Author's response to reviews

Title: Effects of corticosteroid therapy on language, behavior and the Frequency Modulated Auditory Evoked Response (FMAER) in regressive autism: A clinical case control study

Authors:

Frank H Duffy MD (fhd@sover.net)
Aditi Shankardass PhD (aditi.shankardass@childrens.harvard.edu)
Gloria B Mcanulty PhD (gloria.mcanulty@childrens.harvard.edu)
Yaman Z Eksioglu MD (eksiogly@upstate.edu)
David Coulter MD (david.coulter@childrens.harvard.edu)
Alexander Rotenberg MD (alexander.rotenberg@childrens.harvard.edu)
Heidelise Als PhD (heidelise.als@childrens.harvard.edu)

Version: 3
Date: 29 January 2014

Author's response to reviews: see over
Reviewer 1  Todd Richards

This manuscript has very important implications for the treatment of autism. The major compulsory revision that this reviewer suggests is the statistical method used to analyze treatment effects which could change as follows: Test the group by time interaction effect in a mixed ANOVA or do an ANCOVA with condition as the independent variable and the pre-measure as the covariate and post measure as the dependent variable.

The EEG and language measures appear to be appropriate for this study. The paired t-tests and group comparisons are helpful in understanding the data but not enough towards the full statistical test that needs to be done.

Reply: Indeed a variant of ANOVA would be the way to go if groups were comparable. We had considered ANOVA carefully before embarking on our present course. However, given the exploratory nature of this study it would seem less warranted at this pilot level when a simple paired-t test suffices and is easier for readers to grasp as pointed out.

We thank the reviewer for his review and comments.

Reviewer 2  Deborah Fein

This is a paper of great potential importance, which if published in its present form, would doubtless attract a good deal of attention by physicians and parents. It is, therefore, particularly important that conclusions be appropriately cautious and tied tightly to the data. The paper needs substantial revision to meet that criterion.

Reply: We agree with the reviewer and have modified the manuscript accordingly.

Major Compulsory Revisions:

The claims rest on 4 measures: changes in FMAER, changes in EEG, changes in language, and changes in behavior. I am not qualified to comment on the technical aspects of the FMAER; I would only mention two points:

1. First, I believe the only reference to the FMAER method is to a recent paper from the first author’s laboratory; is this a method that has been replicated or accepted by others in the field? If not, perhaps this needs to be mentioned as a limitation of the study, namely, that the study of the meaning of FMAER is relatively new and needs replication by others. Or cite others’ work to this effect,
if I’m wrong about that.

**Reply:** Our referenced FMAER paper re-introduces and carefully explains the FM steady state auditory evoked response, which was developed and published by Green and Stefanatos in 1979 and 1980. Our FMAER paper reviews the past use of FM stimulation, for example, in ‘specific language impairment’ (1995), ‘Landau-Kleffner syndrome’ (1993), and ‘children’s phonological and reading skill’ (1999). The current Introduction section has been modified to include the pertinent references.

2. Second, in Tables 9 and 10, regressions are reported in which for the STAR group, two leads show significant correlation with language, one with receptive language (C3) and one (T7) with expressive, and for the untreated group, one lead (P4) shows correlation with both language measures. These are among 14 leads. It would be most parsimonious to consider that these are chance correlations. The authors argue that the reason for only finding one significant correlation is that all the leads are highly intercorrelated; this is easily checked, by looking at the correlations among leads, and between each lead and the change in language score, separately. If the regression result is because the leads are all intercorrelated, one should see significant correlations between other leads and language, when looked at as simple bivariate correlations. Also, I would appreciate the authors considering in their discussion whether the results found make anatomical sense; would one expect that a left central lead would have the strongest correlation with receptive language, or the mid-temporal lead the strongest correlation with expressive language? Or the right hemisphere lead the strongest correlation with language in the untreated group? I am not saying they don’t make sense, but I would have expected some discussion of that.

