**Supplementary digital content**

Backonja M, et al. Safety and efficacy of neublastin in painful lumbosacral radiculopathy: a randomized, double-blinded, placebo-controlled phase 2 trial using Bayesian adaptive design (the SPRINT trial).

**1. Statistical modeling for adaptive design**

At each interim analysis and at the final analysis, pain and pruritus data were analyzed. The analysis included dose-response models for each endpoint and a longitudinal model for the safety endpoint. The 2 endpoints were combined using a utility function.

*1.1. Primary endpoint modeling*

The outcomes were modeled as:



Where *Yi* is the change in mean average general pain intensity (AGPI) from baseline to 1 week post dose for the *i*th patient, *θd* is the mean response for dose *d*, and the treatment arms are: *d* = 0 (placebo), *d* = 1 (50 µg/kg), *d* = 2 (150 µg/kg), *d* = 3 (400 µg/kg), *d* = 4 (800 µg/kg), and *d* = 5 (1200 µg/kg).

We constructed a dose-response model for mean change from baseline. The placebo arm was modeled separately as:



The active treatment arms were modeled with a first-order normal dynamic linear model (NDLM). The model was a Gaussian random walk model, and the structure was:





While it may appear that these numbers translate to an "informative" prior because the values are small, it may help to remember the scale that the data are on. The pain score is on an 11-point scale, so values for change from baseline will fall between −10 and +10. In this setting, a standard deviation of 5 on the *mean* response is quite reasonable. These values reflect prior belief that mean response on the control (placebo) and the initial dose is within the range of −9.25 to 7.25 with 95% probability. The drift parameter (variance component) *τ2* dictates the amount of smoothing from dose to dose in the model. It was modeled as:



Where *IG(α, β)* is the inverse gamma distribution defined by:



The *τ* parameter is important in the NDLM. Small values of *τ* indicate that the responses of successive doses are likely to be very close, and there is more “borrowing” (smoothing). A feature of the model is that we learn about *τ* based on the data on each of the dose arms. The prior on *τ* helps to dictate the amount of borrowing. The data does help inform this distribution, but the prior matters because we have only 5 doses. We used a prior that corresponds to 1 observation worth of prior data, chosen to give an appropriate amount of smoothing. The variance of the primary endpoint, *σ2* was modeled with prior:



Posterior quantities were calculated using Markov chain Monte Carlo (MCMC) samples from a single chain of length 2500 after a burn-in period of 1000.

*1.2. Secondary endpoint modeling*

Pruritus impact (“the itch is severe enough to cause major problems for me”) was assessed during each week of a patient’s participation in the trial. We let *Xit* be the pruritus impact indicator for the *i*th subject at week *t*, and let *πd* = Pr(*Xi4* = 1 | *di* = *d*) be the probability of pruritus impact at week 4 for dose *d*. We transformed the dichotomous response to a continuous scale for modeling purposes, and let *ξd* be the mean log-odds response, ie:



We constructed a dose-response model for the mean log-odds of pruritus impact on each treatment arm. The placebo arm was modeled separately as:



The dose-response model for the active doses was an NDLM:





When converted from the log-odds scale back to the original probability scale, the resulting prior distribution had 95% probability that the pruritus rate for the placebo arm and the initial dose was between 0.004 and 0.92. The drift parameter *ψ2* had prior distribution:



The parameters of this distribution translated to having 1 prior observation worth of weight on *ψ*.

Posterior quantities were calculated using MCMC samples from a single chain of length 2500 after a burn-in period of 1000.

*1.3. Secondary endpoint longitudinal model*

At the time of each interim analysis, some patients may not have completed the 4-week evaluation period for pruritus impact. A longitudinal model was employed to enable final observations to be imputed for those patients who only had intermediate responses. If *Xit* = 1 at any visit, the final response was imputed as 1. However, suppose that visit *t\** is the last visit for which pruritus impact data is available for patient *i* and that *Xit\** = 0. That is, the patient has not experienced itch that is severe enough to cause major problems as of time t\*. Then, the final response at visit 4 was modeled using a beta binomial model where:



Where *αt*and *βt* are defined as:





In the notation above:



So when *Xit\** = 0, the final endpoint was imputed by inspecting all patients with a known final response (*Xi4*) who also had intermediate response of 0 at visit *t\**. The *αt\** term counts all such patients for whom *Xi4* = 1, while the *βt\** term counts the patients for whom *Xi4* = 0.

These values set the beta prior with 2 observations worth of weight that the prior probability of pruritus impact at week 4 for a patient without pruritus impact in the currently available data is 0.20. This model was used solely for the multiple imputation of final pruritus impact for patients who had not completed 4 weeks at the time of interim analysis. As such, the effect of this prior diminished as data matured and more patients provided final 4-week values.

The longitudinal responses for the placebo arm were modeled using the same setup, but were fit separately from the active doses.

*1.4. Utility functions*

The adaptive aspects of the trial were based on a utility function. The utility function was a bivariate function of the 2 endpoints, constructed by specifying the 1-dimensional component for each endpoint and then combining the 2 multiplicatively. The primary endpoint component of the utility function, *U*1, was defined relative to the placebo arm. We let *M* = *θd* − *θ0*. Then:



The secondary endpoint component of the utility function, *U*2, was defined as:



The utility function was then the product of the 2 component utilities:



**Fig. S1** shows a graphical representation of the 2 component utilities.

**Figure S1.** Component utility for (A) mean change from baseline AGPI score relative to placebo, and (B) severe pruritus impact rate.

****

**Fig. S2** shows a contour plot depicting the 2-dimensional utility function obtained as the product of *U*1 and *U*2. Scenarios along each contour line have equivalent utility.

**Figure S2.** Contour plot of the utility function.



**2. Allocation**

There was an initial burn-in period in which 35 patients were allocated in a ratio of 10:5:5:5:5:5 to placebo and each of the 5 active dose arms, respectively. After this initial burn-in, adaptive randomization began. Adaptive randomization probabilities were updated on a weekly basis. The randomization occurred in blocks such that 2 out of every 7 patients were allocated to placebo. The remaining 5 patients in each block were randomized to the 5 active doses. The randomization probability for each of the 5 active doses was weighted according to the variance components:



Where Pr(*d* = *dUmax*) is the probability that dose *d* is the dose with highest utility, *Var(Ud)* is the posterior variance of the utility at dose *d*, and *nd* is the current number of patients allocated to dose *d*. The *Vd* were normalized to sum to 1. Probabilities <0.03 were set to 0 for all subsequent interims (permanently dropping the dose), and the probabilities for the remaining doses were renormalized.

**3. Statistical power of the study**

The study planned to enroll 165 patients in order to provide 150 evaluable patients with the primary endpoint. The maximum sample size was determined by simulation so that the study would provide ~75% power to detect a treatment difference of 1.5 points between the most efficacious dose and placebo in the primary endpoint if the final data were analyzed using traditional analysis of variance tests. The simulation assumed an efficacy scenario with the most efficacious dose (in this simulation, the highest dose) having 1.5 points difference from the placebo. A common standard deviation of 2.2 and a dropout rate of 10% also were assumed.