

The difficulties in going from P2 to P3 in CNS trials  
*Red flags from a recent CIS program*

ISCTM  
February 2017

# CIS Phase 3 study design (EVP-6124-015/016)

Two identical, global, placebo controlled, 6-month studies

A placebo run-in period preceded the baseline and also had at least one MCCB session

Each study enrolled ~750 patients with stable schizophrenia treated with up to two atypical antipsychotics

Placebo, 1.0 and 2.0 mg doses of encenicline HCl

Co-primary endpoints:

- Neurocognitive component (NCC) of the MCCB (all the tests except for the MSCEIT)
- SCoRS (Clinical function related to cognition: Interviewer total score)

Key Secondary endpoint:

- Marder Factor- measure of negative symptoms (derived from the PANSS)

## Summary of topline CIS data results

Baseline demographics were essentially evenly distributed across all treatment groups

Strong baseline (and pre-baseline) to endpoint cognitive effects were seen – unfortunately this was seen across all three dosing groups.

- MCCB not significantly different from placebo in the ITT population
- Some differences noted between US and ex-US patients in -016

Some modest numerical improvement in SCoRS (functional endpoint), again across all three treatment groups, not statistically significant

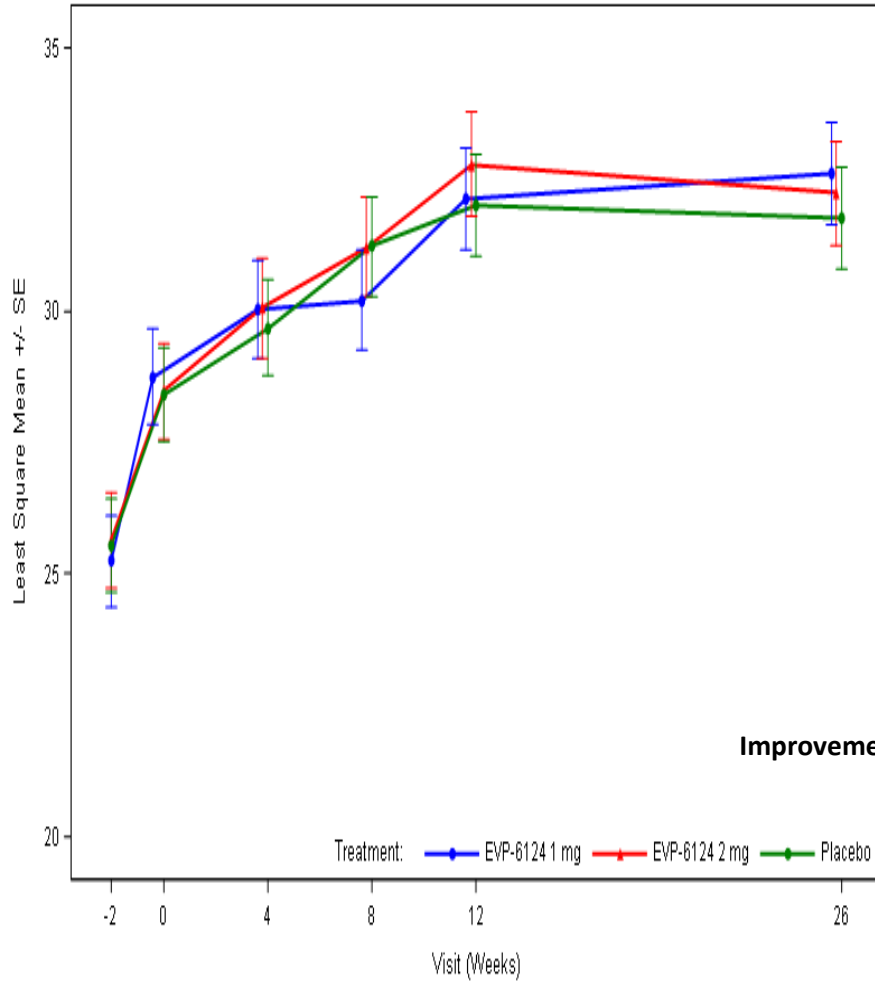
PANSS Negative - key secondary endpoint

- Some differences between encenicline groups and placebo, possible trend, but not quite significant in the ITT study population for either study

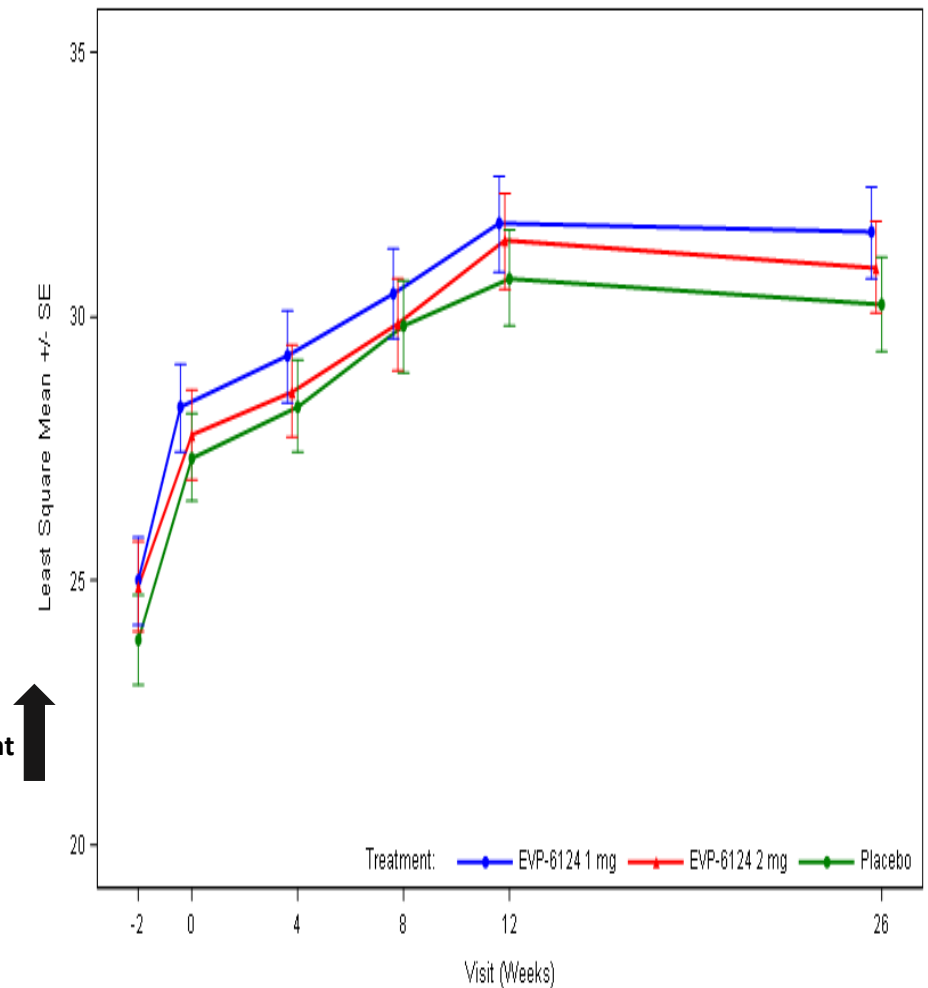
Large changes in the NCC in all arms from baseline (and pre-baseline) to endpoint with little differentiation

Least Square Mean (+/-SE) of neurocognitive MCCB Score over time by treatment group- overall

**EVP-6124-015**



**EVP-6124-016**



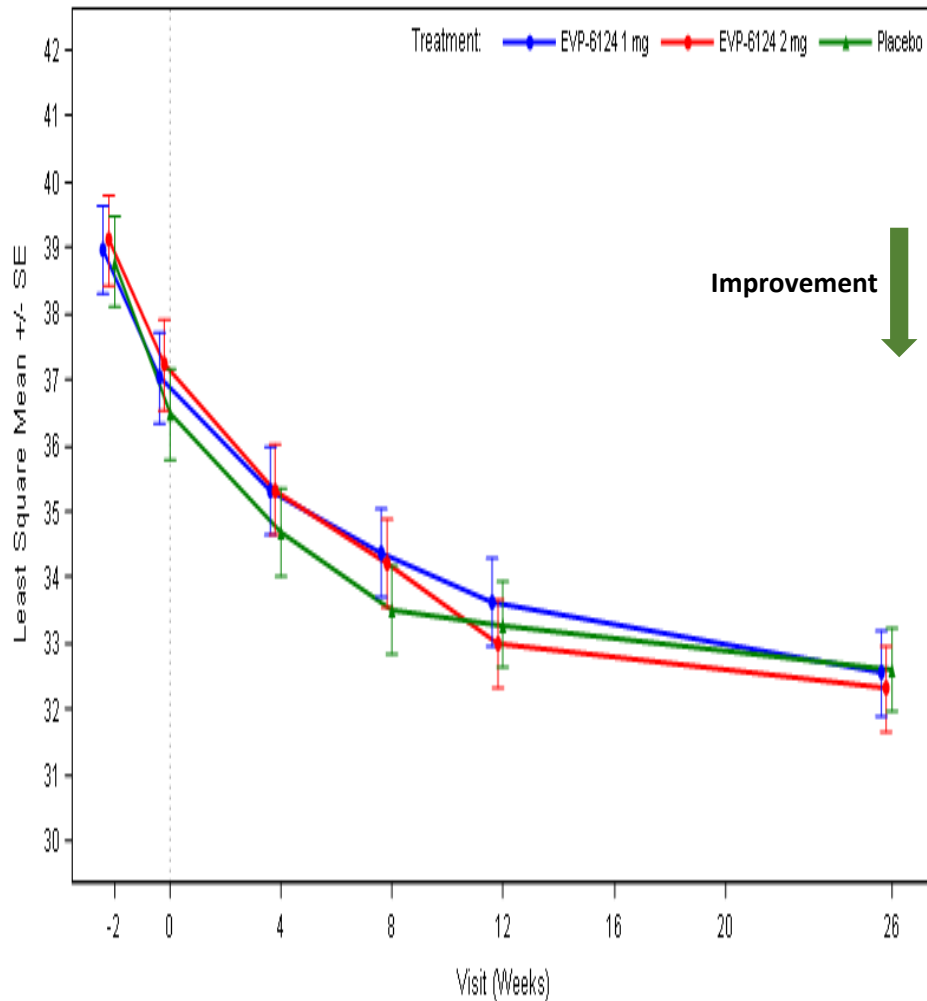
Improvement



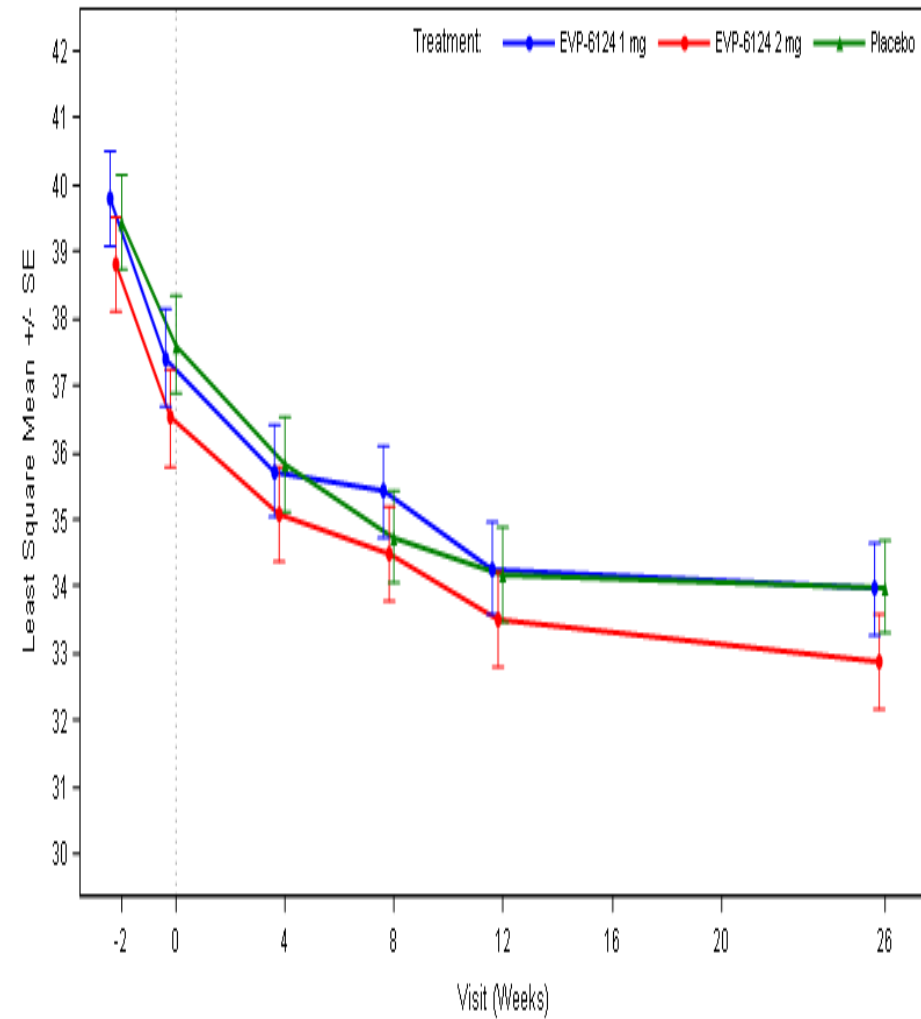
Large changes in the SCoRS in all arms from baseline (and pre-baseline) to endpoint with little differentiation

Least square mean (+/-SE) of neurocognitive SCoRS score over time by treatment group- overall