**Reply:** We have two responses to these important key questions.

(1) When evaluating EEG (aside from epileptic spikes) and also when evaluating spectral analysis of EEG it is reasonable to assume that a finding centered about an electrode represents cortical activity from brain beneath that electrode, e.g., T7 reflects left mid-temporal lobe cortical activity. This is not the case for most evoked response (ER) data that originate as a point source or point sources within brain. In our FMAER paper we demonstrate that the FMAER arises in normal subjects and in non-language impaired patients from the bilateral superior temporal gyri (STG). The response may be modeled by a dipole source (think a double sided flashlight projecting red (+) from one end and blue (-) from the other end). The source (flashlight) is physically oriented in the STGs so that the positive (+) end points medially and anteriorly from the STG and results in a wide anterior, scalp response maximal in the ipsilateral frontal scalp. In contrast the negative (-) end points posteriorly, inferiorly, and laterally so that there is a posterior-inferior negative response typically maximal in the ipsilateral posterior temporal region. There is typically no obvious response in the scalp electrode
lateral to and overlying the STGs as this scalp point is equidistant from the + and – ends of dipole source. Thus, the FMAER is widely projected to the scalp with an ipsilateral frontal to inferior temporal distribution with a poor representation on scalp just over the STG despite the fact that it actually arises from the STG. In ‘normals’ the FMAER arises bilaterally with overlapping projection to the medial frontal regions. One would predict, therefore, that all ipsilateral electrodes would show signals that are highly correlated since they all arise from the same single dipole source and differ from one another only as regards amplitude of scalp response and therefore also to signal-to-noise ratio (i.e., Evoked Response (ER) to background residual EEG amplitude). In deciding how to take measures for quantification we were guided by the fact that in pathology the FMAER dipole source may not only be low amplitude and/or distorted but also pathologically physically dislocated to another cortical site —such cases are graphically illustrated in our FMAER paper. As pathology may induce dipole sources that do not always arise in the STG and may not have ‘normal’ orientation (e.g., point to unexpected scalp regions), we elected to base analyses on 7 electrodes per hemisphere to maximize response detection. As a result of the nature of scalp recorded ER data findings, one cannot make clear clinical correlations between electrode of maximal statistical difference and the function of underlying cortex. One can, however, view differences in left and right sided findings as hemispherically specific and draw left vs. right hemispheric correlations. We have extended the methods section to clarify these complex points.

(2) Also, please, see Table 9 and 10 and associated text in Results. This was a step-wise regression. In BMDP 2R, in the first step the correlation between all 14 neurophysiological variables and the single language variable is calculated. The strongest neurophysiological variable (largest F statistic) is taken for the ‘First Step’. Then BMDP 2R removes the degree to which information in the chosen neurophysiological variable is contained within all 13 remaining neurophysiological variables. After removing the first variable’s effects from all remaining 13 variables new correlations are formed between these 13 and the language variables. At that point BMDP 2R chooses a member of the 13 remaining variables; however, the F test did not reach selection criterion for any one of the remaining 13 variables. In step-wise discriminant and step-wise regression analysis this so-called ‘suppression of variables’ by previously chosen variables means that remaining variables contain no new information compared to those (in this case the one variable) chosen earlier. In this study it means that once the ‘best’ correlating variable was chosen no additionally useful information was left in the remaining 13 variables. This results from the fact that all 14 neurophysiological variables are highly correlated — as anticipated.

The paper has been modified to reflect the above important issues.

3. The second result concerns change in EEG, and the authors appropriately do not claim that such changes are due to the treatment. It was also a very good aspect of their method that the EEG assessor was blind to group membership.
However, in Table 5, the Fisher’s exact test for how many participants in each group had EEG’s that were worse, unchanged, or better was p=.03. I ran a Fisher’s exact test, as well as a chi-square with and without Yates correction, for the data in this table, and I got nonsignificant results for each. If the Fisher’s exact test is based on running ‘changed’ vs. ‘unchanged’ cells in a 2 x 2, I still get nonsignificant Fisher’s values, for either one-tailed or two-tailed. Finally, the text states that ‘there was a significant difference (p = 0.033) in the type of EEG abnormalities’ – there were no data about actual type of EEG abnormalities presented.

Reply: Table 5 is in error. The ‘3’ and ‘8’ were swapped in error. When corrected the groups now add up to the correct group counts. The 2 x 3 Fisher exact test is not significant. Thank you for identifying this error. Additionally, the EEG abnormalities as tabulated were all of the ‘benign sort’ that are commonly considered inconsequential in normal subjects. The number of events was small; there were no group differences and neither group showed any epileptiform activity. Thus, adding ‘counts’ appears unnecessary. Note that all subjects had had a ‘normal’ clinical EEG including sleep EEG as an entry criterion to this study so as to rule out any consideration of Landau Kleffner Syndrome (LKS). The text has been modified accordingly.

