Reviewer’s report

Title: Neurodevelopmental outcome of Extremely Low Birth Weight Infants at 24 months corrected age: a comparison between Griffiths and Bayley Scales

Version: 1 Date: 3 July 2015

Reviewer: Dianne Creighton

Reviewer’s report:

Major Compulsory Revisions

1. There is confusion in how the research questions are stated. At first, the purpose was stated as “to define which edition [of the Bayley] would show the highest agreement with the Griffiths…” (line 33, bolding mine). Later, a hypothesis was added, “that Bayley-III would display significantly higher Composite Scores compared to both the Bayley-II and the Griffiths Scales” (lines 89-91, bolding mine). But then, in the Discussion section, they state that “Contrary to our hypothesis, the comparison with the Griffiths Scales revealed a higher agreement with the Bayley-III mean composite scores than with the previous edition ones” (lines 254-255). Note that there was no hypothesis which stated that the agreement between the Griffiths and the Bayley-III would not be higher than the agreement between the Griffiths and the Bayley-II, only that the Bayley-III composite scores would be higher than the Bayley-II and Griffiths scores. Thus, there is confusion about what their hypothesis is. Correction of the inconsistencies about the research question would be a Major Compulsory Revision.

2. There is confusion about psychometric terminology. In the statement of the aim of the study, “to define which edition of the Bayley Scales better agrees with the Griffiths one” (lines 87-88) the authors continue, “and, thus, can be considered the more reliable assessment procedure for the ELBW follow-up” (lines 88-89, bolding mine). I would argue that it is the validity of the Bayley assessment, not the reliability that is being evaluated, using the Griffiths as the gold standard. Next, the authors confuse validation with norming. “Since neither the Bayley-II nor the Bayley-III have been validated in Italy, the USA norms…were used” (lines 145-146). This sentence should read “neither… has been normed in Italy”. Also, they state that “Since an Italian validation of the Griffiths…is not available we referred to the …UK norms” (lines 160-161). Again, instead of validation the authors should refer to the development of norms. Accurate use of psychometric terminology would constitute a Major Compulsory Revision.

3. I have a question about Table 1, which states that “Italian as mother language, %” is 19.6 for both groups. However, in the text it states that “all infants were exposed to Italian as primary language” (line 212). What does this discrepancy mean? Was there one “mother language” that was more common than Italian, or was 19.6% the largest language group in the samples? Given that the child’s
language development was a key component of the data for this article, I think that clarification of the primary language of the children and the language used in test administration is a Major Compulsory Revision.

Minor Essential Revisions

4. The formal name of the ethics committee that approved the study should be stated, and not just “the department ethics committee” (line 96).

5. The definitions of impairment should be stated as research definitions, since they are not part of the interpretive guidelines of the Griffiths or Bayley tests. For example, “an index composite score <70…indicates severe impairment” (line 142) should be re-worded to say something like “an index composite score <70 … is defined here to indicate severe impairment.”

6. There is a mix of terminology used for lower scores, with the term “impairment” (which is defined) used sometimes, and the term “delay” (which is not defined) used at other times. I would advocate for using the term “impairment” consistently.

7. It is not stated what language was used in the administration of the tests, given that “all infants were exposed to Italian as primary language” (line 212). Was an Italian translation of the Griffiths and the Bayley used? If so more details are needed.

8. Note that the list of abbreviations has an error, with ROP defined as Bronchoulmonary Dysplasia, and BPD missing.

9. I found the first sentence of the Statistical Analyses section difficult to understand: “The homogeneity between the two groups of children has been verified through the confidence interval at 95% of the differences between either the means or the proportions of the variables taken into consideration” (lines 178-180). I also found it unusual to read comparisons of percentage rates to be presented in the form of pluses or minuses within the text of the paper. For example, “Bayley-II showed consistently higher rates of severe impairment both in cognitive and language abilities (+14.1%) and in motor skills (+15.3%)” (lines 232-234). Another example: “Bayley-III Motor Index highly agreed with the Griffiths … but identified -9.8% of severe impairment…” (lines 242-243). These are only two examples of a presentation style commonly used in this manuscript. I recommend that revisions to the wording of the data analyses and discussion of the data be considered Minor Essential Revisions.

10. There are a number of unusual English language usages that need attention. Examples include: “the Bayley-III showed -3% of severe impairment” (line 44). The preposition “of” is used incorrectly. Also, the pronoun “one” is used awkwardly several times, such as “Bayley-III composite scores are up to 10 points higher than Bayley-II ones” (line 69-70). Editing of the manuscript is needed.

11. Note a typographical error in Table 5: “Jouden” should be “Youden”.

Discretionary Revisions

12. The conclusion that “the Bayley-III… tends to underestimate neurodevelopmental delay (line 48-49)” is somewhat biased. The results indicate
that the Bayley-II yields higher rates of neurodevelopmental delay than the Griffiths, so a conclusion could be that the Bayley-II tends to overestimate neurodevelopmental delay.

13. I would add a limitation referring to the use of US Norms for the Bayley versus UK Norms for the Griffiths. As well, it would seem to be a limitation that an Italian translation, especially of the language scales, was not used, assuming Italian to be the primary language of most of the infants.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.