**EVP-6124-015**



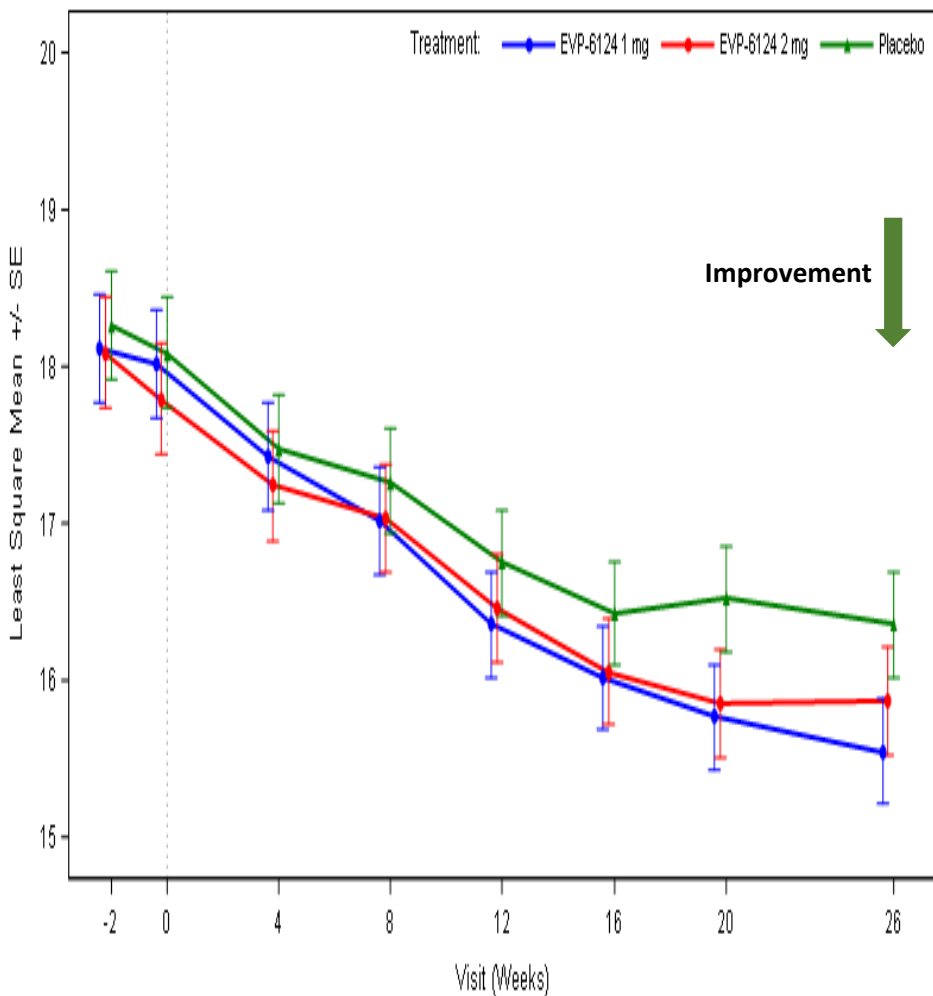
**EVP-6124-016**



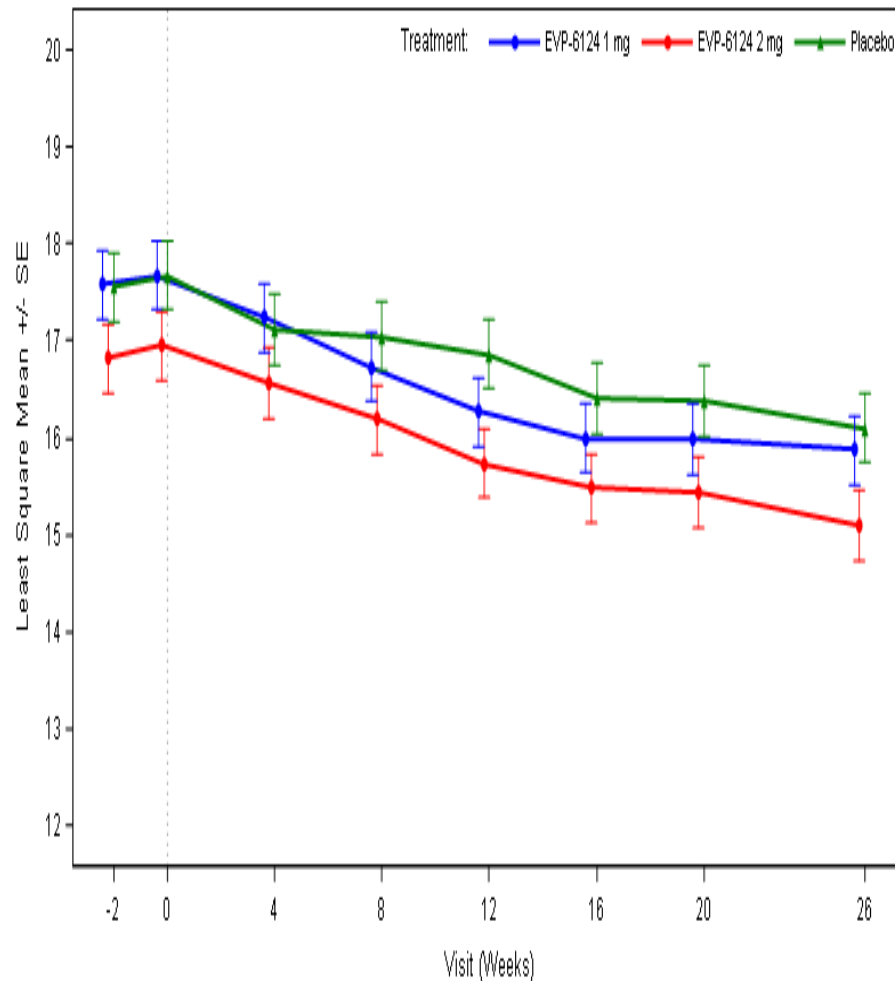
# Some evidence of a trend of at least one drug arm in the scale for negative symptoms

Least square mean (+/-SE) of the Marder factor for the PANSS Marder score (negative symptoms) over time by treatment group- overall ITT

### EVP-6124-015



### EVP-6124-016



# In-depth analysis prompted by anomalies identified in initial trial results

## Initially:

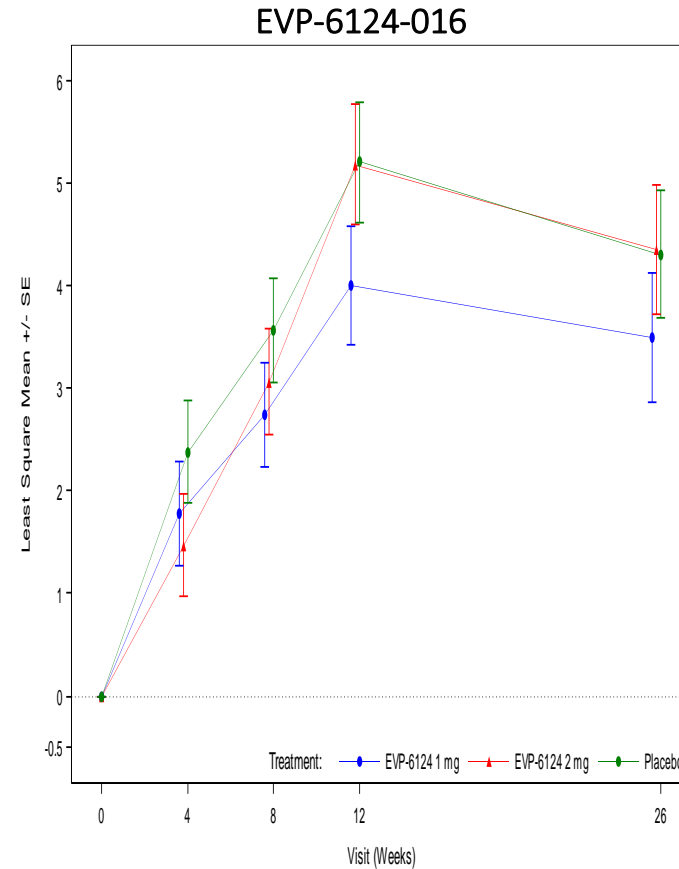
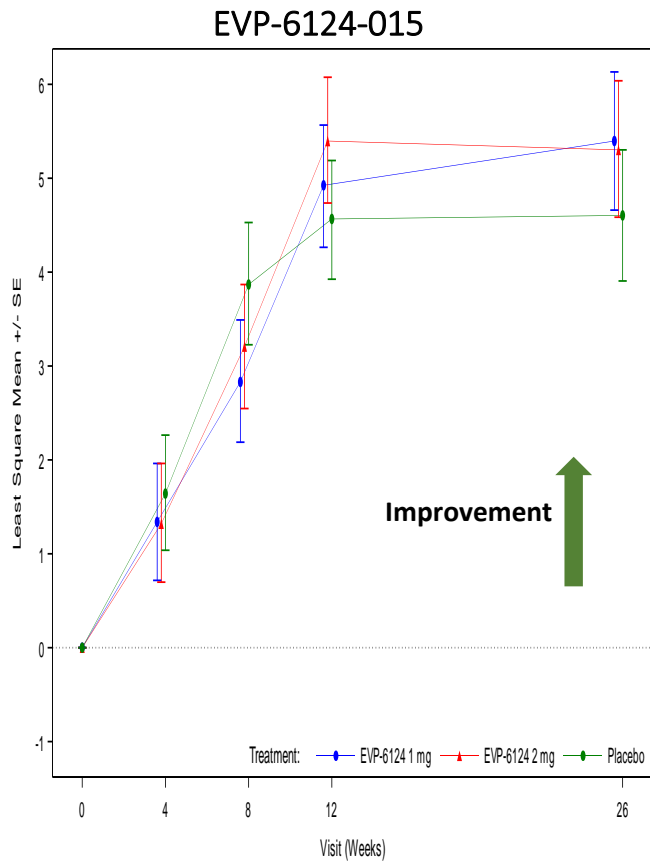
- Unexplained regional variation in co-primary and secondary endpoints
- Large Baseline to endpoint changes in all groups (too large to be believable)
  - Unprecedented learning/placebo rates,
- Observed subgroups that did well in the trial (e.g. African Americans)

## Soon after:

- High rate of non-compliant subjects (5%-18% depending on cutoff)
- Placebo/Learning rate was much lower early in the trial than later in the trial (where it actually exceeded drug response)
  - Relationship to trial amendments
- Patients stabilized on two antipsychotics were found to have a very high placebo response relative to drug
- Suspicious outlier values seen (e.g. pre-baseline changes)

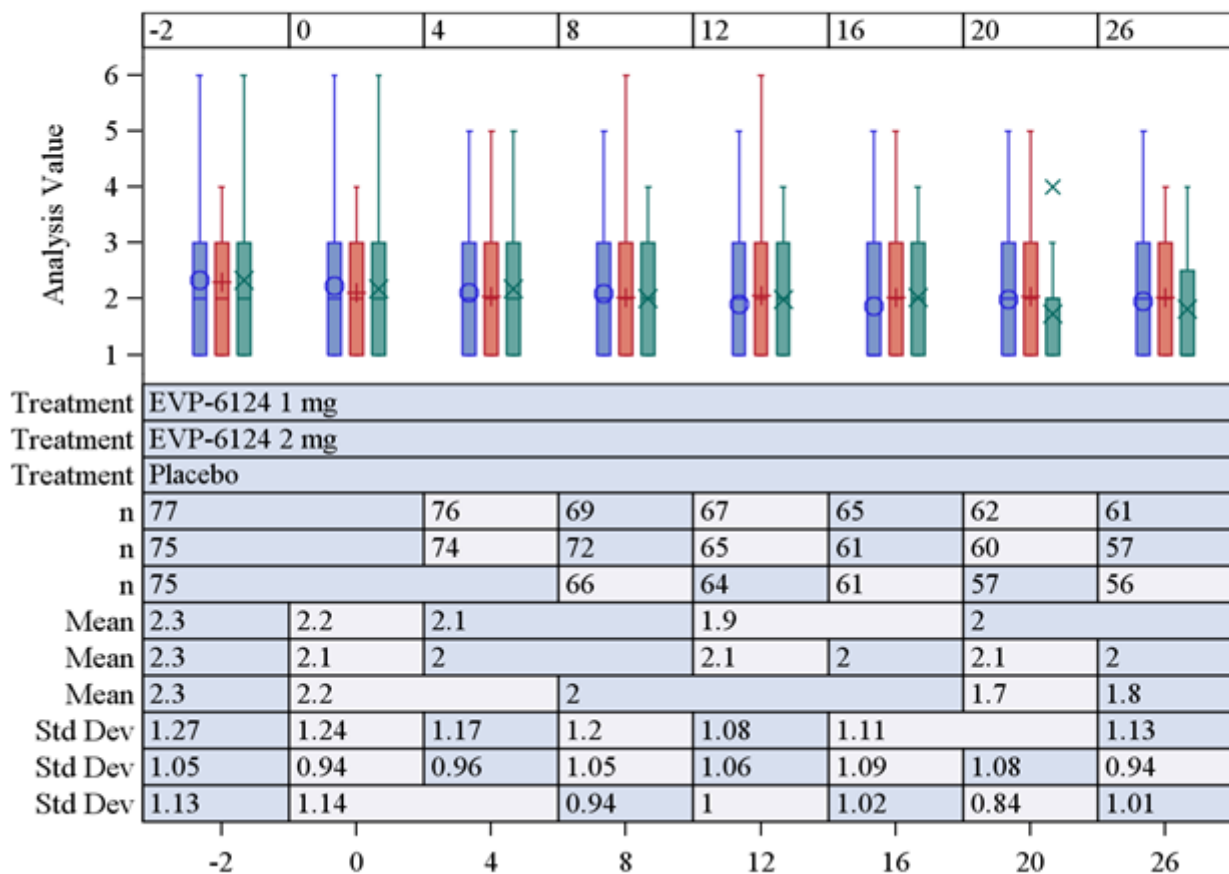
In the overall ITT population for the NCC, European sites average a gain of ~4.5 points in the placebo arms

Least squares means ( $\pm$  SE) of change from baseline in neurocognitive MCCB score over time by treatment group- Europe