The issue of R-ASD as reflecting a covert form of LKS remains open and we are planning to explore additional R-ASD patients with combined MEG and 128 channel EEG in order to investigate whether spike discharges might be identified by MEG or denser electrode arrays, while not apparent by standard EEG. This is now mentioned in the Conclusion.

4. The third metric is that of improved language, and here I do have some major concerns. This measure of language is problematic on several levels: first, it has never been assessed for reliability or validity, or published.

Reply: The reviewer is correct. Steroid therapy discontinuation (see text) rested upon information about language change or plateau/stability; since there appeared to be no ‘published’ test that covers the range of changes expected in such subjects as based on past clinical experience, and no test that may be given in less than 10 minutes by a neurologist in the context of a clinical visit, the CLSQ was developed. Note that the CLSQ was developed before not after the reported study was undertaken. It was scored at the end of each therapeutic trial without knowledge of FMAER results and of course without knowledge of the overall group analysis. The NSA group’s language assessment was still less rigorous than the treated group’s language assessment. Nevertheless the clinical ratings appear to offer the opportunity for preliminary investigation of possible language effects. We have added text in the revised manuscript that addresses the legitimate concerns raised by the reviewer.

5. Second, some of the distinctions seem difficult to make: how is the parent to
know the difference b/t meaningful 1-2 word phrases vs. short meaningful phrases? The scale doesn't distinguish b/t combinations that may be frozen or routines (bye bye grampa) vs. real indications of language growth. What does 'nearly normal' or 'appears normal' in terms of expressive language mean? Did the clinicians or parents take the child's cognitive level into account? It's not entirely clear whether the clinician or parent judgments were the primary basis for the rating.

Reply: The reviewer’s points are well taken. The final scoring was decided by the clinician on the basis of one or both parents’ responses to a set of standardized questions coupled with further exploration of adequate supporting evidence offered by the parents in elaboration of their responses. Experienced Child Neurologists typically are quite astute regarding the differentiation of scripted, imitated, visually prompted, etc. language performance. The parent(s) were challenged clinically sensitively yet firmly to provide full evidence to support their contention(s). There were certainly occasional circumstances where, on one hand, the parent(s) were so happy to have any treatment for their child that they reported improvement which failed verification upon questioning. There were also cases where parent(s) were unhappy with a transient behavioral complication and on that basis initially failed to report improvement that was identified as present after further discussion. The term 'appears normal' was defined as the clinician and parents’ judgment that the child spoke normally in the office and at home, i.e., appeared to speak like others his/her age – a score that was not achieved in this study by any of the children. ‘Nearly normal’ was defined to mean that the child responded verbally to questions and/or initiated ostensibly normal speech appropriate to the current context yet the trained clinician identified mispronunciations, poor or odd word choice, unusual fluency or sparseness, unusual grammatical errors such as odd tense errors, pronoun gender errors, or pluralization errors. The difference between ‘short meaningful 1-2 word phrases’ and ‘short meaningful phrases’ lies in the fact that the latter required at least 3 or more words per phrase. This distinction required a dialog with specific probing for examples between clinician and parent(s). The revision addresses the definitions, means of collection, and limitations of the CLSQ more specifically.

6. The scores were obtained for the control group in such a different way as to seriously damage the appropriateness of comparing scores between these two groups. In addition - where are the scores for the untreated group? Why was mean change reported in scores only for the treated group, and not for the untreated group?

Reply: Unfortunately, for the non-steroid treated (NSA) group, CLSQ scores were not be obtained as these subjects did not receive pharmacological treatments, for which the CLSQ had been developed. The NSA subjects were identified retrospectively. The only systematic language information available for
them were the neurology notes at the two time points. This is now explained more clearly.

7. Most important, scores were not made blind to group membership or backed up by any objective data that might be more immune to subjectivity. I have seen many parents report on subjective improvement in language, behavior, mood, etc. when the child is involved in a treatment in which the parent is highly invested, for whatever reason, while objective testing shows no change in functioning. The paper claims that ‘All study participants had two sequential formal language and behavioral’ – this language assessment is not a formal assessment.

