From: Loring, David W
Sent: Monday, September 30, 2013 7:11 AM
To: Dan Drane
Subject: Lorazepam WMT reanalysis comments

Hi Dan – Thanks for passing along Paul Green's request for me to comment on Marty Rohling's reanalysis of my Lorazepam WMT RCT data.

Could you please forward this to Paul's list? Thanks, David

_____

Group –

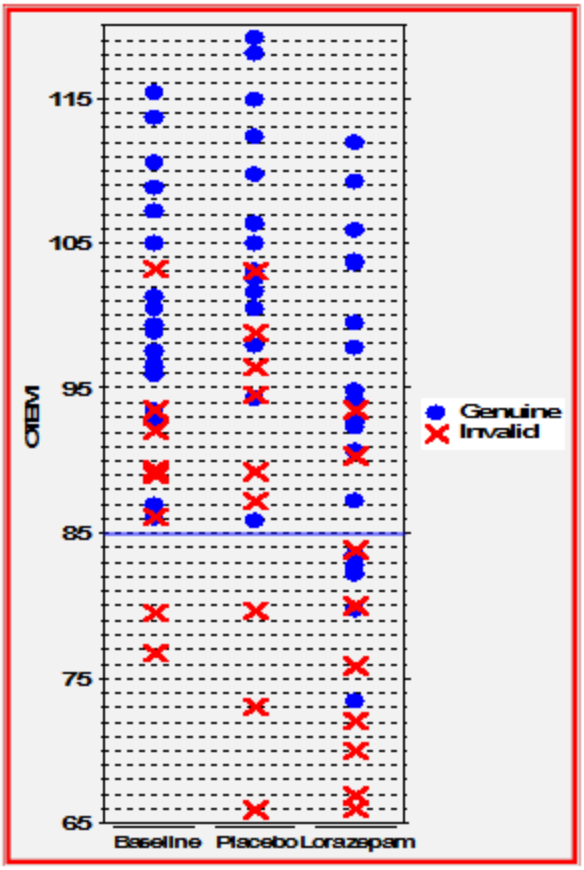
It seems premature to address Marty Rohling's conclusions from his reanalysis of my lorazepam WMT RCT data since it's a poster and not a peer-reviewed publication, particularly since I actively avoid posting on symptom validity where, like politics and religion, there is little likelihood to move pre-conceived notions. However, the backchannel email chatter has raised many points that I cannot allow to pass without formal comment. Although Marty already shared with me his initial findings, it was not until this weekend that I gave the poster a careful read. Having done so, there are multiple concerns that I will detail in this single post.

I am not a member of this listserv, so forgive me if my points have already been addressed.

Although the precise number exceeds my competence to compute it, I think it is sufficient to note that 28 (subjects) * 16 (embedded validity indices) * 3 (conditions) is pretty big (1344) and is associated with an uncomfortably high alpha error rate.

More importantly, however, is that the poster's bottom line is based on an assumption that the reported performance validity measures actually mean something. Regardless of how they are derived, these indices have not been characterized in either patient or simulator studies. Although I am not a fan of dichotomizing continuous variables, at least the WMT and MSVT cut-points have been described in multiple peer-reviewed papers, and their error rates are relatively well known from independent review (e.g., Sollman & Berry, ACN, 2011). I think it is misleading to label performance as INVALID in Marty's scatter gram without the newly derived performance indices having some validation themselves.

The scattergram (below) itself nicely make this point. I personally find it hard to reconcile that the number of "Invalid" performances, particularly for Baseline and Placebo, for OTMB scores in the "normal" range, which I broadly define as OTMB ≥ 85.



The number of "invalid" scores is greater in the lorazepam group (n=7) with OTMB < 85 reflects drug effects on task performance and is greater compared to placebo (n=3) or baseline (n=2). Subject 20 in Marty's Table 2 is classified as "Global Invalid" due to failure during lorazepam only, but with "valid" scores during baseline and placebo. Antiepilepsy drugs are known to increase performance variability, and CNS Vital Signs performance is sensitive to the effects of 2 mg lorazepam (this was the entire reason the study was funded: http://www.sciencedirect.com/science/article/pii/S1525505012005379. Derived measures from this dataset suffer from contamination artifacts that cannot be disentangled.

From the poster: "Data were then examined using a mixed repeated measures ANOVA, with the main effects being condition and validity status. This analysis revealed a main effect for trial ($p < .0001$) and an main effect for validity status ($p < .0001$); however, the interaction term was not significant ($p = .37$). **Therefore, the S's who obtained the lowest scores during the lorazepam trial also obtained an equivalent level of low scores during baseline and placebo trials."**

I disagree with this interpretation. The nonsignificant interaction demonstrates that the relative (emphasis on "relative") performance of "valid" versus "invalid" subjects did not change across study conditions. The inference that "It would appear that the invalid scoring S's would have done just as badly during the baseline testing on the WMT and MSVT if they had been administered as they did during the lorazepam trial" inaccurately interprets the lack of interaction. There is no evidence suggesting that "invalid" subjects would have performed at the same absolute level in baseline as in

lorazepam phase – only that the relative performance is the same across the three phases (see Figure 1) in which both "valid" and "invalid" subjects perform better in baseline and placebo conditions.

For fun, I correlated the data from 11 subjects presented in Table 2 and found no significant pair-wise correlations in the expected direction for "validity" classification (see attached; 1=baseline, 2=lorazepam, 3=placebo). Whatever is being measured, it is not consistent in these subjects. It fails a primary criterion of validity, namely consistency, and the algorithms used to characterize performance as "invalid" or not do not have statistically significant reliability in this dataset.

Independent from these specific data, it is important to note that there is an opposite perspective to Marty's suggestion that clinical trials need to carefully weed out all subjects with poor cognitive validity scores. At Emory, we have observed poor WMT in epilepsy patients with active left-hemisphere discharges. If these subjects are excluded from study recruitment, then the sample being studied is not representative of the population, and any cognitive benefit derived successful treatment intervention may be missed (Type II) or its magnitude underestimated. If performance more directly reflects "effort" rather than diminished attentional resource allocation from active EEG discharges, this too can introduce systematic sample bias. For example, if patients with major depressive disorder are screened and excluded from study participation based solely upon poor validity scores on cognitive assessment, then there is a very real risk that patients with greater depressive symptoms will be excluded, a group with perhaps the greatest need for new therapeutic benefit to be developed.

I've always considered Marty's poster to reflect an interesting academic exercise and nothing more.

I think any suggestion that a double blind RCT demonstrating a differential effect is somehow flawed whereas a post-hoc analysis deriving an extraordinary large number of variables with uncertain validity somehow represents more accurate characterization of the dataset will be met with skepticism by most reviewers and editors.

To those who may disagree, go home, take 2 mg lorazepam, and email me your experience <g>.