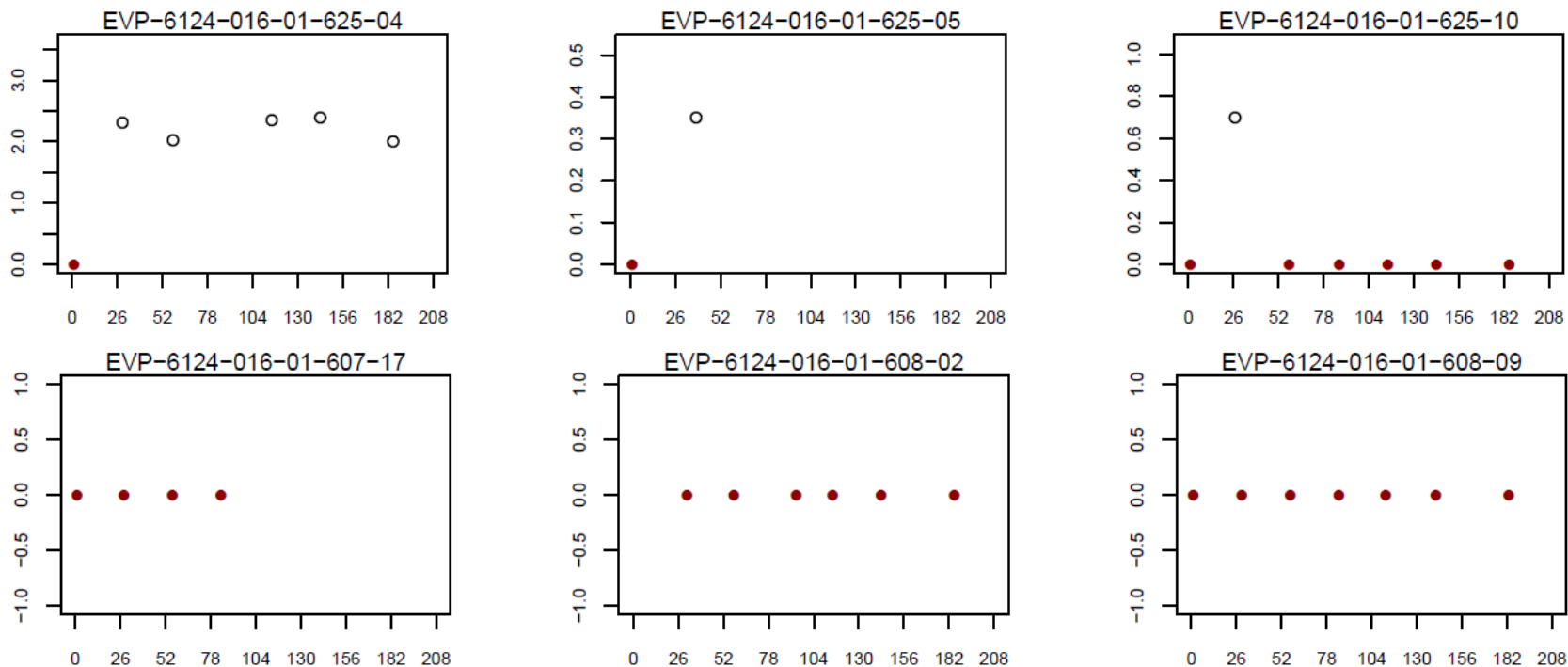
## Sites that Recruited less than 6 subjects Show almost no differences between Treatment Groups

*Boxplot of PANSS Score over Time by Treatment Group  
All Sites with Less Than 6 Patients*



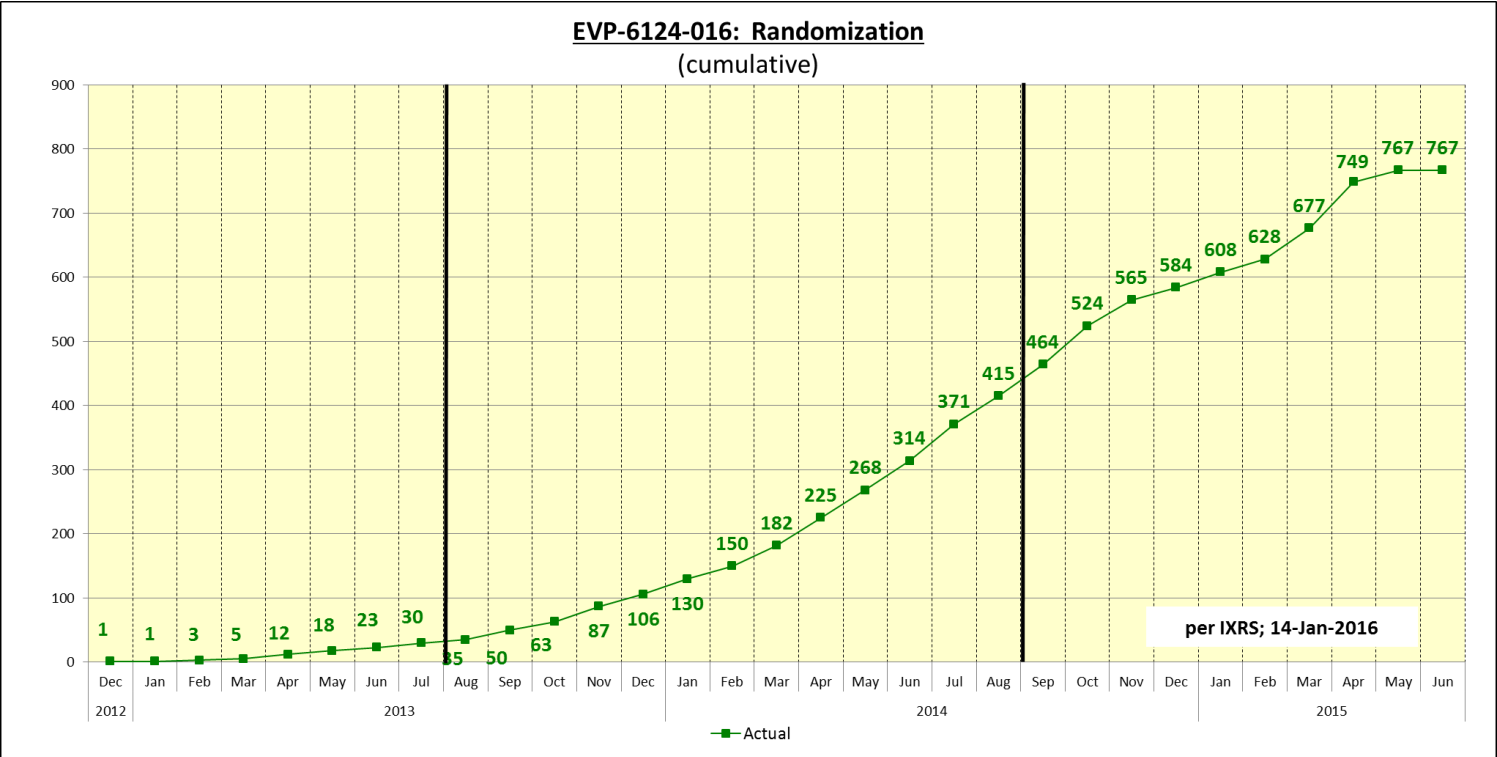
Trial subject compliance by PK showed a number of patterns, including about 5% with apparently zero compliance

Plasma encenicline concentrations during Phase 3 Studies; examples of individual subjects patterns



- 74 (17%) and 89 (18%) of active-treatment subjects in Studies -015 and -16, respectively, had a post-Day 1 concentration of “0” at some time in the study
- 19 (4%) and 26 (5%) of active-treatment subjects in Studies -015 and -016, respectively, had a concentration of “0” at every sampled time in the study

# 016 Randomization: one of the 2 CIS trials, both with essentially the same pattern



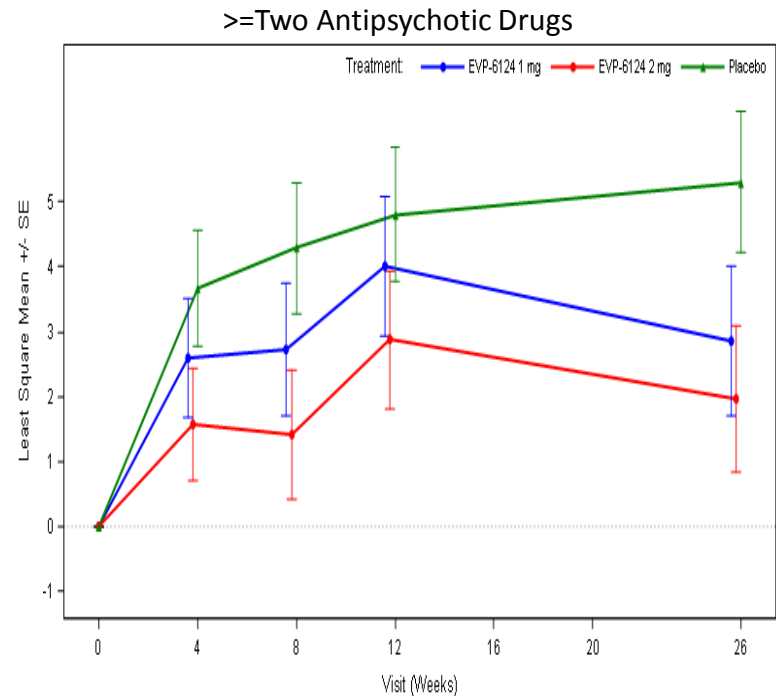
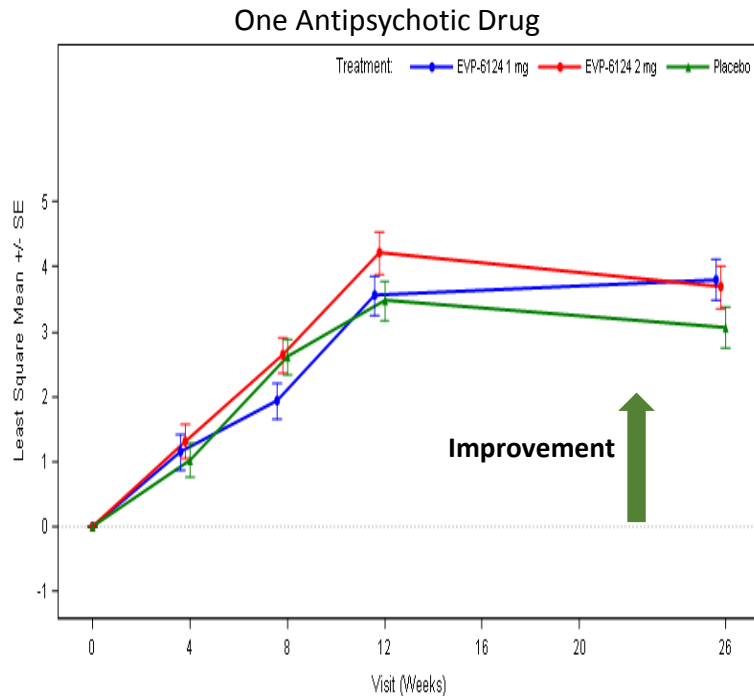
In an effort to accelerate enrollment, after Aug 2013, subjects could be enrolled if their regimen included 2 antipsychotics (prior to that they had to be stable on 1 Antipsychotic)

At the end of the trial, the breakdown of how many Antipsychotics subjects were on was:

- Stable on One Antipsychotic, N= 1343(90.55%)
- Stable on Two Antipsychotics, N= 140 (9.44%)

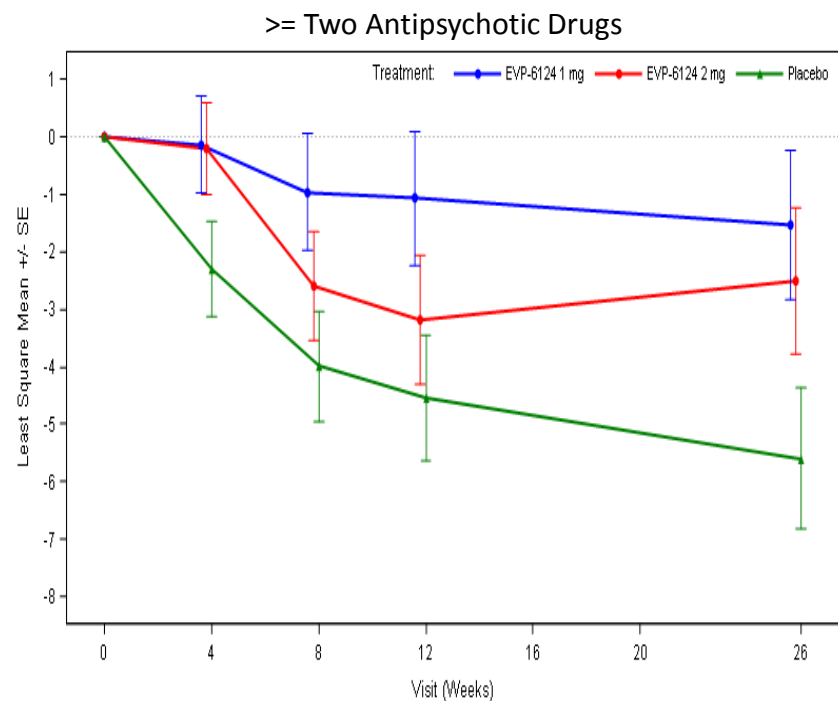
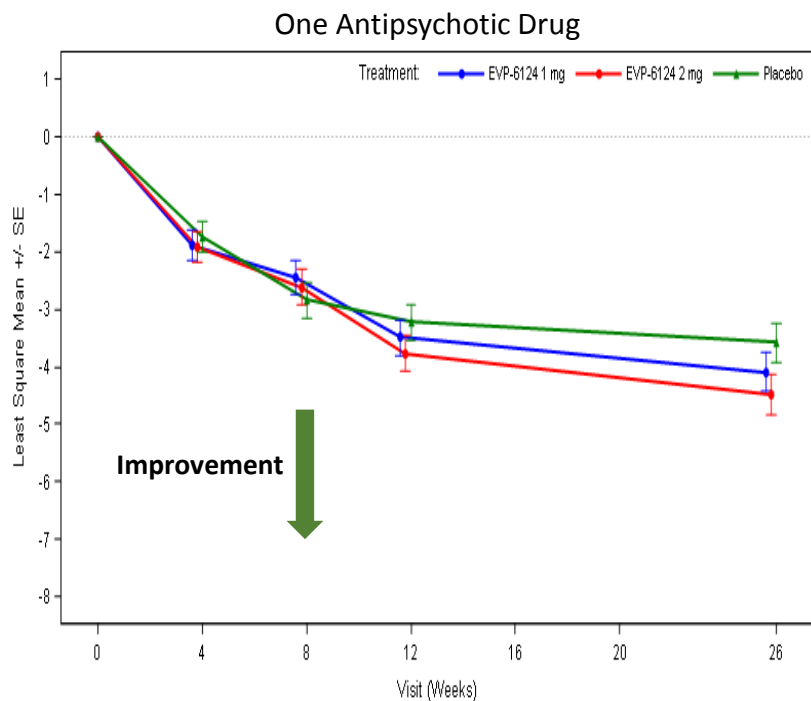
Despite the request from the sites to allow this, only a little less than 10% of subjects enrolled were actually stable on 2 antipsychotics; though there was no reason to believe this group would differ from those stable on 1 Antipsychotic, post hoc analysis showed this not to be the case.