Reply: Scores for STAR patients are discussed above in Item 5. STAR group patient scoring was done before the current retrospective study was conceived and before the FMAER data were analyzed. As for the NSA group, the scorer did not know which of the untreated 75 subjects would meet NSA group study criteria and thus would be included in the current study. Nevertheless we agree that this is not a truly ‘blinded’ evaluation. It was unbiased only in terms of the final identity of the study subjects to be utilized.

Due to proofing oversight on our part, inclusion of the incorrect statement that all study participants had two sequential formal language and behavioral assessments was unfortunately overlooked in the paper as submitted. The statement is in error and has been removed from the revision.

8. My concerns about the claim that behavior improved dramatically and that 16 of the 20 children no longer met criteria for ASD are similar. The measure of behavior and diagnosis is not adequately described. The only description is that DSM-IV criteria for ASD are rated as 0 thru 3. How were DSM-IV symptoms assessed? By interview? By observation? By performing a structured or quantitative measure such as the ADOS? What is ‘traditional scoring’? Overall score = per DSM-IV manual – I am not aware of any overall score described in the DSM-IV. Again, one presumes that these judgments were made by the neurologist in consultation with the parent, who is likely to be invested in a successful outcome.

Reply: DSM-IV symptomatology-based questions were posed by the clinicians to the parents and subsequently confirmed or modified by the clinicians’ observation. Diagnosis of ASD was made by patterns of DSM-IV symptom item abnormalities as described in the DSM-IV manual. Since an additional, quantitative measure of behavior was desired for measurable detection of improvement or decline, each DSM-IV item was given a scaled score ranging from 0 to 3 (exclusive of A2b – language based item), as noted in Table 2 (0-3)). By adding up the scaled score values a total DSM-4-based symptom severity score was derived. This score was obtained at the first of the treatment group visits in collaboration of the clinician and the parent, and it was repeated at the termination of treatment, again obtained by the clinician in collaboration with the
parent. Although the DSM-IV-based scoring occurred without knowledge of the FMAER score, the DSM-IV measure cannot be considered ‘blinded’. This is now further emphasized in the revision.

9. Looking at the data, it appears that the average item score goes from a 2 (definitely present) to about a 1 (mildly present). The authors then appear to have decided not to count a ‘1’ as a symptom being present. All of these aspects of this method need to be described, including the non-blinded assessment (on the part of physician and parent), since this is the most startling claim of the study (that 16 of 20 children lost the diagnosis).

Reply: Given this valid criticism we have removed the description of loss of ASD diagnosis from the paper yet have retained the presentation and discussion of the scaled DSM-IV summary scores as estimates of ‘behavior’ only.

10. ‘Regression’ is not defined. It turns out not to be so easy to decide which children showed a convincing regression, which a plateau, and which some other pattern of development. Most papers reporting on regression require a certain level of skill before regression (for example, a certain number of words) and define which areas of function have regressed. In addition, recent prospective research no longer supports the clear distinction between regressive, plateau, and early impairment onset (see Ozonoff et al, 2010), although I have heard the convincing regression story often enough through the years that it is hard to dismiss. But in any case there should be some defined criteria for regression.

Reply: ‘Regression’ implies loss of age appropriate cognitive, language, and behavioral milestones; children with regression are typically first referred to child neurologists who seek to assess and confirm the validity and nature of the regression. This includes ruling out a host of potential diagnostic possibilities including encephalitis, tumor formation, epileptic encephalopathy, neuro-metabolic abnormality, genetic abnormality, and more. LKS is one form of regression that is associated with behavioral changes deemed autistic-like in younger children. None of the children in the STAR group met any of the likely relevant diagnoses, with special emphasis of a clear rule-out of LKS. By neurological history all STAR group children showed obvious regression. Patients with a ‘plateau’ rather than regression were excluded. Exceptions were children who showed an abrupt and profound regression followed by a low functioning plateau; these children were included in the STAR group. Most child neurologists are highly experienced in diagnosing a true regression. In earlier years such children were diagnosed as having a Landau-Kleffner Syndrome variant (LKSv), i.e., autistic like regression but without evidence of epileptic encephalopathy such as many EEG discharges in sleep, diagnostic of LKS. However, most epileptologists objected to the LKSv label since the child did neither show LKS nor epilepsy. Because of the autistic symptoms after regression of function the term regressive autism (R-ASD) came about and remains in active use to describe this group. More recently, autism specialists
have questioned the ‘autism’ portion of the name since this diagnosis typically does not get established by one of the ‘gold standard tests’ such as the ADOS. Furthermore, concerns have been voiced that regressive autism has been ‘over diagnosed’ by retrospective chart review and that regression and plateauing are often confounded. Some day there may be a better name for this entity. As concerns this study patients in the STAR group were clinically referred to as R-ASD. All the STAR group subjects had been referred for neurological evaluation due to unequivocal regression.

