Improving evaluation of obstetric interventions
van 't Hooft, Janneke

Citation for published version (APA):
van 't Hooft, J. (2016). Improving evaluation of obstetric interventions

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 5

PREDICTING DEVELOPMENTAL OUTCOMES IN PREMATURE INFANTS BY TERM EQUIVALENT MRI: SYSTEMATIC REVIEW AND META-ANALYSIS

Janneke van ’t Hooft, Johanna H. van der Lee, Brent C. Opmeer, Cornelieke SH. Aarnoudse-Moens, Arnold GE. Leenders, Ben Willem J Mol, Timo R. de Haan.

ABSTRACT

Background
This study aims to determine the prognostic accuracy of term MRI in very preterm born (≤32 weeks) or low-birth-weight (≤ 1500 g) infants for long-term (>18 months) developmental outcomes.

Methods
We performed a systematic review searching Central, Medline, Embase, and PsycInfo. Two independent reviewers performed study selection, data extraction and quality assessment. We documented sensitivity and specificity for three different MRI findings (white matter abnormalities (WMA), brain abnormality (BA), and diffuse excessive high signal intensity (DEHSI)), related to developmental outcomes including cerebral palsy (CP), visual and/or hearing problems, motor, neurocognitive, and behavioral function. Using bivariate meta-analysis, we estimated pooled sensitivity and specificity and plotted summary receiver operating characteristic (sROC) curves for different cut-offs of MRI.

Results
We included 20 papers published between 2000 and 2013. Quality of included studies varied. Pooled sensitivity and specificity values (95% confidence interval (CI)) for prediction of CP combining the three different MRI findings (using normal/mild vs. moderate/severe cut-off) were 77% (53 to 91%) and 79% (51 to 93%), respectively. For prediction of motor function, the values were 72% (52 to 86%) and 62% (29 to 87%), respectively. Prognostic accuracy for visual and/or hearing problems, neurocognitive, and/or behavioral function was poor. sROC curves of the individual MRI findings showed that presence of WMA provided the best prognostic accuracy whereas DEHSI did not show any potential prognostic accuracy.

Conclusions
This study shows that presence of moderate/severe WMA on MRI around term equivalent age can predict CP and motor function in very preterm or low-birth-weight infants with moderate sensitivity and specificity. Its ability to predict other long-term outcomes such as neurocognitive and behavioral impairments is limited. Also, other white matter related tests as BA and DEHSI demonstrated limited prognostic value.

Systematic review registration. PROSPERO CRD42013006362
BACKGROUND

Preterm birth is associated with an increased risk of neurodevelopmental problems. Magnetic resonance imaging (MRI) is increasingly being used to identify cerebral white matter lesions in the brain of preterm infants at term equivalent age. It is claimed to be a valuable tool to predict neurodevelopmental outcomes in very preterm infants and its clinical use is, therefore, being promoted. However, the prognostic accuracy of white matter related MRI abnormalities for long-term developmental outcomes is debatable and its use as a standard of care is not yet recommended by the American Academy of Neurology Quality Standards. The lack of meta-analytic synthesis of the primary studies reporting prognostic values, which tends to show conflicting results, hampers the debate.

Subsequently, the lack of knowledge about the prognostic accuracy of term MRI hampers an adequate interpretation of this test. This may invoke unwanted effects, as parents may worry unnecessarily about the possible abnormal development of their child. However, if term MRI can predict neurodevelopmental outcomes accurately, the use of this expensive diagnostic procedure as part of standard care could be justified as it may select high risk infants for prolonged and intensive supportive care.

Our study aims to evaluate the following two questions:

1. What is the prognostic accuracy (in terms of sensitivity and specificity) of white matter related abnormalities seen on term MRI for long-term developmental outcomes of infants born very preterm or with low birth weight?

2. Is there a difference in prognostic accuracy between the three types of white matter abnormalities as seen on term MRI including white matter abnormality, a combination of cerebral white matter lesions defined as ‘brain abnormality’, and diffuse excessive high signal intensity? To answer these questions, we performed a systematic review and meta-analysis on the subject.
METHODS

We performed a systematic review following the guidance of the PRISMA statement, Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy and other recommendations found in the literature,7–9 with a prospectively published protocol at the Prospero database (www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013006362#.VVMAX47tIBc).