Surprisingly, the 10% of study subjects who were stabilized on 2 Antipsychotics showed a very different placebo response than those on only one antipsychotic



A similar pattern is seen in the functional co-primary with the placebo arm doing markedly better than the drug arms

Least squares means ( $\pm$ SE) of change from baseline overtime in SCoRS scores by treatment group- EVP-6124-015+016



## “Clean” Study Population: No subjects on 2 antipsychotics, non-compliant with Study meds, or with large pre-baseline changes

This data set shows the clearest separation between the drug arms and the placebo arm, particularly for the co-primary outcomes

- Which drug arm is stronger varies between the outcome measures

The placebo arm here behaves like a more typical placebo arm does in trials

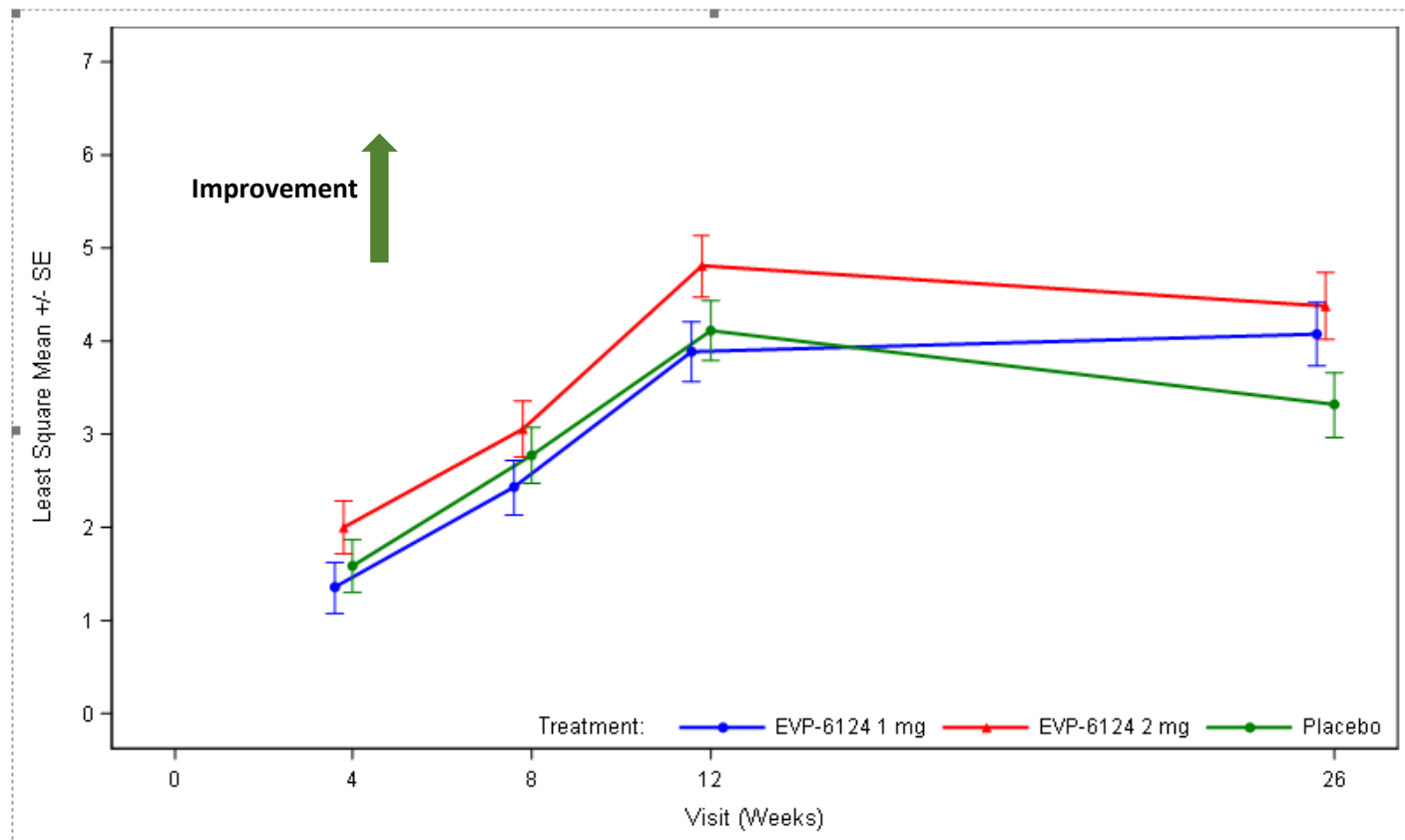
- Less extensive Baseline to endpoint gains
- A somewhat flatter curve over time, particularly at later time points

Though there is separation, the amount of separation is relatively small

- Combining the exclusion of subjects who were noncompliant and those who had large pre-baseline changes was clearly not additive; possibly the 2 groups may have had considerable overlap
- Not sure the difference can be called “clinically significant”, though directionally this analysis is encouraging that additional trials, if conducted carefully, could be successful
- Not clear why there is less “enhancement” with the Negative symptom scale (Marder factor), though this was the one scale that showed trends even with the ITT population (and thus less enhancement possible); it is possible that it was emphasized less in the clinic as it was not a co-primary

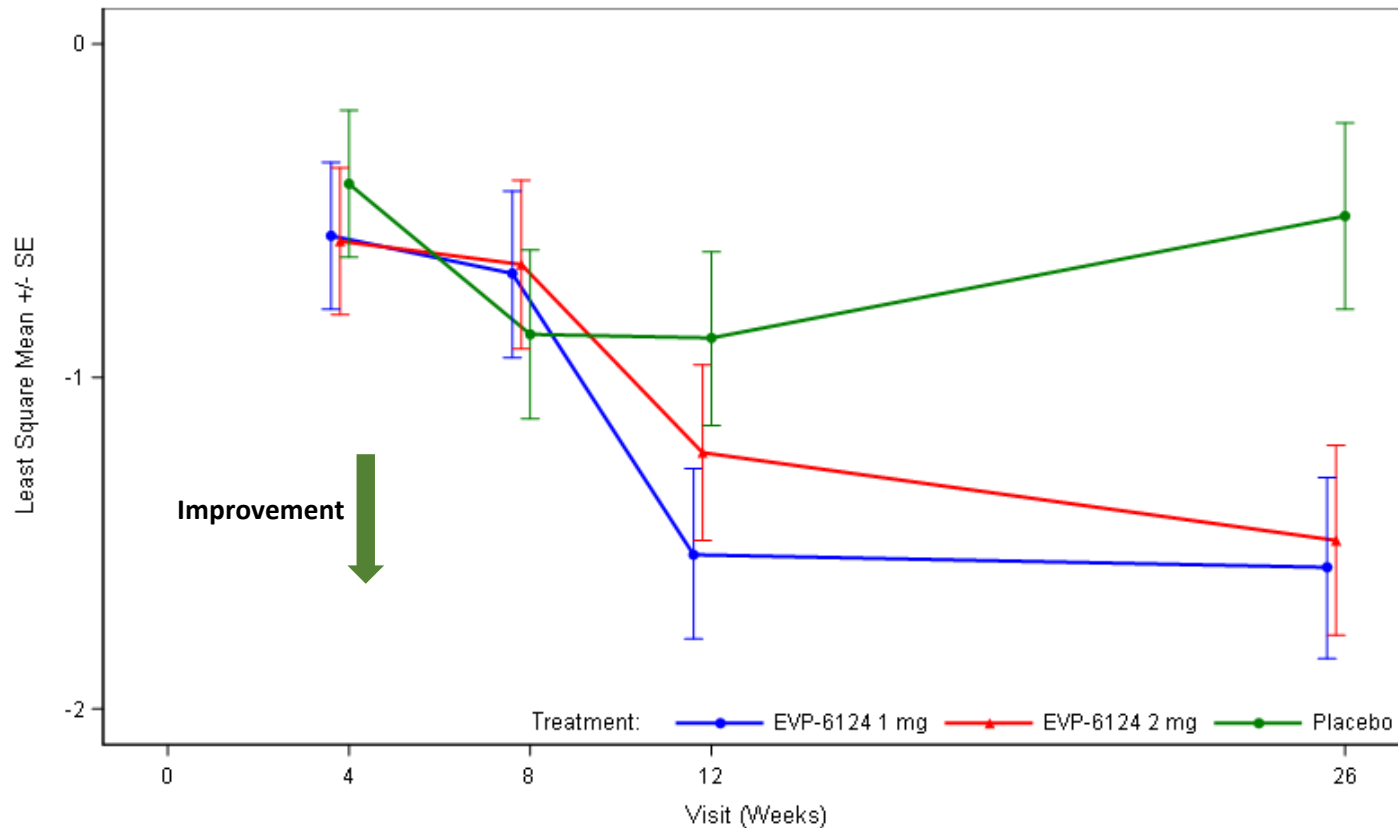
For the primary outcome measure, this analysis shows separation in both drug arms, though it is stronger in the 2 mg Arm

*Figure: Least Square Mean (+/-SE) of Neurocognitive MCCB MCCB Neurocognitive Score over Time by Treatment Group (Clean Patient Population)*



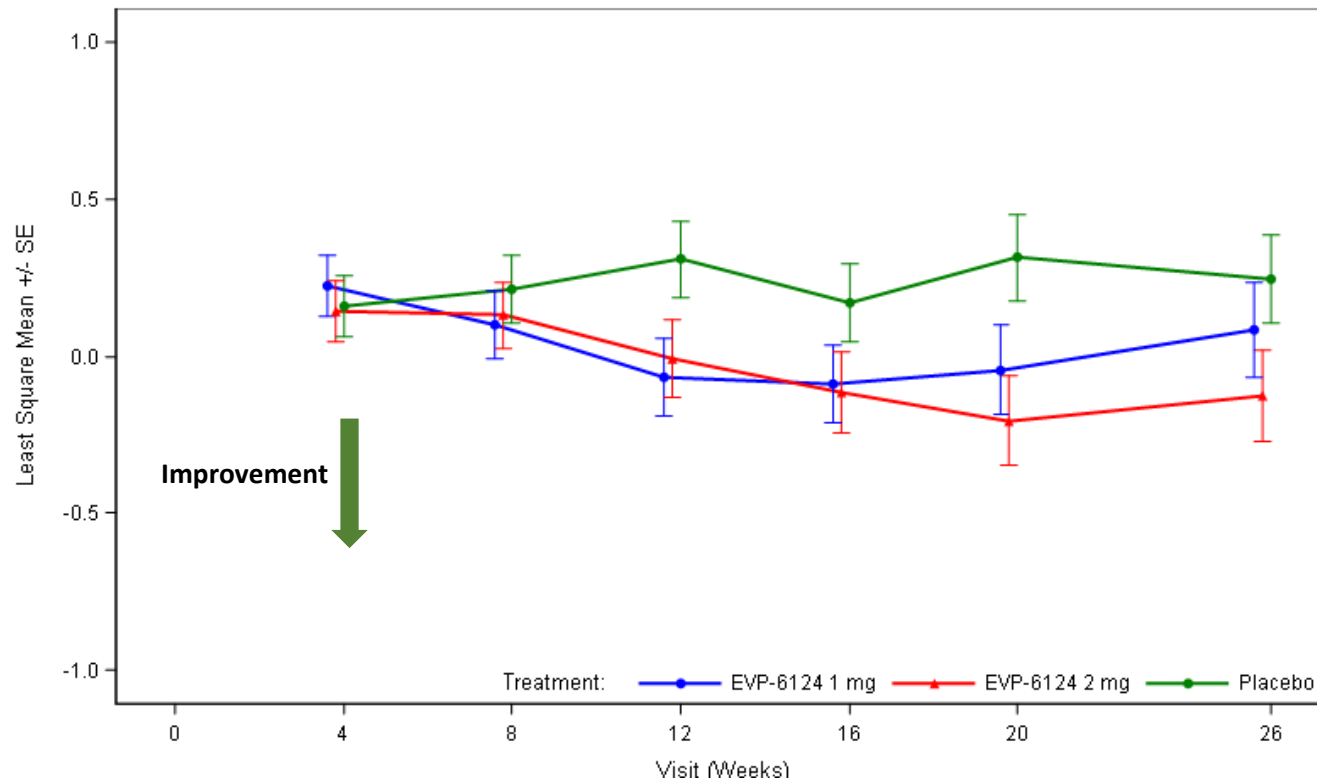
For the functional co-primary, one sees clear separation by the end for both drug arms though here the 1 mg drug arm looks stronger

*Figure: Least Square Mean (+/-SE) of Total SCoRS Score over Time by Treatment Group (Patients with 1 Antipsychotics and Excluding Pts with >=8 points reduction from Screening to Baseline and Without Any Undetectable PK During study) Clean Population*