The paper has been modified to define ASD with regression more thoroughly.

11. The follow-up seems quite important for clinical implications, but is left very vaguely defined and reported.

Reply: We have modified and clarified the statement in accordance with this valid criticism.

12. How many patients’ records would have qualified for the STAR group? Were all that qualified included or were they selected in some way?

Reply: All patients who met selection criteria were included in the study. We have clarified this point.

13. What were the clinical features that led to steroid recommendation? Can that be described any more completely? Were all regressive cases recommended for such treatment, and parents’ willingness the deciding factor? I understand that this may be too complex and individual to describe in any detail, but it raises the question of characteristics on which the non-treated group might have differed.

Reply: Some patients were referred by their clinicians for consideration of steroid treatment. Some parents requested steroid treatment based upon their reading of the popular literature, and for some children steroid treatment was recommended due to the convincing history of regression and the identification of an absent FMAER. Patients who refused steroids obviously were not included in the STAR group. Other patients not included dropped out for various reasons, such as distance to the hospital, etc; other not included in the study may have received steroid therapy yet did not have the before and after neurophysiological testing recorded in the existing database; this makes it impossible to answer the question of referral bias. The 20 subjects included in the study are all those for whom all required study data were available. These points have been emphasized in the revision.

14. ‘NSA subjects’ language scoring was performed ‘quasi-blinded’ in that approximately three times as many language reports were scored retrospectively as were declared eligible for inclusion in the analyses’ – this is rather confusing.
How were the reports selected? What makes this a blinding procedure?

Reply: As mentioned earlier (see point 7. above) 75 records were reviewed of children who met the criteria as having ASD and being in the study’s desired age range. All were language scored. Twenty four of the 75 met the additional study criteria including the existence of two neurophysiology investigations. All 75 subjects had been scored for language performance prior to the selection for this study and without knowledge of the children’s FMAER results. While this may not constitute complete blinding, it provides a description of the degree of protection form bias as regards the language scores. The term ‘quasi-blinded’ has been eliminated from the revision.

15. ‘Although the utilized database did not contain information as to the families who choose not to proceed with the above steroid protocol, an informal estimate identified approximately 1 of 5 or 20% of families who declined steroids and instead chose anticonvulsant therapy, or declined all pharmacotherapy. ‘ This is rather confusing as well– are the authors saying that 4/5 of families to whom steroids were offered adhered to the protocol, or something else?

Reply: Since we did not record the actual data on this issue the 20% figure represents an educated guess only; therefore the comment has been eliminated from the revision.

16. In the title of Table 6. ‘Effect of steroids on CLSQ difference scores for STAR group’, this is really misstated – these data are not the effect of steroids – they are change in score over time, which the authors are attributing to the use of steroids.

Reply: This is correct and the Table title and text have been modified accordingly.

17. There is also an error in the footnote to this table.

Reply: Thank you for this pick-up; the error has been corrected.

18. The authors also say ‘The results revealed a very significant improvement for the STAR group as compared to the NSA group both in terms of receptive as well as expressive language ‘ but the scores for the untreated group are not given – what they are referring to is a greater frequency of children reported as improved in the treated group.

Reply: This point is now clarified in the revision: The NSA group was, indeed, not numerically scored for analysis at each study point. For the quantification of language change in the NSA group a five point rating scale was developed in order to quantify language change between the NSA group members’ two study points. For Fisher exact analysis these 5-point ratings were reduced to two,
namely ‘Better’ or ‘No difference/Worse -. The NSA group’s 5-point scores of change were used to determine their correlations with the FMAER change scores. Change score means are shown, separately, for expressive and receptive language. Additionally, the incorrect wording regarding the number of subjects per group that showed change has been corrected in the revision.