Search strategy
We searched Central, Medline, Embase, and PsycInfo from their inception to November 2013 for relevant studies. The search was performed by a trained clinical librarian (AL) and two other authors (TdH and JvH). Broad text and MeSH terms were used. Also, keywords of eligible papers were screened and included in the final search. We did not apply any language restrictions. The search was limited to studies including humans. The full search in all these databases can be seen in Additional file 1. References from included studies were checked. Abstracts and reports from meetings were included only if they related directly to previously published work.

Eligibility criteria
The following inclusion criteria were used to select studies: (1) the study pertained to infants born at a gestational age ≤ 32 weeks and/or birth weight ≤ 1500 g; (2) MRI should be planned at term equivalent age (37-42 weeks) with a maximum range of 3 weeks earlier or later (34-45 weeks); (3) MRI findings should be related to any developmental outcome; and (4) developmental follow-up should be performed ≥18 months postnatal age. Isolated single case studies and review articles were not included.

Abstracts were screened for eligibility by two independent reviewers (JvH and TdH). Full-text articles were retrieved if applicable to the core research question, or if the abstract did not supply sufficient information. Any disagreement was set by discussion until consensus. The same two reviewers appraised the methodological quality and performed the data extraction. Any disagreement at this stage was resolved by a third reviewer.
Methodological quality
Due to lack of existing quality assessment tools for prognostic accuracy studies, we developed a modified version of the QUADAS-2 assessment tool\textsuperscript{10} to evaluate the risk of bias (see Additional file 2).

Data extraction
A standardized data extraction form (see Additional file 3) was used to record study information. The results of white matter abnormalities (WMA) and brain abnormalities (BA) are usually expressed as either no, mild, moderate or severe abnormalities as described by Inder and Woodward et al.\textsuperscript{11,12} Where possible we defined two cut-offs, i.e., (1) no abnormality vs. mild, moderate or severe abnormality, reported as ‘normal vs. any’ and (2) no or mild abnormality vs. moderate to severe abnormality, reported as ‘normal/mild vs. moderate/severe’. BA was defined as a combination of WMA plus presence of other brain abnormalities such as ventricular haemorrhage or increased ventricle size. For diffuse excessive high signal intensity (DEHSI), the results are usually expressed as either present or absent. Therefore, only one cut-off was used in the 2x2 tables presenting the results for these MRI findings.

The cut-off point for unfavorable developmental outcome was defined as a minus 2 standard deviations (-2 SD) difference from the mean for each MRI finding. If this cut-off was not reported (but for example, only a -1.25 or -1 SD), we used the reported cut-off in the meta-analysis. In cases of duplicate reporting, i.e., the same cohort was described in two papers or one paper reporting developmental outcomes at different time points of age, we used data from the paper that reported the developmental outcome at a comparable age with the other included papers. For example: if two papers reported motor skills at 2 years of age and one paper reported at 2 and 6 years of age, the reporting at 2 years of age was used. In case two papers reported the same cohort at similar ages, the study with the largest sample size and least quality concerns was selected. If the required data could not be extracted from the publication, authors were contacted by email. All data were entered in Review Manager (RevMan) version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012.
Statistical analysis
We performed a meta-analysis using a bivariate modelling approach. In view of the observed heterogeneity, a random-effects model was used. We compared pooled sensitivity and specificity (95% confidence intervals); likelihood ratios of positive and negative test results (LR+/LR-) were calculated from the pooled sensitivity and specificity; diagnostic odds ratios (DOR), and posttest probabilities of three different MRI findings (WMA, BA and DEHSI), for all types of developmental outcomes. Sensitivity and specificity for individual studies and summary ROC curves (sROC) were plotted to visualize possible heterogeneity of data and overall test accuracy.

RESULTS
Our search strategy yielded 1,311 citations after removal of duplicates (Figure 1). A total of 44 papers met the inclusion criteria, of which 27 papers provided 2×2 tables. One more relevant paper was identified by contact with the authors. After excluding multiple publications from the same cohorts (8 papers), a total of 20 papers were available for the meta-analysis.

