0.6 (0.0) [0, 3] {0.0, 1.0} <47>	0.7 (0.0) [0, 3] {0.0, 1.0} <47>	0.6 (0.0) [0, 2] {0.0, 1.0} <48>	0.3 (0.0) [0, 2] {0.0, 1.0} <48>	0.6 (0.0) [0, 3] {0.0, 1.0} <48>	2.6 (2.0) [0, 10] {0.0, 4.0} <47>
1 year	0.6 (0.5) [0, 3] {0.0, 1.0} <48>	0.9 (1.0) [0, 2] {0.0, 1.0} <48>	0.8 (0.0) [0, 3] {0.0, 1.0} <47>	0.3 (0.0) [0, 1] {0.0, 1.0} <47>	0.8 (1.0) [0, 2] {0.0, 1.0} <47>	3.4 (3.0) [0, 11] {0.0, 5.0} <47>