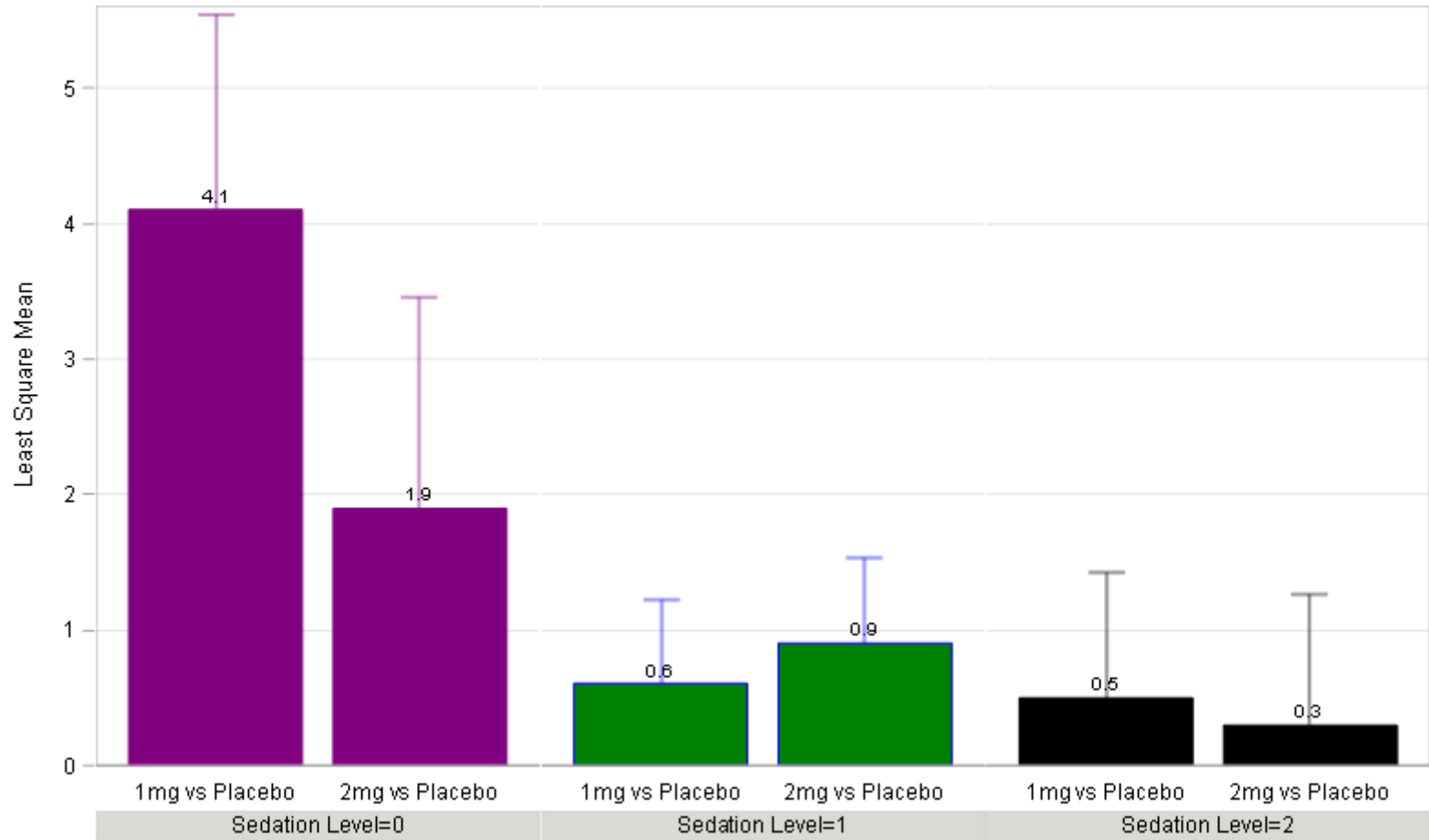
For the Key Secondary Outcome (Marder Factor) there is strong separation for the 2 mg arm, but the 1 mg arm trails off toward the end of the study

*Figure: Least Square Mean (+/-SE) of PANSS Score over Time by Treatment Group (Patients with 1 Antipsychotics and had  $\geq 3$  points decrease from screening to baseline and Without Any Nondetectable PK During study)  
Clean Population*



# Sedating Antipsychotics blunt the Response of Encenicline to the MCCB\*

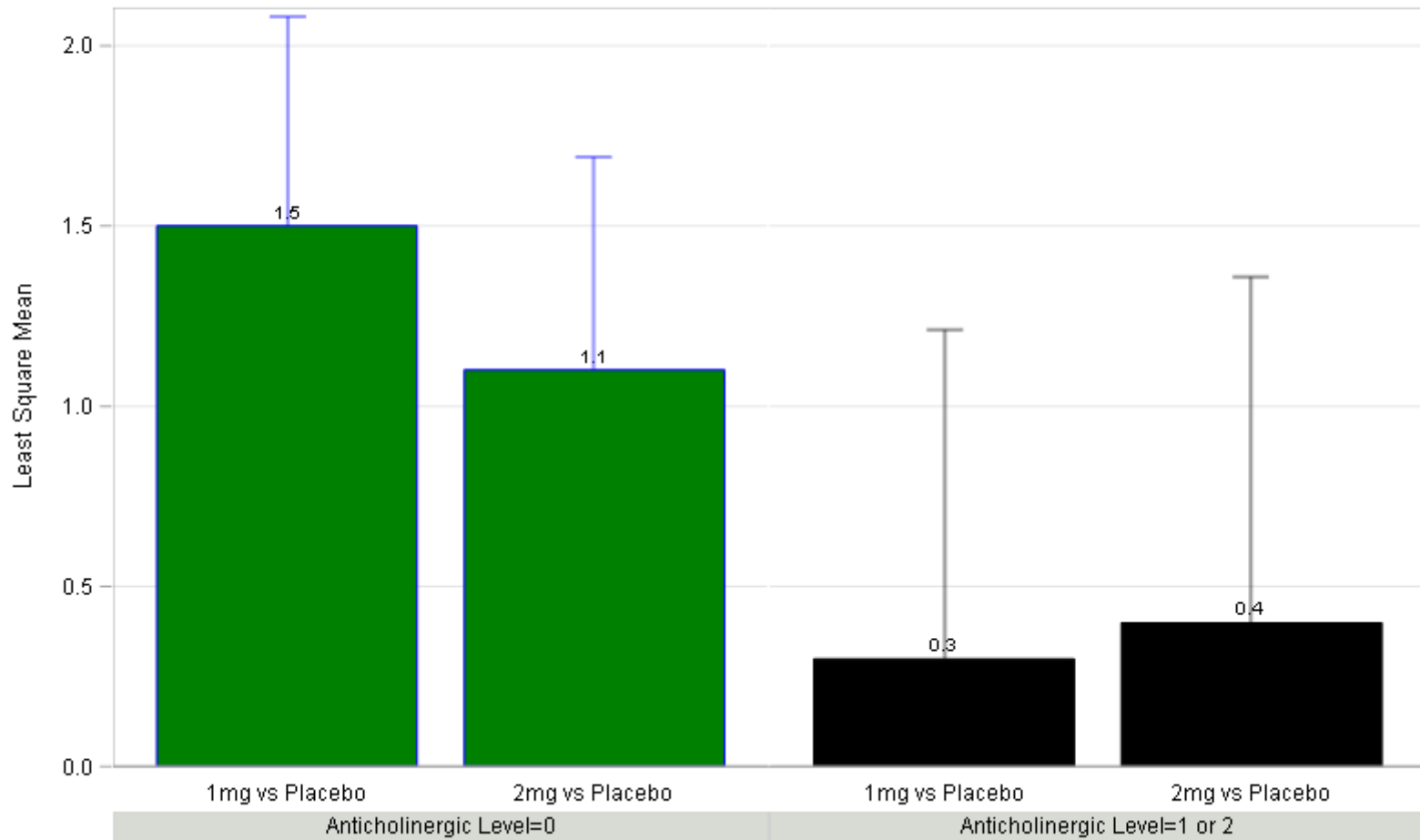
Difference of Least Square Mean of Change from Baseline in Neurocognitive MCCB T-Score at Week 26 Compared to Placebo



Analysis done on the "Cleanest" population  
Sedation of antipsychotics determined by staff using Stahl's ranking

# Antipsychotics with stronger ACh effects blunt the Response of Encenicline to the MCCB\*

Difference of Least Square Mean of Change from Baseline in Neurocognitive MCCB T-Score at Week 26 Compared to Placebo



Analysis done on the "Cleanest" population  
ACh effect of antipsychotics determined by staff using Stahl's ranking

# Summary of post-hoc analysis findings (1)

Significant separation can be seen in the pooled (n=1515) data when removing:

- Subjects on second antipsychotic (n=142) & Subjects who were rarely compliant [define with no plasma exposure] (n=272) or Subjects showing high variability from screening to baseline (n=125, 175, 240)
- Doing all three above exclusions results in a data set shows the greatest separation (and the placebo arm behaves more as expected)

Substantial difference in pooled data (n=1,515) between patients enrolled early in the trial and later

- Partly confirmed with Baseline differences seen between the 1<sup>st</sup> time cohort and the rest of the study

Some regional variation

- Europe has higher response to drug and a quite robust response to placebo
  - Russia & the Ukraine were specifically looked at, but do not explain the overall differences seen
- North America has lower drug response and much lower placebo/learning response
- No specific country created havoc

No one medication or combination of medications appears to be “toxic” or helpful

- Though differences, nothing substantial seen overall in the study that seemed to sway the data
- Though differences, nothing substantial seen in the subgroup that was stable on 2 antipsychotics

When Antipsychotics were looked at regarding how sedating or anticholinergic they were, a signal emerged that the more sedating or anticholinergic they were, they less likely they were to show a separation from placebo

Outlier values were apparent when looking at scatterplots; much of our interest has focused on the pre-baseline changes (e.g. an improvement of 24 points on the SCoRS)

- When cutoffs were established for the top 10% of pre-baseline changes for each outcome measure, eliminating those subjects from the analysis decreased variance and enhanced the separation of drug over placebo for all outcomes in the combined data sets, equivalent to eliminating those subjects who rarely took the drug

# Thoughts on design of a potential follow-on study

- 1) Control for high-noise sub-groups
  - Exclude patients stabilized on two antipsychotics from enrollment criteria
  - Exclude patients without sufficient drug exposure
  - Exclude patients with high screening-to-run-in variability
- 2) Use 25-35 sites in the US only to reduce variability
  - Sites scrubbed for outlier sites in the prior trials
  - Do not appear to need to avoid urban sites
- 3) Return to 3 month duration for the trial (same as P2)
  - 3 months is sufficient for FDA, it was EMA that wanted 6 months
  - Shorter duration may help with dropouts, etc
- 4) Have fewer outcome measurements during the trial
  - The multiple testing used in P3 did not decrease the “learning” effect
  - Still do a pre-baseline, only 1 or 2 post randomization
- 5) Though one can consider changes to endpoints, endpoint blinding may be sufficient
  - One would still need to pre-specify in the SAP, but do not need to tell the sites
  - Doing some of the MCCB, rather than all of it may be attractive depending on analyses
  - The Marder factor (Negative symptoms) looks fairly good and does not require a co-primary
- 6) Procedural changes to reduce noise and reduce placebo/learning response
  - Less frequent assessment with less extensive oversight (than recent P3 studies)
  - Maintain a run-in design which can also be needed to eliminate those not taking the drug &/or those with high pre-baseline changes (FDA opined that this would likely be acceptable done prior to randomization)

## Other opportunities to consider in evaluating potential paths forward

Genetic markers for placebo response: several groups working on this and several claims, but waiting for confirmation

- If can identify a biomarker, then exclude such subjects from a trial

New technologies for compliance (e.g. AI Cure, Proteus, etc)

- Some of these are now being used in studies and again could help in identifying noncompliant subjects

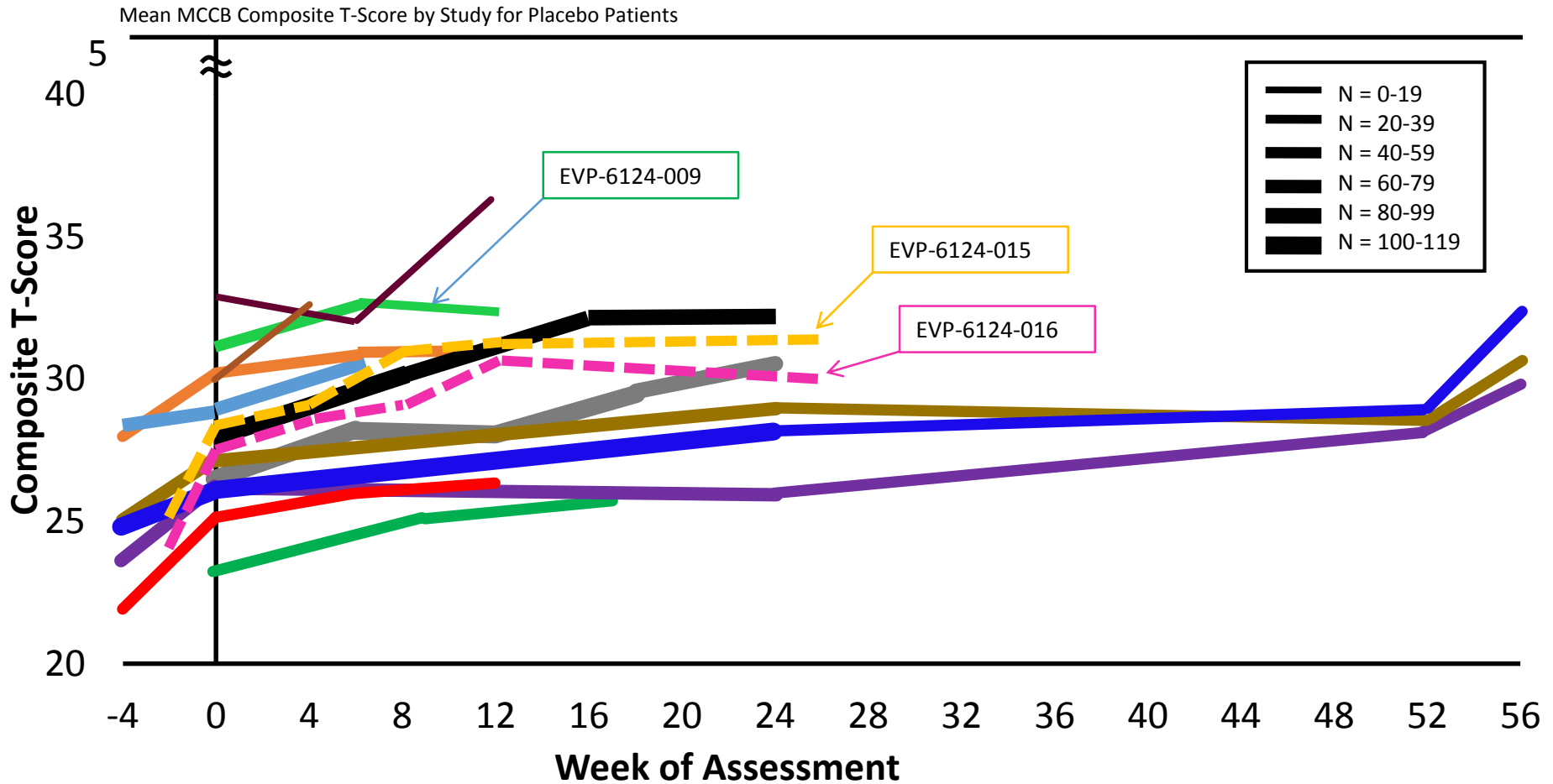
New technologies for large screening of Mismatch negativity, etc.