The 20 papers were all published between 2000 and 2013. These papers reported on 12 different cohort studies (2 retrospective and 10 prospective) including 1,287 patients (682 male and 605 female). The extracted data provided 54 2x2 tables for WMA, BA or DEHSI. These three MRI findings were used for the prediction of various developmental outcomes: cerebral palsy (CP), visual and/or hearing problems, motor, neurocognitive, and behavioral function, as well as a combination of problems in these domains defined as ‘neurodevelopmental impairment’ (NDI). Study characteristics are shown in Additional file 4: Table S1.

Studies from which 2x2 tables could not be derived (n=17 papers, not reported in this manuscript) reported continuous data with no cut-offs. These studies mostly reported the following MRI tests: cerebellar abnormalities, volumes and diameter measures of the brain (total brain or specific regions as hippocampus, corpus callosum or ventricles).
Methodological quality of included studies

In general, 70 to 90% of the included studies scored positive on each of the QUADAS-2 quality assessment items (Figure 2). For example, 90% of the studies included in the meta-analysis used a consecutive sample of very preterm born and/or low-birth-weight neonates over a specific period of time in their clinic (Figure 2). In general, a good description of the MRI test and reference standard was provided, as well as a verification process to all neonates who had a MRI performed. However, almost 50% of the papers did not report blinding of the test results, i.e., results of the MRI findings are not (made) available to the person performing the follow-up neurodevelopmental test.
Chapter 5

Figure 2. Quality assessment of included studies in meta-analysis (n=20).

Meta-analysis

The reported sensitivity and specificity were generally higher for the WMA tests when compared to BA or DEHSI findings (Table 1). Figure 3 shows the sROC curves for prediction of four different developmental delays related to any MRI abnormality (combination of WMA, BA or DEHSI tests) using a 'normal/mild vs. moderate/severe' cut-off. The sROC curve for prediction of CP shows a curve that lies the most towards the (optimal) upper left corner of the ROC space. Also the sROC curve for prediction of motor function has a tendency to the upper left corner. The sROC curves for mental impairment and neurodevelopmental impairment, which are visualized in Figure 3, are heading more towards the diagonal (non-discriminating) line of the ROC space.
Table 1. Results from bivariate analysis on sensitivity (Sens), specificity (Spec), 95% confidence interval (95% CI), diagnostic odds ratio (DOR), positive/negative likelihood ratio (LR+ and LR-), and pretest and posttest probabilities