19. ‘The process clearly put stress on the family’ – could that be explained a bit more? What kind of stress was reported by families?

Reply The main stressors (see Table 12) were as follows: (1) Careful management of food intake in the face of increased appetite and the need to limit monthly weight gain to hold off development of hypertension – many children were hungry and demanding of constant food intake (100% of subjects), (2) Management of behavioral changes such as increased irritability in 50% of subjects; and (3) Management of sleep disturbances. This has been described more clearly more in the revision.

20. Finally, it needs to be at least mentioned that the children were probably receiving other treatments (such as behavioral or educational treatments), and these would need to be equated or partialled out between groups to be able to draw firm conclusions.

Reply: Actually none of the children in the study received any other specific treatments of any specialized kind during the time of the steroid administration and its equivalent for the non-pharmacologically treated group. This clarification has been added to the revision.

Conclusion:

I do believe the authors are reporting on a possible treatment that rests on firm theoretical and clinical grounds and needs to be carefully studied (as they suggest in their conclusions), and that might have tremendous import for families and children. What they have here is a series of cases that could be described in more detail and presented as such. The control group differs on so many methodological and clinical features that it’s not of too much usefulness, in my view. The biggest concerns are the reliance on parent report of improvement, when parent surely are invested in the success of the treatment, the very weak measures of language and behavior, and the overstated claims and conclusions especially in the abstract, which have the potential to lead parents and even doctors to think that these conclusions are firm enough to act on on a widespread basis.

Reply: Our response to these observations is four-fold:
1. The primary goal of the current paper is to provide adequate evidence to encourage a full-fledged study of the effectiveness of steroid treatment for the children in question. We fully agree that the current evidence which results from
a retrospective pilot investigation is not strong enough to indicate the use of steroids without a larger confirmatory a priori trial.

2. The retrospective development of the NSA group served the evaluation of FMAER changes over time in a non-steroid-treated comparison group. The CLSQ which had been developed and used clinically and not originally intended as a research tool allowed demonstration of association between the FMAER changes and language improvement.

3. Based on another reviewer's criticism, an EEG ‘noise analysis’ was newly performed and is now included in the manuscript’s revision. It further serves to strengthen the current findings of this preliminary retrospective study. The noise analysis demonstrated that the FMAER, which preceded steroid treatment, showed a two-fold deficit, namely (1) a low amplitude 4 Hz response, and (2) a documented response distortion (spurious frequencies production). The post-treatment response showed reversal of both these defects. This is now detailed in the paper.

4. We have taken care to emphasize, in the revised manuscript, the retrospective, preliminary nature of the study and its findings. Hopefully the revision will be considered as much improved and adequately responsive to the reviewer’s many well-taken and incisive comments.

Reviewer 3: Editors Comments

1. The paper is certainly tackling an important target – What can be done to treat and hopefully reverse the regression that happens to autism after early language and social development appear to be on a normal track. The lead author leads an experienced team working with childhood neurodevelopmental disorders. Clinical and EEG/ERP outcomes are considered. ERP measures are quantitative and objective, and outcomes are quite promising. These are all important strength

   Limitations however are significant, stemming from the comparison group that provides a weak basis for group comparison, very weak measurement of language skills, relatively weak diagnosis of ASD.

Reply: We agree with the characterizations of the study’s limitations; the revised manuscript addresses the limitations clearly and emphasizes the preliminary nature of the study.

Treated and untreated patients are identified retrospectively. It is of concern that the groups were not the same pre-treatment with different problems, so comparing them post hoc risks identifying people with different problems rather than differential treatment outcomes. Most obviously, the treated group had regression and this was not a requirement in the comparison group. This limits inferences that can be made from group comparison. A comparison group of
untreated cases with regression would provide much stronger basis for inferences from group comparison.

**Reply:** The Editor is quite correct. These limitations are now emphasized.

> The lack of healthy controls, especially on the EEG measures, makes it hard to know what is even abnormal.