- This appears to be a reliable BM for Schizophrenia and could be used as an inclusion criteria (if logistics and price prove manageable)

**BACKUP**

# Historical placebo arm results in CIS studies using the MCCB

Relative to the 009 study (P2), the P3 studies' placebo arms had lower baselines and larger improvements over time



# Situation summary

Encenicline (PHX-6124) is a partial  $\alpha 7$  agonist in developed by FORUM for treatment of cognitive impairment in multiple indications

Drug advanced into four large Phase 3 registration trials based on positive Phase 2b data

- Alzheimer's disease: 877 patients x 2 mirrored global trials (placed on Clinical Hold by FDA midway through due to a potential Safety Signal)
- Schizophrenia (CIS): 1516 patients x 2 mirrored global studies

Lower GI SAEs led the AD trials to be put on clinical hold and later stopped by FORUM

- Related to 5-HT3 inhibition in the context of age, Alzheimer's dementia & diverticular disease
- Little expected impact on younger populations such as schizophrenia where no signal observed

Schizophrenia trials did not meet primary efficacy endpoints

- Favorable safety profile
- Very large Baseline (and pre-baseline) to Endpoint changes in all groups resulting in a particularly robust placebo/learning effect on the co-primary outcome measures (MCCB and SCoRS)

FORUM was wound down quickly in Q2 of 2016 based on unfavorable data and shift in the shareholder's portfolio strategy

Subsequent post-hoc analysis of the schizophrenia data suggests a drug effect and possible path for continued clinical development

- Elimination of subgroups with unusual responses results in a large data set that shows separation of the drug arms from placebo in all the outcome measures
- African American subgroup shows separation of drug from placebo in the outcome measures
- Potentially informs a new study design to confirm post hoc analyses of completed Phase 3 studies

## Topline Safety Summary- essentially well tolerated

	015 Study	016 Study
Overall TEAEs	Evenly distributed across treatment groups 1mg (50.0%), 2mg (53.8%), Plc (52.4%)	Plc (58.3%) > 1mg (48.1%), 2mg (49.2%)
Incidence of TR AEs	1mg (26.6 %) > 2mg (18.5%), Plc (19.4%)	Similar across treatment groups 1mg (18.6%), 2mg (22.0% ), Plc (21.3%)
SAEs	Total: 16 subjects (2.1%) 1mg (2.4%), 2mg (1.2% ), Plc (2.8%) Most frequent: <ul style="list-style-type: none"> <li>Psychiatric decompensation</li> </ul>	Total: 25 subjects (3.3%) 1mg (4.3%), 2mg (2.8% ), Plc (2.8%) Most frequent: <ul style="list-style-type: none"> <li>Psychiatric decompensation</li> </ul>
Discontinued due to AE	Total: 1mg (4.8%), 2mg (4.4%), Plc (5.9%) Most common reason: <ul style="list-style-type: none"> <li>Psychiatric decompensation (2.1%)</li> </ul>	Total: 1mg (5.0%), 2mg (3.5%), Plc (5.1%) Most common reason: <ul style="list-style-type: none"> <li>Psychiatric decompensation (2.1%)</li> </ul>
Most frequent AE	Headache (7.3%), constipation (4.4%), insomnia (4.1%) and psychiatric decompensation (2.8%)	Headache (5.9%), constipation (5.0%), psychiatric decompensation (3.7%), and insomnia (3.4%)
Deaths	Two deaths (1mg and 2mg); Both accidental overdoses considered unrelated	No deaths

## EVP-6124-016 serious adverse events

Few overall and roughly equally distributed

Twenty-five subjects (3.3%) overall experienced a serious adverse event: 11 subjects (4.3%) in the 1 mg group, 7 (2.8%) subjects in the 2 mg group, and 7 (2.8%) in the placebo group.

The most frequently occurring SAE was psychiatric decompensation and was essentially evenly distributed across treatment groups: 1mg (3.9%), 2mg and placebo (3.5%)

Total of 5 out of 29 SAEs were assessed as treatment related

- 1mg
  - ECG signs of myocardial ischemia
  - Psychiatric decompensation
- 2mg
  - Myocardial infarction
  - Psychiatric decompensation (2)
- Placebo
  - No Serious TR AEs

SOC	1 mg	2 mg	Placebo
Any SAE	11	7	7
Psychiatric	8	6	6
Cardiac	1	1	0
Injury	0	0	1
Renal	1	0	0
Infections	1	0	0

# EVP-6124-015 serious adverse events

Few overall and roughly equally distributed

Sixteen subjects (2.1%) overall experienced a serious adverse event: 6 subjects (2.4%) in the 1 mg group, 3 (1.2%) subjects in the 2 mg group, and 7 (2.8%) in the placebo group.

The most frequently occurring SAE was psychiatric decompensation and was essentially evenly distributed

Total of 3 out of 25 SAES were assessed as treatment related

- 1mg
  - Psychiatric decompensation
- 2mg
  - No Serious TR AEs
- Placebo
  - Psychiatric decompensation

SOC	1mg	2mg	placebo
Any SAE	10	7	8
Injury, poisoning, procedural complications	4	2	
Psychiatric	6	5	5
Gastrointestinal			1
Infections and infestations			1
Musculoskeletal/ Connective Tissue			1

## Safety summary - areas of special interest

Essentially no unexpected findings

	015 Study	016 Study
Cardiac	<p>2mg (0.4%) &lt; 1mg, plc (2.0%)                      Overall incidence is 1.5 %                      No SAEs</p>	<p>Similar across treatment groups:                      • 1mg, 2mg, Plc (1.2%)                      Two SAEs:                      • ECG signs of myocardial ischemia                      • Myocardial Infarction</p>
GI	<p>Varied across treatment groups:                      • 1mg (13.9%), 2mg (14.9%) and Plc (8.7%)                      Overall incidence 12.5 %                      Constipation suggests dose dependence:                      • 1mg (5.6%), 2mg (7.3%), Plc (0.4%)                      One SAE: Abdominal pain / placebo arm                      Seven non-serious AEs across treatment groups:                      Rectal hemorrhage, Hemorrhoidal hemorrhage and Hematochezia</p>	<p>Similar across treatment groups:                      • 1mg, 2mg, Plc (12.9%)                      Overall incidence 14.8 %                      Constipation suggests dose dependence:                      • 1mg (5.0%), 2mg (6.7%), Plc (3.1%)                      No SAEs</p>
Psych	<p>2mg (8.1%) &lt; 1mg (13.5%), plc (13.4%)                      Most common:                      Insomnia                      • 1mg (4.8%), 2mg (3.6%), Plc (4.0%)                      Anxiety                      • 1mg (3.2%), 2mg (2.0%), Plc (4.7%)                      Psychiatric decompensation                      • 1mg (2.4.9%), 2mg (2.8%), Plc (3.2%)</p>	<p>Plc (14.6%) &gt; 1mg (12.0%), 2mg (11.0%)                      Most common:                      Psychiatric decompensation                      • 1mg (3.9%), 2mg (3.5%), Plc (3.5%)                      Insomnia                      • 1mg (3.1%), 2mg (2.8%), Plc (4.3%)                      Anxiety                      • 1mg (3.1%), 2mg (3.1%), Plc (2.0%)</p>

In the ITT analysis, correlations between the MCCB Domains looked close to what expert neuropsychologists anticipated

*Most correlations between .4 and .6, except for social cognition, which is what is typically seen in MCCB datasets*

EVP-6124-015 & 016

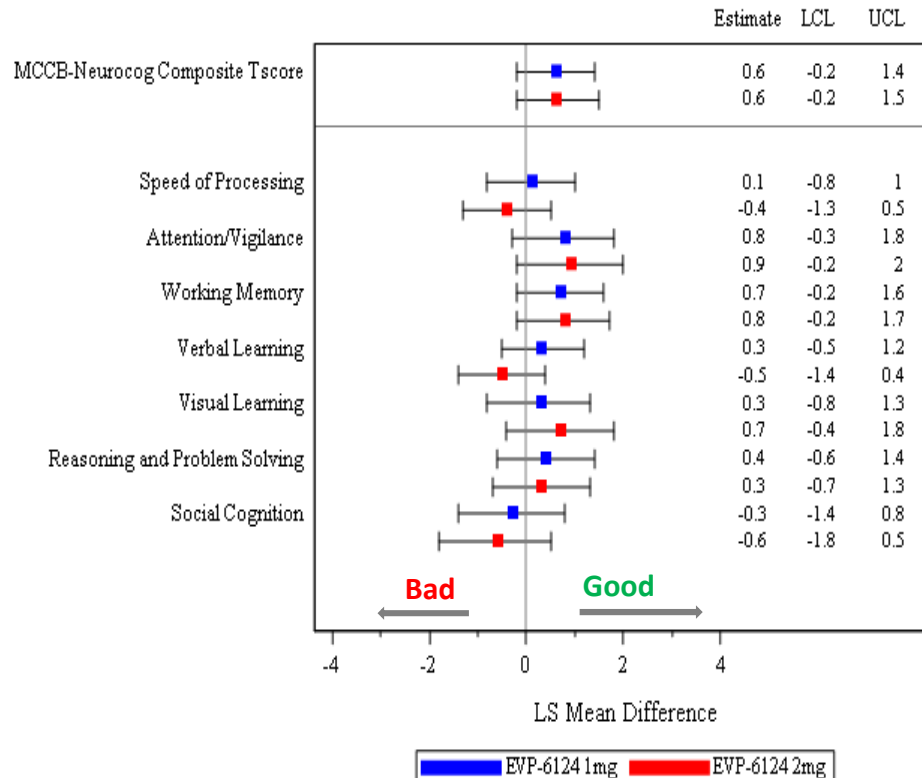
Correlation between MCCB Domains  
ITT Population

	Speed of Processing	Attention/Vigilance	Working Memory	Verbal Learning	Visual Learning	Reasoning and Problem Solving	Social Cognition
Speed of Processing	1.00	0.57	0.60	0.45	0.53	0.63	0.29
Attention/Vigilance	0.57	1.00	0.59	0.40	0.41	0.40	0.27
Working Memory	0.60	0.59	1.00	0.49	0.58	0.46	0.30
Verbal Learning	0.45	0.40	0.49	1.00	0.49	0.28	0.24
Visual Learning	0.53	0.41	0.58	0.49	1.00	0.41	0.22
Reasoning and Problem Solving	0.63	0.40	0.46	0.28	0.41	1.00	0.21
Social Cognition	0.29	0.27	0.30	0.24	0.22	0.21	1.00

In the original ITT analyses, there were not any Components that performed particularly well or poorly  
 Forest plots of MCCB: Week 26 and by Components- Combined EVP-6124-015 & 016