<table>
<thead>
<tr>
<th>MRI test with used cut off</th>
<th>Developmental Outcome</th>
<th>No. of studies</th>
<th>No. of neonates</th>
<th>Sens (95% CI)</th>
<th>Spec (95% CI)</th>
<th>DOR</th>
<th>LR+</th>
<th>LR-</th>
<th>Pretest probability (%)</th>
<th>Posttest probability Positive</th>
<th>Posttest probability Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMA -'normal vs. any'</td>
<td>CP</td>
<td>1 [35]</td>
<td>125</td>
<td>100 (70-100)</td>
<td>81 (73-87)</td>
<td>&gt;100</td>
<td>5.27</td>
<td>0.01</td>
<td>7.2</td>
<td>29.0</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>IQ</td>
<td>2 [36, 37]</td>
<td>283</td>
<td>79 (65-88)</td>
<td>41 (18-69)</td>
<td>2.61</td>
<td>1.34</td>
<td>0.51</td>
<td>16.6</td>
<td>21.1</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>Language</td>
<td>2 [37, 38]</td>
<td>283</td>
<td>87 (69-97)</td>
<td>30 (23-39)</td>
<td>2.78</td>
<td>1.24</td>
<td>0.44</td>
<td>5.3</td>
<td>6.5</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>Mental Development</td>
<td>3 [12, 35, 39]</td>
<td>448</td>
<td>81 (59-93)</td>
<td>49 (26-73)</td>
<td>4.13</td>
<td>1.60</td>
<td>0.39</td>
<td>13.8</td>
<td>20.4</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>Motor</td>
<td>3 [12, 35, 40]</td>
<td>485</td>
<td>87 (74-94)</td>
<td>51 (26-76)</td>
<td>7.29</td>
<td>1.80</td>
<td>0.25</td>
<td>13.3</td>
<td>27.4</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>Vision/hearing</td>
<td>2 [12, 35]</td>
<td>125</td>
<td>62 (13-95)</td>
<td>53 (23-82)</td>
<td>1.88</td>
<td>1.33</td>
<td>0.71</td>
<td>31.2</td>
<td>37.6</td>
<td>24.4</td>
</tr>
<tr>
<td>WMA-'normal/mild vs. moderate/severe'</td>
<td>CP</td>
<td>2 [41, 42]</td>
<td>164</td>
<td>67 (38-87)</td>
<td>92 (85-96)</td>
<td>22.35</td>
<td>8.11</td>
<td>0.36</td>
<td>6.7</td>
<td>36.8</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>IQ</td>
<td>2 [36, 37]</td>
<td>283</td>
<td>53 (39-67)</td>
<td>83 (77-87)</td>
<td>5.41</td>
<td>3.06</td>
<td>0.57</td>
<td>16.6</td>
<td>37.9</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Language</td>
<td>2 [37, 38]</td>
<td>283</td>
<td>47 (24-71)</td>
<td>86 (82-90)</td>
<td>5.46</td>
<td>3.38</td>
<td>0.62</td>
<td>5.3</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Mental Development</td>
<td>3 [12, 39, 41]</td>
<td>398</td>
<td>38 (26-52)</td>
<td>87 (83-91)</td>
<td>4.21</td>
<td>2.98</td>
<td>0.71</td>
<td>13.8</td>
<td>32.3</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Working memory</td>
<td>2 [43, 44]</td>
<td>258</td>
<td>24 (17-32)</td>
<td>88 (78-94)</td>
<td>2.26</td>
<td>1.96</td>
<td>0.87</td>
<td>47.7</td>
<td>64.1</td>
<td>44.2</td>
</tr>
<tr>
<td></td>
<td>Motor</td>
<td>3 [12, 40, 41]</td>
<td>435</td>
<td>54 (30-77)</td>
<td>90 (84-94)</td>
<td>10.59</td>
<td>5.37</td>
<td>0.51</td>
<td>17.5</td>
<td>53.2</td>
<td>9.7</td>
</tr>
<tr>
<td></td>
<td>NDI</td>
<td>1 [12]</td>
<td>167</td>
<td>38 (26-53)</td>
<td>86 (78-91)</td>
<td>5.28</td>
<td>2.70</td>
<td>0.72</td>
<td>28.1</td>
<td>51.4</td>
<td>22.0</td>
</tr>
<tr>
<td>BA-'normal vs. any'</td>
<td>CP</td>
<td>2 [36, 45]</td>
<td>277</td>
<td>90 (68-98)</td>
<td>60 (54-66)</td>
<td>13.71</td>
<td>2.27</td>
<td>0.17</td>
<td>7.2</td>
<td>15.0</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Mental development</td>
<td>3 [46, 48, 49]</td>
<td>273</td>
<td>90 (74-97)</td>
<td>80 (75-85)</td>
<td>37.83</td>
<td>4.56</td>
<td>0.12</td>
<td>11.4</td>
<td>36.9</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Behavior</td>
<td>1 [47]</td>
<td>177</td>
<td>76 (61-88)</td>
<td>37 (29-46)</td>
<td>1.88</td>
<td>1.21</td>
<td>0.64</td>
<td>23.7</td>
<td>27.3</td>
<td>16.6</td>
</tr>
<tr>
<td></td>
<td>Hearing</td>
<td>2 [36, 46]</td>
<td>397</td>
<td>100 (51-100a)</td>
<td>58 (53-63)</td>
<td>&gt;100</td>
<td>2.49</td>
<td>0.00</td>
<td>2.0</td>
<td>4.9</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>NDI</td>
<td>2 [36, 46]</td>
<td>424</td>
<td>81 (69-89)</td>
<td>68 (61-75)</td>
<td>9.44</td>
<td>2.57</td>
<td>0.27</td>
<td>13.9</td>
<td>29.3</td>
<td>4.2</td>
</tr>
<tr>
<td>BA-'normal/mild vs. moderate/severe'</td>
<td>CP</td>
<td>3 [46, 48, 49]</td>
<td>273</td>
<td>90 (74-97)</td>
<td>80 (75-85)</td>
<td>37.83</td>
<td>4.56</td>
<td>0.12</td>
<td>11.4</td>
<td>36.9</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Mental development</td>
<td>2 [46, 50]</td>
<td>216</td>
<td>82 (8-100)</td>
<td>75 (69-81)</td>
<td>13.80</td>
<td>3.31</td>
<td>0.24</td>
<td>8.3</td>
<td>23.1</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>NDI</td>
<td>3 [36, 51, 52]</td>
<td>405</td>
<td>63 (31-86)</td>
<td>73 (54-865)</td>
<td>4.53</td>
<td>2.32</td>
<td>0.51</td>
<td>23.5</td>
<td>41.7</td>
<td>13.6</td>
</tr>
<tr>
<td>DEHSI</td>
<td>CP</td>
<td>2 [41, 45]</td>
<td>421</td>
<td>46 (9-89)</td>
<td>39 (3-92)</td>
<td>0.55</td>
<td>0.76</td>
<td>1.38</td>
<td>6.2</td>
<td>4.8</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>Mental development</td>
<td>3 [35, 41, 53]</td>
<td>362</td>
<td>87 (72-94)</td>
<td>19 (9-35)</td>
<td>1.53</td>
<td>1.07</td>
<td>0.70</td>
<td>10.5</td>
<td>11.1</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>Motor</td>
<td>3 [35, 41, 53]</td>
<td>362</td>
<td>86 (70-94)</td>
<td>20 (8-40)</td>
<td>1.48</td>
<td>1.07</td>
<td>0.72</td>
<td>10.5</td>
<td>11.1</td>
<td>7.8</td>
</tr>
</tbody>
</table>

