

Sustained Pain Response—Statistical Analyses of this Endpoint in Study 301 and Study 304

Study MT100-301 specified that logistic regression would be used for the analysis of the primary endpoint-sustained pain response at 24 hours. This analysis found a non-significant difference ($p=0.08$) for this endpoint in the comparison of MT 100 and naproxen sodium. During a teleconference between POZEN and the FDA, POZEN presented a supplemental analysis of the primary endpoint in Study MT100-301 using ordered logistic regression. This post-hoc analysis was performed because logistic regression of the sustained pain-free endpoint indicated that MT 100 was statistically significantly superior to naproxen sodium (15% vs 11%; $p=0.05$). Since sustained pain-free is a subset of sustained pain response, the statistical analysis of sustained pain response was refined to weight the proportion of sustained pain-free responders differently from the proportion of sustained pain responders. Ordered logistic regression maintains the inherent order from “no response” to “pain response” to “pain-free”.

The FDA minutes of the meeting of March 27, 2000 stated:

“The sponsor agreed to use the 2-hr sustained response rate as primary endpoint in their second study. The Division reiterated that we consider the first study as negative. The sponsor proposed using an ‘almost positive’ study 301 and a positive second study to demonstrate superiority over its components.....The division agreed to this approach.”

There have been many discussions between POZEN and the Division concerning the protocol-specified analysis of the primary efficacy measure (sustained pain response) for Study 304. POZEN submitted results in NDA 21-645 using ordered logistic regression (OLR) with pooled sites as strata and baseline pain as a covariate. Results of this analysis yielded a p-value of 0.03 for the comparison of MT 100 to naproxen for sustained relief of headache pain, indicating that metoclopramide contributes to the effectiveness of MT 100. The Division requested that POZEN reanalyze and resubmit the results according to the protocol-specified method. The Statistical Methods Section of the protocol for Study 304 states that the primary efficacy endpoint for the study “will be analyzed by methods for 3 ordered categories such as extended Mantel-Haenszel statistic with score of 0, 1, and 2 for the three ordered categories and using a model that controls for center, baseline pain and gender.” POZEN, using NPARCOV, a macro developed by Gary Koch, Ph.D., obtained $p=0.038$ for the comparison of MT 100 to naproxen. The Division’s analysis using NPARCOV yielded $p=0.063$. As pointed out in the NAL, this difference arises from POZEN’s decision to weigh the selection of pooled sites equally (i.e., choosing the option $c=0$ in NPARCOV) while the Division used unequal weights ($c=1$). POZEN agrees that $c=1$ may be more consistent with the fact that the pooled sites, which are arbitrarily pooled to be more equal in size, vary somewhat in relative contributions of patients to each treatment group. Therefore, POZEN does not dispute the findings of the Division regarding results of the analysis for sustained pain response for Study MY100-304 using NPARCOV ($p=0.063$). However, the NAL states that evaluating the study results by stratifying only by pooled sites leads to $p=0.09$ by both NPARCOV and by the Division’s own SAS program. (POZEN assumes that the Division’s SAS program is SAS PROC FREQ.) The fact that both apparently dissimilar

methods yielded the same result prompted POZEN to question the assumptions underlying the macro. POZEN has found that both NPARCOV and the method the Division used to obtain $p=0.09$ assume that, rather than being merely rank ordered, the ordinal outcome values are spaced at equal intervals.

Data from Study MT100-304, with only the stratifying variable (i.e., pooled site), indicate that NPARCOV produces p-values equivalent to the CMH test for row mean score differences (obtained from SAS PROC FREQ with "Table Scores" as the scoring option). PROC FREQ, however, can only adjust for additional covariates by creating additional strata, while NPARCOV can adjust for them as continuous variables. Hence, with more than one adjustment variable, the two methods produce different p-values. Both NPARCOV and PROC FREQ with Table Scores assume the outcome is not just an ordinal, but an "interval" variable. The actual outcomes are simply coded ordinally as 0=no sustained pain response, 1=some sustained pain response and 2=sustained pain free. The values assigned do not have interval meaning, (i.e., the three-point scale does not assume that "some sustained pain response" is exactly equidistant between "no sustained response" and "sustained pain free" as equally spaced values (0, 1, 2) imply). To illustrate this point, suppose that "sustained pain free" were 50% better than "some sustained pain response" and that "some sustained response was 100% better than "no sustained response," the outcomes would be given values with differently spaced intervals, (i.e., 0, 1 and 1.5). Similarly, scores of 0, 1 and 3 would imply "some sustained pain response" was 100% better than "no sustained response" and 200% worse than "sustained pain free." NPARCOV and CMH (using Table Scores) would both find different p-values for these different scorings. Table 1 summarizes p-values obtained from NPARCOV and CMH (with Table Scores) when outcome values are spaced differently for Study 304.