**Reply:** No ‘epileptiform’ EEG transients (spikes, spike waves) were recognized in either group. The sharp waves, sharp theta, and paroxysmal theta features found are considered benign transients. Subjects, however, who manifested epileptiform transients on routine EEG with sleep were excluded from the study in order to avoid confounding the study subjects with likely LKS patients. Part of the study’s goal was to investigate whether benign EEG transients were affected by steroids; such effects were not identified. There was an error in Table 5 with transposition of 8 and 3. Once corrected the Fisher 2 x 3 test was not significant. Thus, we concluded that there was no evidence to suggest that benign EEG transient activity contributes to the treatment response.

> 3) Patients were not diagnosed by gold standard diagnostic procedures (ADOS/ADI).

**Reply:** The STAR group subjects were all classified as ASD by the intake DSM-IV criteria. The NSA group subjects were diagnosed as autistic by their referring child neurologists or child psychiatrists. Please, also see our response to Reviewer 2, (10.) regarding ‘regression’ and ‘autism’ designations. Child neurologists are well trained in clinical documentation of ‘regression’ which for this study included complete loss and/or deterioration of language, cognitive, and behavioral function. Subjects who on intake manifested a developmental plateau were excluded. Regression due to other common causes, in particular LKS, was excluded. All subjects in both groups met the autism spectrum disorder classification by DSM-IV criteria. The subjects who presented with an acute and obvious deterioration as agreed upon by parents and physicians in past years were referred to as Landau-Kleffner Syndrome ‘variants’ (LKSv). Epileptologists have long objected to this term since true LKS requires either seizures, a diagnosis of epilepsy and/or an epileptic encephalopathy, which all had been ruled out for the current study’s subjects. Once diagnosed the subjects were referred to as presenting with Regressive Autism (R-ASD). Some autism specialists have objected to this term since typically for such patients the current gold-standard ADOS is not available as they are diagnosed by neurologists who have ruled out all other encephalopathies as etiology for the presenting symptoms.

> 4) The approach for assessing language is weak – gross clinical ratings base apparently on chart notes of clinical observation.
**Reply:** The STAR group subjects were prospectively evaluated by the CLSQ. The CLSQ had been developed to be clinically helpful in the evaluation of ongoing change in language in R-ASD treatment. No other instrument appeared to be available to Child Neurologists that is suitable for assessment and clinical decision making during office visits. Use of the CLSQ for correlation with FMAER change was decided well after all subjects’ CLSQ data had been collected. The CLSQ data were all obtained and entered into the child’s clinical data base prior to any investigation of FMAER changes. The abstractions of language function from clinical reports as used for the NSA group is more questionable, indeed, yet it was performed in order to demonstrate that in the case of untreated ASD of the NSA group subjects no obvious improvement was identified. The NSA subjects spanned the same age range and complied with the same separation epoch between the two study points investigated as did the STAR group. Nevertheless, this and other weaknesses are now additionally emphasized in the paper.

5) *The quantitation of the resting and evoked ERPs is not as complete as it might be, and examination of whether there is a change in global resting or evoked power (vs. at 4 Hz) is not presented.*

**Reply:** Spectral analysis was not performed on the EEG during stimulation – although this is an alternative to signal averaging. The FMAER was formed and the spectral analysis of this resulting FMAER was contrasted with the residual EEG noise in the same data segment. In order to accomplish this signal averaging was repeated using the plus-minus approach, which we describe in detail. Our findings suggest that the FMAER ‘defect’ at analysis time 1 is two-fold, and consists both of a low amplitude 4 Hz response, and furthermore it is coupled with consistent noise responses (distortion) at other than 4 Hz. At time point 2, the 4 Hz response was much larger than at time 1; moreover the time 2 data failed to show any evidence of significant distortion. Thus, treatment appeared to augment the desirable 4 Hz response and diminish unwanted spectral distortion components. This information has been added to the paper’s revision.

Also, as the authors know, advocating new treatments without sufficient evidence has led to diverse polypharmacy treatment of ASD with often little benefit and adverse consequences. We want promising pilot findings to be universally valid and robust, but this is an area requiring far more caution than is evident in wording of the paper.

**Reply:** We agree and hope that the changes made in the revised manuscript are considered responsive to these sensitive and important issues. Our prime goal is to indicate that well designed prospective studies must be undertaken to assess definitively the effectiveness of steroid therapy in regressive autism (R-ASD, LKSv).

Specific issues:
Abstract

I think the estimate of regression in abstract is too high, unless it is very (too) loosely defined.