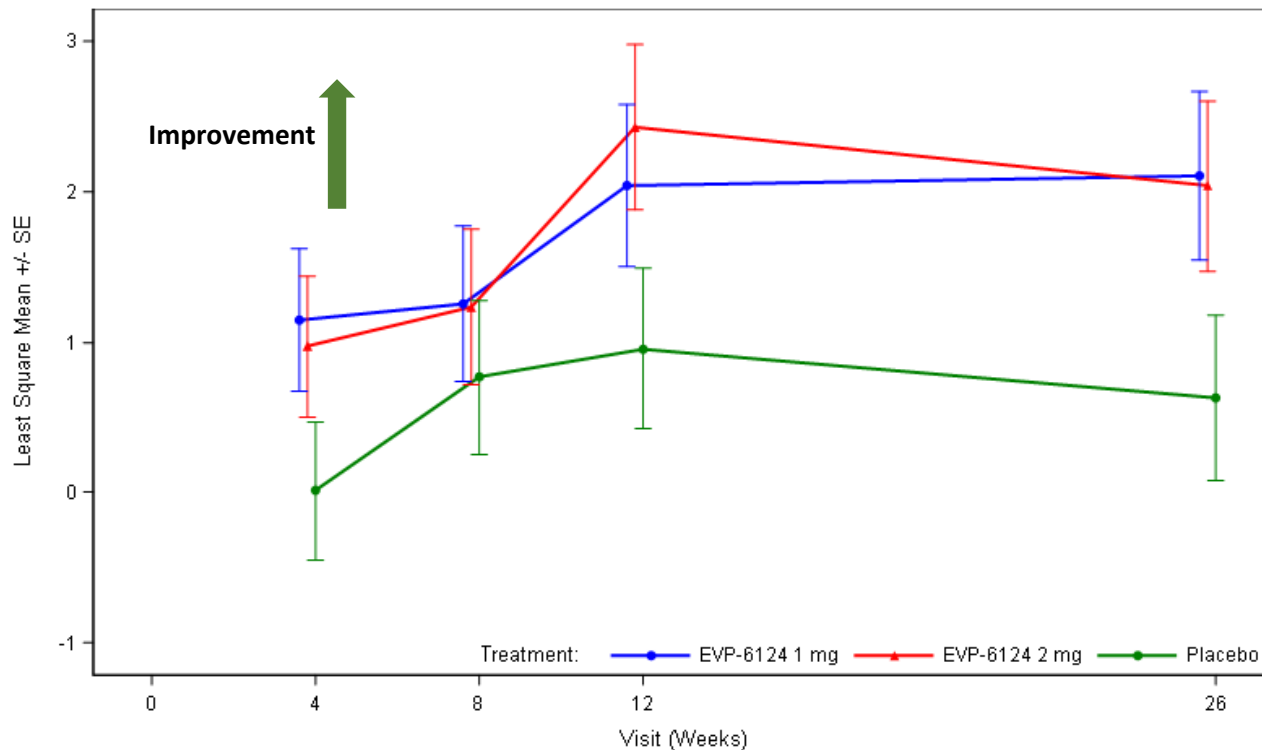
EVP-6124-015 & 016

MCCB Components - Treatment Difference and 95% CI at Week 26



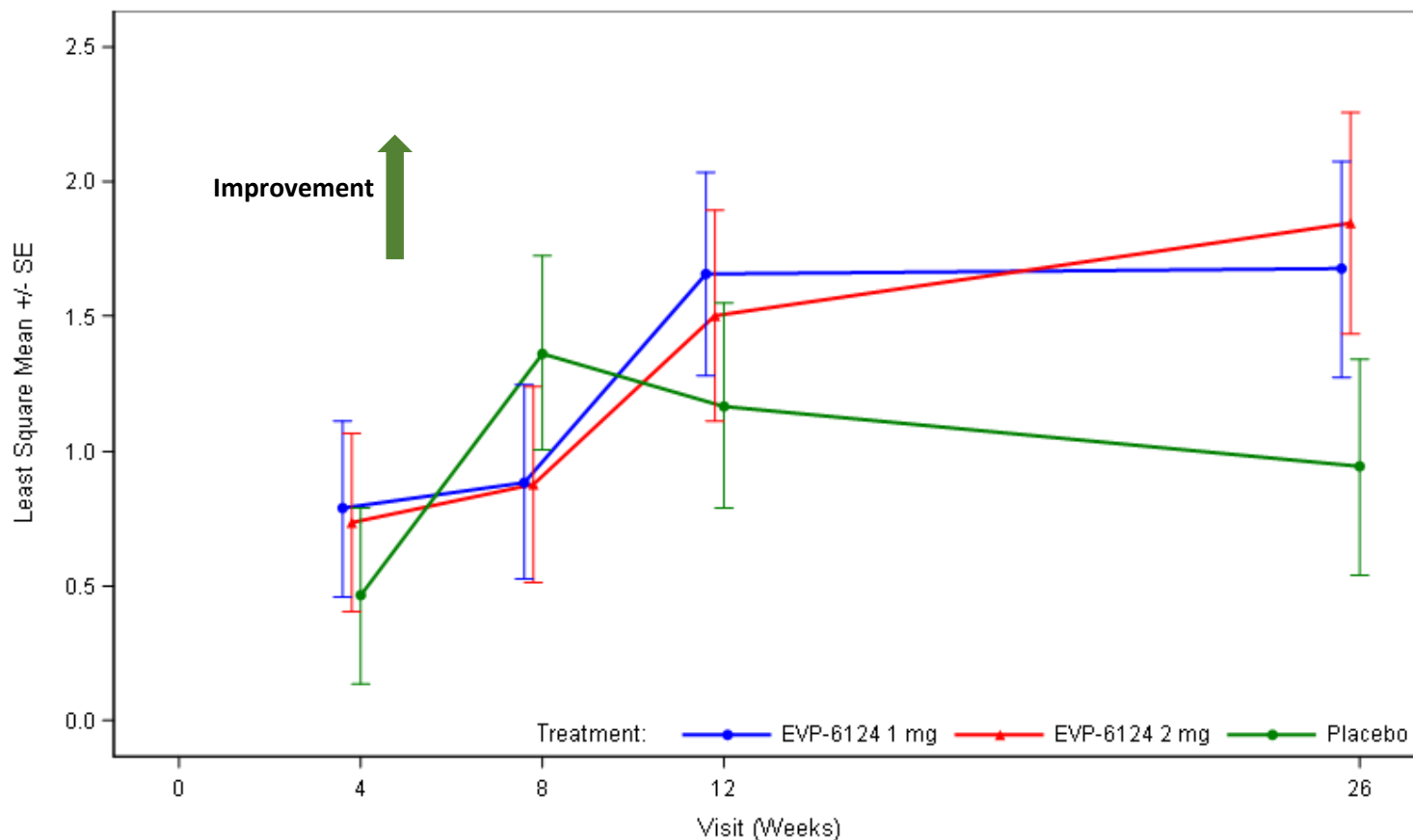
Analysis on a “cleaner” data set (here those on 2 antipsychotics were excluded) shows some possible advantage in one of the components

*Figure: Least Square Mean (+/-SE) of Attention/Vigilance Domain T-score over Time by Treatment Group (Patients with 1 Antipsychotics) modified ITT population*



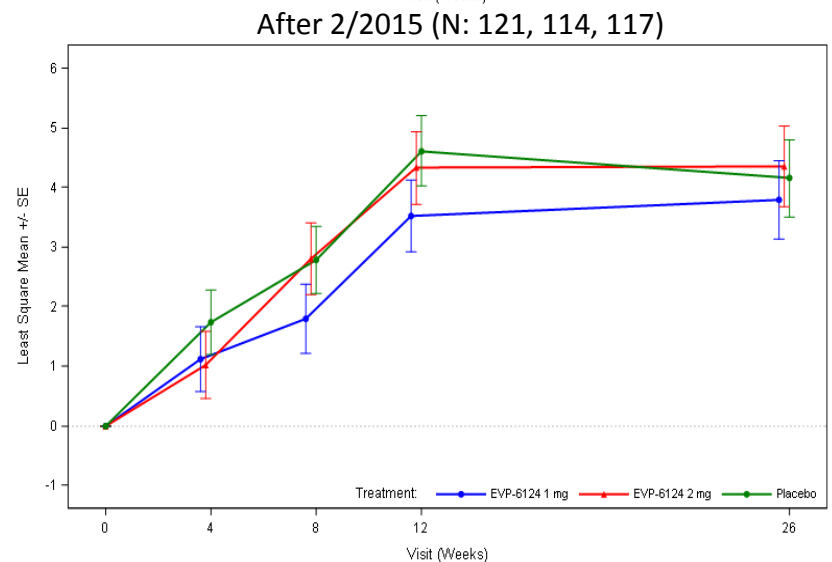
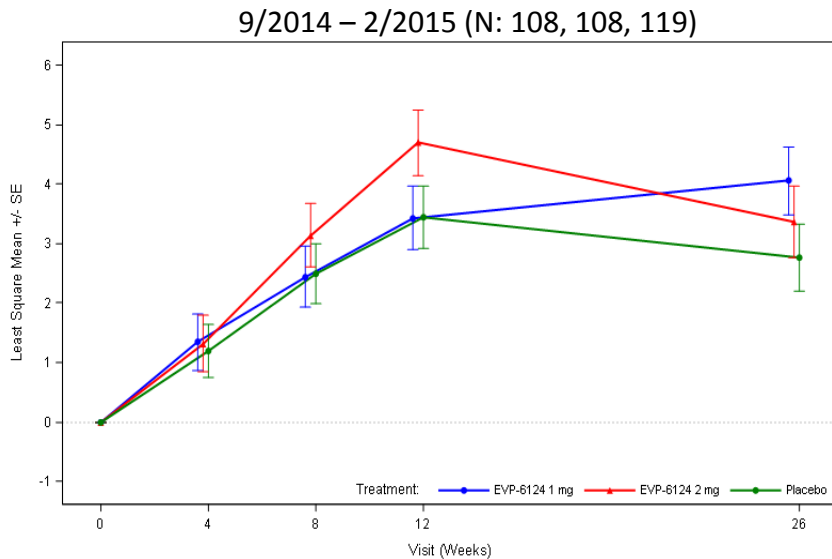
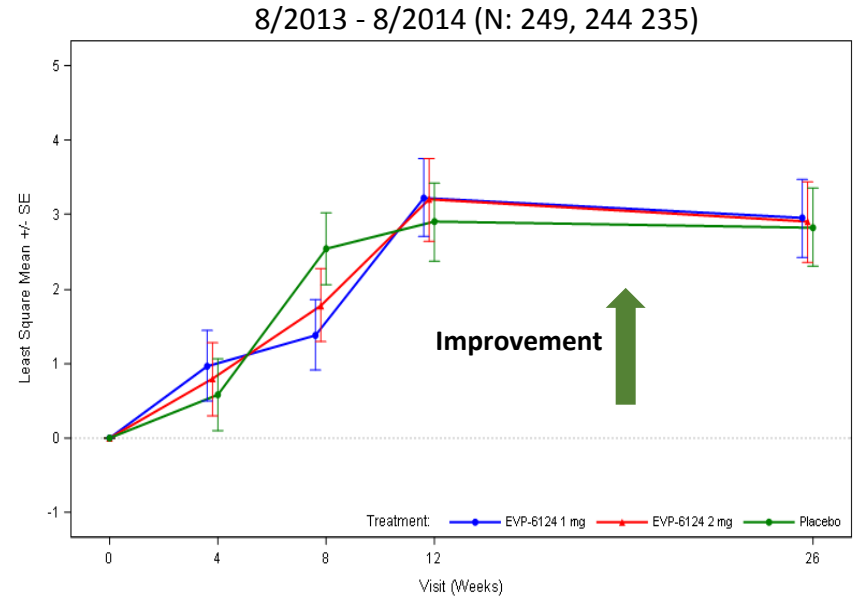
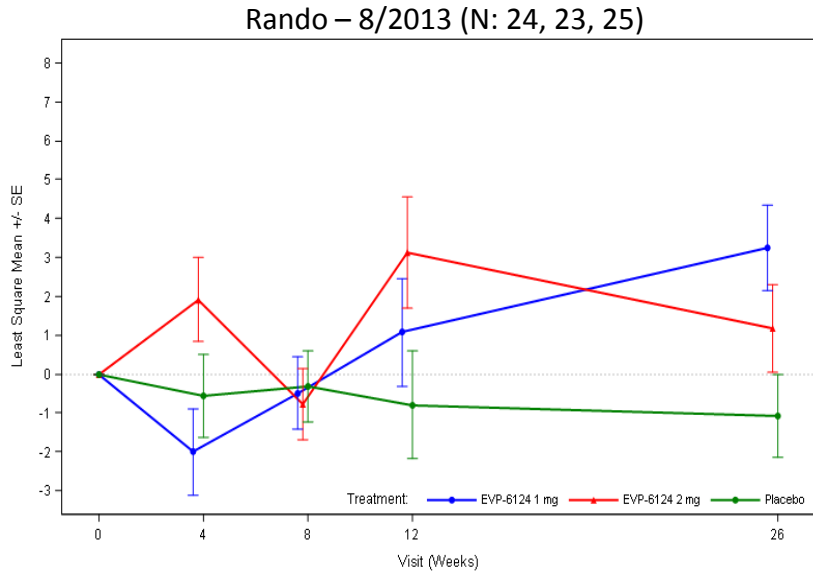
Analysis on the “cleanest” data set (here those on 2 Antipsychotics, who were noncompliant, or who had large pre-baseline improvements were excluded)

Figure: Least Square Mean (+/-SE) of Subdomain Neurocognitive MCCB T-Score (Attention/Vigilance) over Time by Treatment Group (Clean Patient Population)  
ITT Population



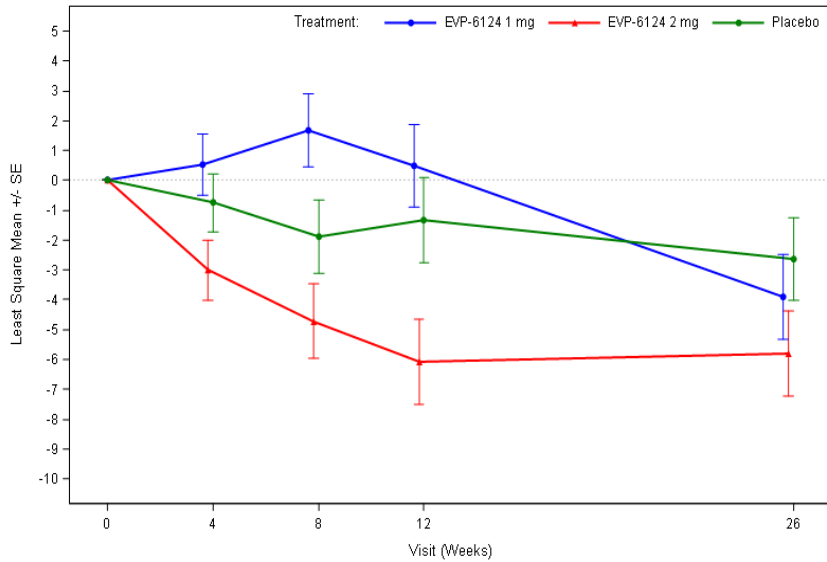
Notable differences, particularly in the placebo arm, in the time windows;  
both in amount and in relationship to the drug arms

Line plot of NCC score over time by treatment group combined studies (016/015)