Table 1: P-values for Sustained Pain Response Comparisons of MT 100 to Naproxen Sodium in Study 304--When the Response Values Have Interval Meaning

Method (adjusted for pooled sites, baseline pain and gender)	Assumptions	Outcome Values Assigned		
		All outcomes are Equal Distance (0, 1, 2)	Unequal, Pain Free not much better than pain response (0, 1, 1.5)	Unequal, Pain free twice as good as pain response (0, 1, 3)
NPARCOV (c=0)	Strata are given equal weights	0.038	0.029	0.061
NPARCOV (c=1)	Strata are weighted by the number of patients	0.063	0.043	0.11
CMH (with Table Scores option)	Strata are weighted by the number of patients	0.086	0.043	0.19

POZEN believes that there is no basis for making an assumption about the relative relationship of the outcomes other than rank ordering and made no such assumption in drafting the Study 304 protocol. Neither NPARCOV, as used in POZEN's (c=0) and the Division's (c=1) analyses, nor CMH with Table Scores is an appropriate analysis method.

There are three score options available within SAS PROC FREQ for CMH that rank the outcome values, but differ in how weights are applied to the strata: (1) Rank scores, (2) Ridit scores, and (3) modified Ridit (MODRIDIT). Because each of these score options utilizes different strata weighting, p-values obtained using the three options will not necessarily agree. However, the interval distances per the outcome values do not affect the p-values.

The NPARCOV macro can also accommodate ordered categories or outcomes by grouping the outcome values into dichotomous outcomes and then performing the analysis for proportional odds (using options `hypoth=1`, `c=1` and `method=prosmall`). Because the outcomes are reduced to dichotomous groupings, NPARCOV will give the same p-values regardless of the scores used for the original outcome values.

Ordered Logistic Regression (OLR), which was used in the original submission of NDA 21-645, is a parametric regression analysis of ordered categorical outcomes. Like CMH with rank, ridit or modified ridit scores, OLR makes no assumptions about the interval distances for the outcome values. In addition, unlike CMH, OLR adjusts for covariates in a regressive process which minimizes the potential for over-stratification.

Table 2 summarizes p-values obtained from NPARCOV (using proportional odds with options `hypoth=1`, `c=1` and `method=prosmall`), OLR, and CMH with ranks, ridits and modified ridit scores.

Table 2: P-values For Sustained Pain Response Comparisons of MT 100 to Naproxen Sodium in Study 304--When the Response Values Have Ordered Meaning (Ranks)

Method (adjusted for pooled sites, baseline pain and gender)	Assumptions	Outcome Values Assigned		
		All outcomes are Equal Distance (0, 1, 2)	Unequal, Pain Free not much better than pain response (0, 1, 1.5)	Unequal, Pain free twice as good as pain response (0, 1, 3)
NPARCOV (options hypoth=1, c=1 and method=prosmall)	Outcome values are reduced to dichotomous groups, strata are weighted by the number of patients in each stratum	0.038	0.038	0.038
Ordered logistic regression	Outcome values are ordered, strata are adjusted as covariates in a regression approach	0.029	0.029	0.029
CMH (with RANK scores)	Outcome values are ordered, strata are given equal weights	0.008	0.008	0.008
CMH (with RIDIT scores)	Outcome values are ordered, strata are given weights based on strata size	0.019	0.019	0.019
CMH (with MODRIDIT scores)	Outcome values are ordered, strata are weighted by the number of patients	0.016	0.016	0.016

When all other protocol-specified strata are applied, all of these methods consistently show MT 100 to be statistically superior to naproxen sodium for Study MT100-304. The same methods can be applied to data from Study 301. For the comparison of MT 100 to naproxen sodium, CMH with rank scores yields $p=0.124$, while CMH with ridit and modridit scores yields $p=0.053$ and $p=0.059$, respectively. Results from analysis using the NPARCOV macro with proportional odds were similar ($p=0.05$), while results from OLR yielded $p=0.025$. Given these analyses, POZEN believes that Studies MT100-301 and MT100-304 can each be considered to demonstrate a statistically significant contribution of metoclopramide to the claimed effect of MT 100.