Reply: The source of the regression estimate reported is the Simon Foundation Autism Research Initiative (SFARI), which states: “About one in three children with autism abruptly lose language, social or other developmental skills in their second year of life, according to a meta-analysis published 2 August in the Journal of Autism and Developmental Disorders” by Barger et al. We have reviewed this topic more thoroughly in the introduction.

Starting results and conclusion sections saying “steroid treatment produced ..…” and “Steroid treatment greatly improves …” are far too aggressive an interpretation of the data given all the study design weaknesses. One can talk about group differences and say “they suggest that steroids may” to interpret them very modestly/carefully, but to infer causality in this context is greatly premature.

Reply: We agree with the Editor and have made corrections accordingly in the revision.

In summary, my view is that different reasonable reviewers could see the problems and say it should not be published, and others who could see the strengths and feel otherwise. My view is rather in the middle. There are important weaknesses that need to be addressed, but that because the findings especially with the ERP are sufficiently promising and because there is such high need for treatments for those with regression, with important revisions the data can be a useful contribution.

1) a very careful effort is needed to tone down inferences that they have shown this is a useful treatment. This is no more than a set of promising observations from a post-hoc retrospective exploratory project. It is not only scientifically wrong to say more, but it is dangerous to imply that the data show this to be an effective treatment for those showing regression. ASD families do not need another treatment added on to drug regimens that already are too large and guided by far too little evidence.

Reply: We agree with the Editor. We emphasized in the revision that our preliminary data might serve to spur a well-designed prospective study.

2) An explicit section at the end is needed on study limitations, listing all the limitations and restating that the findings are post hoc, no more than suggestive, and are presented to spur RCT studies to learn whether this approach is useful. And for studies to better clarify just who the treatment will
help in much more objective terms than are now available so that the treatment is used selectively and not inappropriately. Desperate families are grasping for straws and children’s treatment can suffer.

Reply: We fully agree with the Editor and have elaborated the discussion section substantially regarding the limitations of the study.

> I had a postdoc in my lab doing frequency modulated ERP studies of ASD to take a look at the study methods, and attach her comments below.

> While the frequency analysis of the EEG itself is rather simplistic, there is nothing inherently wrong in the quantification of the 4 Hz frequency response. One question must be asked, however, given Figure 1: the before and after FFTs are done with slightly different frequency resolution, according to the figure (upper right bar in the FFT window, .99 and .98 Hz). While this is certainly not enough to cause the differences seen between the before and after frequency-wise, it suggests (since BESA normally sets the frequency resolution on an FFT based on the width of the window) that there is a ~10 ms discrepancy in the size of the windows used in the before and after averages. While this itself, again, would not cause the changes seen here, it does suggest that the before and after analyses were slightly different and should probably be double-checked.

Reply: The Postdoc identified correctly a reporting error. In fact, the frequency resolutions for both the right and left were identical. This has been corrected.

> It would be more convincing that steroid usage had a specific effect on tightening the FMAER frequency response if the overall distortion were in some way quantified and compared. While the single subject in Figure 1 is certainly a very nice example of the researchers’ hypothesis, it cannot be ruled out with the current analytical methodology that steroid usage simply increases power in many low frequencies, including the one of interest. The authors make the point many times that distortion is decreasing with steroid treatment and stress that the FMAER is an objective test, but do not actually quantify distortion in any way, therefore these conclusions are unjustified with the current statistical evidence. The figure depicting a single subject does appear to show a decrease in distortion in this particular subject, however a statistical handling of these observations would be much more beneficial.

Reply: We have added a section, which shows the quantification of the distortion and statistically assesses the distortions identified.

> Overall the findings are interesting and potentially very valuable to treatment development in this population; however the claims of reduced distortion in the FMAER response with steroid treatment are largely unsubstantiated by the current statistical methods and would benefit from some additional quantification of the change in “noise” after treatment.
Reply: We have included evidence that indicates both an augmented 4 Hz FMAER response and also significant reduction in spectral distortion of response as associated with steroid treatment.

In summary, we thank the reviewers for their detailed and incisive criticisms and observations and hope that the changes made in the revised manuscript are considered responsive to the important issues raised. Our prime goal is to indicate that well designed prospective studies must be undertaken to assess definitively the effectiveness of steroid therapy in regressive autism (R-ASD, LKSv).