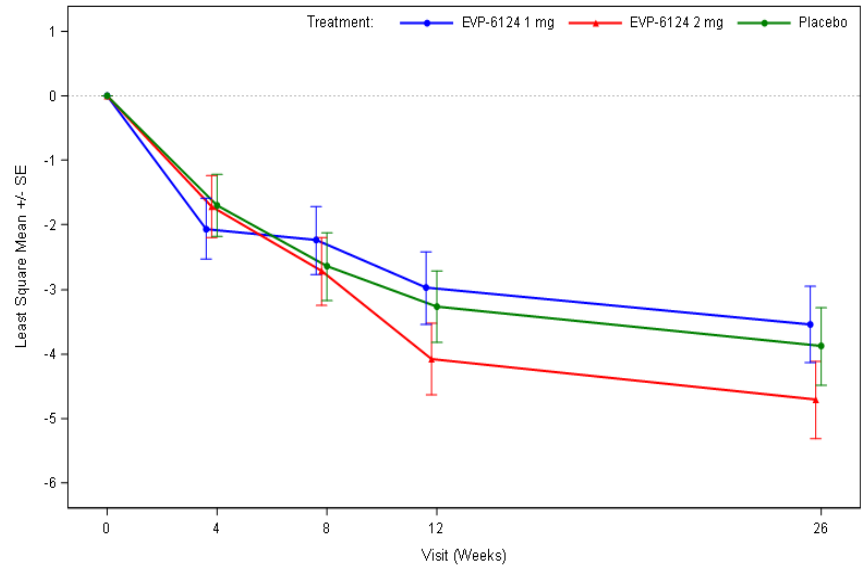


# Similar findings are also seen in the functional co-primary Line plot of SCoRS score over time by treatment group (016/015)

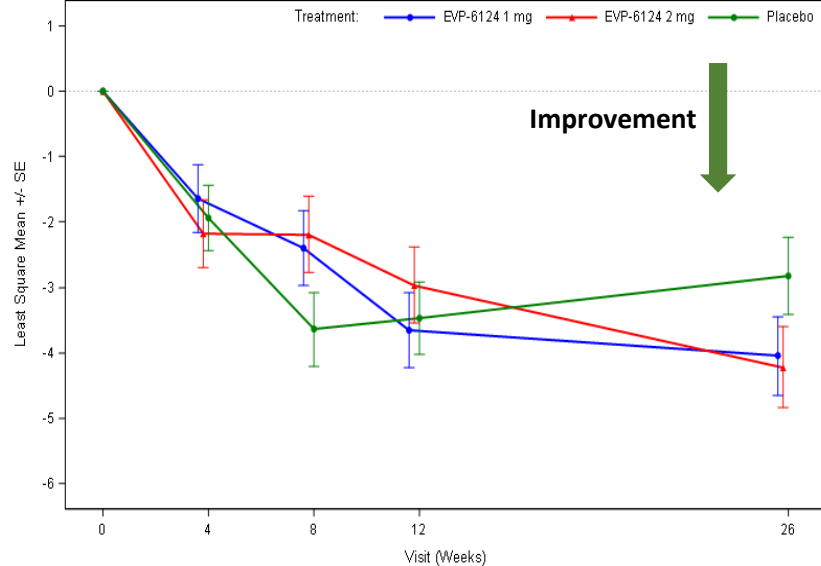
### Rando – 8/2013



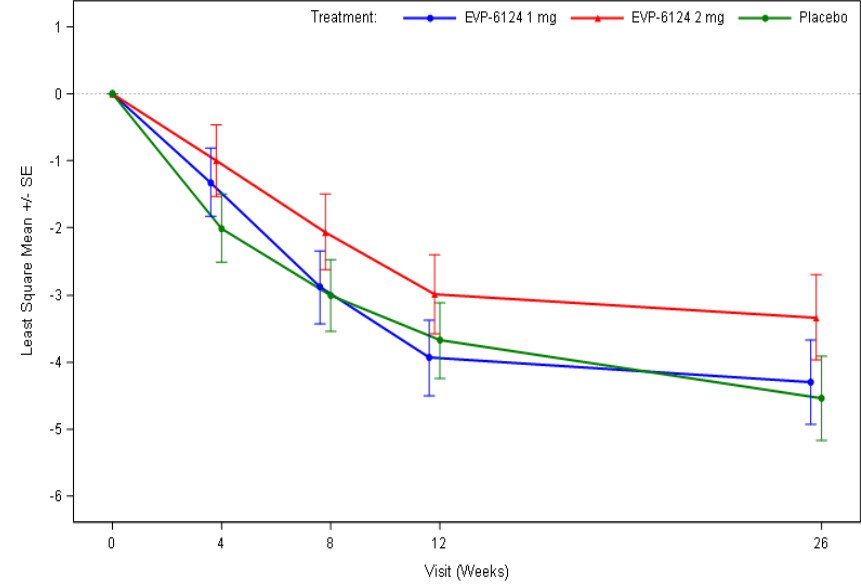
### 8/2013 – 8/2014



### 9/2014 – 2/2015



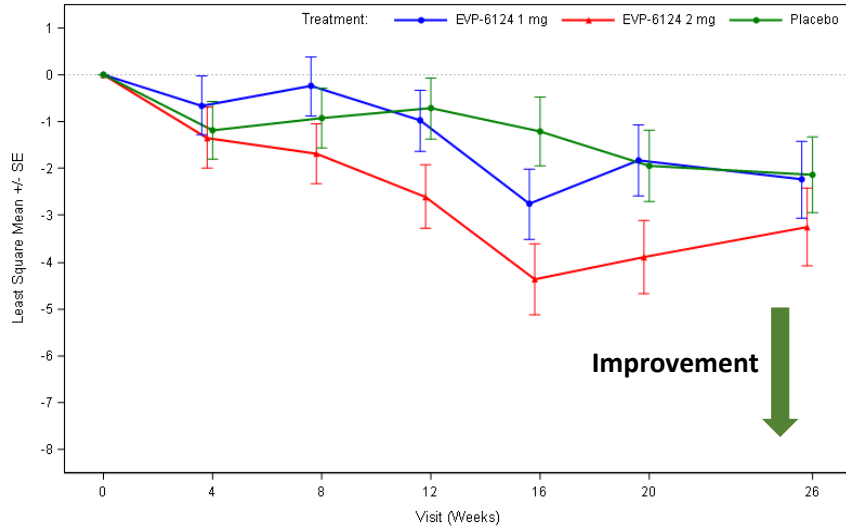
### After 2/2015



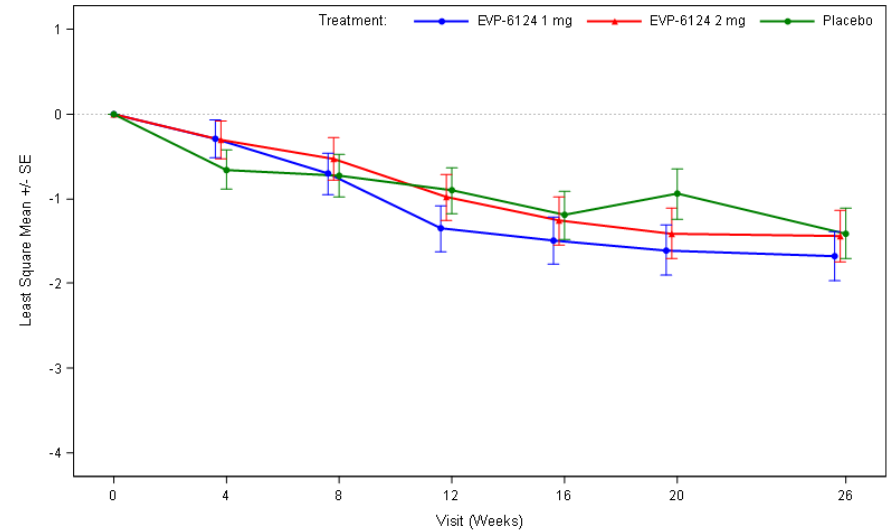
These findings are not seen in the key secondary outcome measure for negative symptoms (Marder Factor)

Line plot of PANSS Marder score over time by treatment group (016/015)

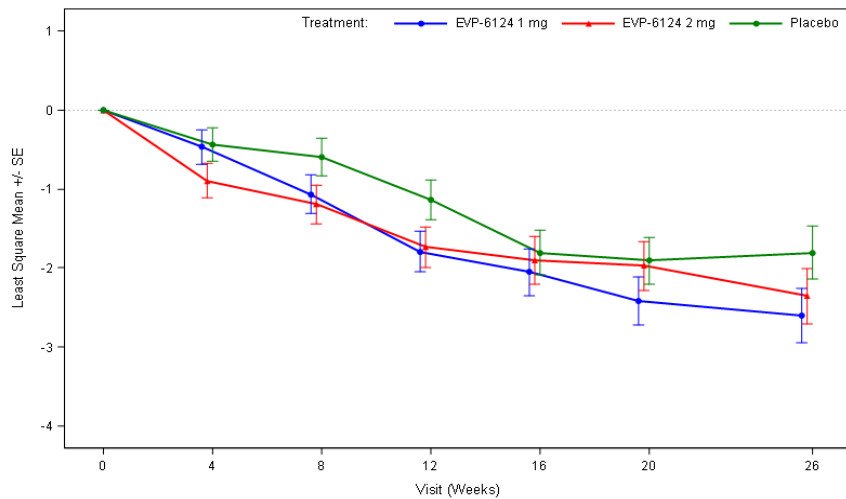
Rando – 8/2013



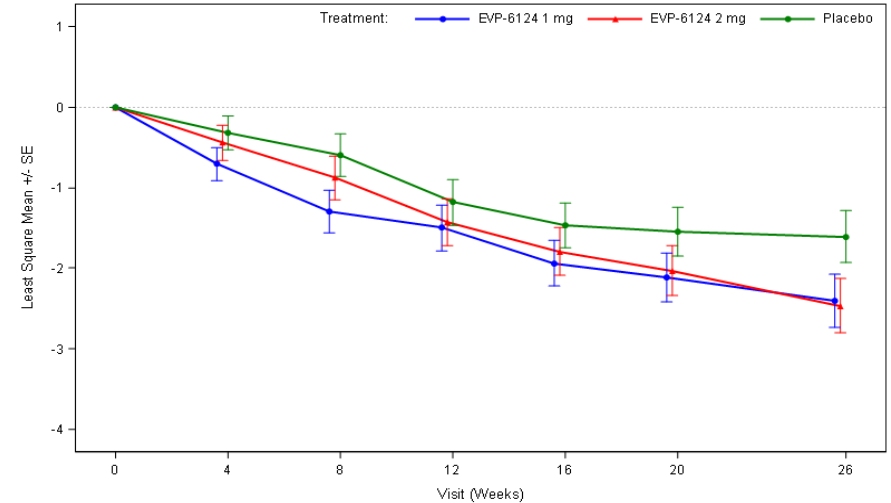
8/2013 – 8/2014



9/2014 – 2/2015

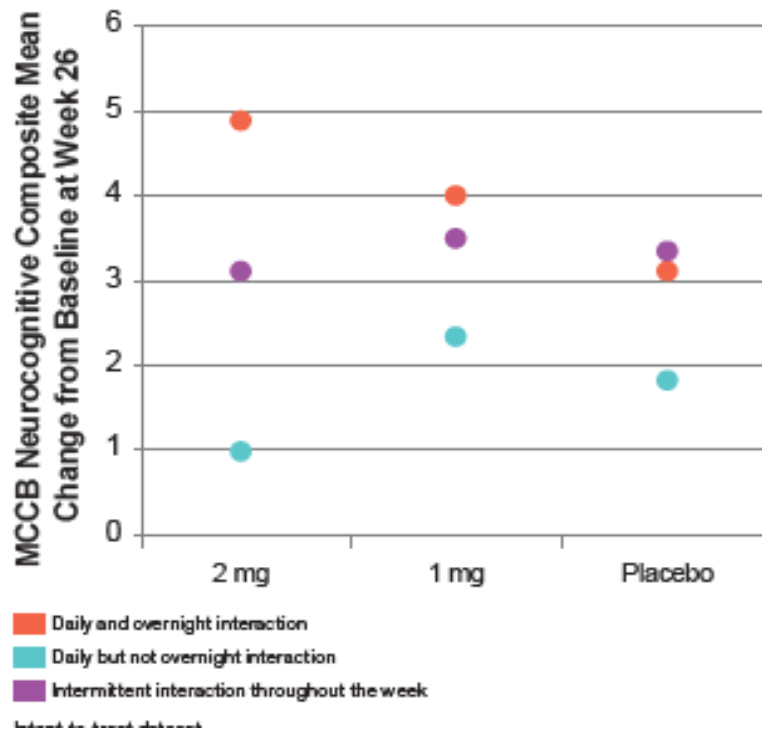


After 2/2015



# Impact of informant contact frequency on primary trial analysis in the ITT population (less is worse)

Figure 4: MCCB Change from Baseline by Dose Group and Informant Contact Frequency



## Summary of post-hoc analysis findings (2)

No one medication or combination of medications appears to be “toxic” or helpful

- Though differences, nothing substantial seen overall in the study that seemed to sway the data
- Though differences, nothing substantial seen in the subgroup that was stable on 2 antipsychotics

When Antipsychotics were looked at regarding how sedating or anticholinergic they were, a signal emerged that the more sedating or anticholinergic they were, they less likely they were to show a separation from placebo

Outlier values were apparent when looking at scatterplots; much of our interest has focused on the pre-baseline changes (e.g. an improvement of 24 points on the SCoRS)

- When cutoffs were established for the top 10% of pre-baseline changes for each outcome measure, eliminating those subjects from the analysis decreased variance and enhanced the separation of drug over placebo for all outcomes in the combined data sets, equivalent to eliminating those subjects who rarely took the drug

## “Clean” study population evaluated

This data set shows the separation between the drug arms and the placebo arm, particularly for the co-primary outcomes

- Which drug arm is stronger varies between the outcome measures

The placebo arm here behaves like a more typical placebo arm does in trials

- Less extensive Baseline to endpoint gains
- A somewhat flatter curve over time, particularly at later time points

Though there is separation, the amount of separation is modest

- Combining the exclusion of subjects who were noncompliant and those who had large pre-baseline changes was clearly not additive; possibly the 2 groups may have had considerable overlap
- Not sure the difference can be called “clinically significant”, though directionally this analysis is encouraging that additional trials, if conducted carefully, could be successful
- Not clear why there is less “enhancement” with the Negative symptom scale (Marder factor), though this was the one scale that showed trends even with the ITT population (and thus less enhancement possible); it is possible that it was emphasized less in the clinic as it was not a co-primary