Reviewer's report

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

Version: 2 Date: 18 November 2013

Reviewer: Deborah Fein

Reviewer's report:

STEROID REVIEW

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.

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.

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.

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.

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.

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.

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?

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.

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.

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).

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.

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

Minor Essential Revisions

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

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.

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?

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?

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.

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

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.

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

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.

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.

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests: 

i have no competing interests.