If bivariate model could not estimate 95% CI for pooled sensitivity for two studies, estimate is based on study with largest sample size

BA brain abnormality, DEHSI diffuse excessive high signal intensity, NDI neurodevelopmental impairment, WMA white matter abnormality

data derived after contact with author
Figure 3. Pooled sensitivity and specificity with sROC reporting four developmental outcomes detected by any MRI abnormality (including White Matter Abnormality, Brain abnormality or Diffuse Excessive High Signal Intensity using 'normal/mild vs. moderate/severe' cut-off).

a. Cerebral palsy
   - (n=seven studies): pooled sensitivity 77% (53 to 91%) and specificity 79% (51 to 93%).

b. Motor function
   - (n=seven studies): pooled sensitivity 72% (52 to 86%) and specificity 62% (29 to 87%).

c. Mental impairment
   - (n=seven studies): pooled sensitivity 66% (41 to 84%) and specificity 53% (35 to 71%).

d. Neurodevelopmental impairment (NDI)
   - (n=four studies): pooled sensitivity 61% (34 to 83%) and specificity 85% (75 to 92%).

The individual studies are visualized as squares with the horizontal axis corresponding to the total non-diseased neonates and vertical axis the total diseased neonates of that particular study population, i.e., a flat square represents a low prevalence of the disease and the surface of the square represents the size of the study population.
The pooled sensitivity and specificity values (95% confidence interval (CI)) for prediction of CP were 77% (53 to 91%) and 79% (51 to 93%), respectively. Almost similar values were found for the prediction of motor function with a sensitivity of 72% (52 to 86%) and specificity of 62% (29 to 87%). Lower values were found for mental development and NDI with sensitivity of 66% (41 to 84%) and 53% (35 to 71%) respectively and specificity of 61% (34 to 83%) and 85% (75 to 92%). Using a ‘normal vs. any’ cut-off pooled sensitivity and specificity values were 84% (45 to 97%) and 58% (27 to 84%) for prediction of CP; 76% (48 to 92%) and 26% (8 to 57%) for prediction of motor function; and 85% (74 to 92%) and 36% (20 to 56%) for prediction of mental development, respectively.

Figure 4 shows the sROC curves corresponding to the two different cut-offs: ‘normal vs. any’ and ‘normal/mild vs. moderate/severe’ when only the results of WMA are taken into consideration for prediction of various developmental outcomes. If only moderate to severe WMA lesions are coded as abnormal (‘normal/mild vs. moderate/severe’), the specificity increases and the sensitivity decreases.

The spread of the individual studies alongside the sROC curves in Figures 3 and 4 shows a substantial heterogeneity of the collected data explained by a threshold effect. The threshold effect is similar to the shift in sensitivity and specificity as described above, yet without an explicit change in cut-off levels. The shift is presumably the result of an implicit use of a different threshold, e.g., following from subjective judgments or calibration of diagnostic devices.
**DISCUSSION**

This study shows that the presence of moderate/severe WMA on MRI performed around term equivalent age can predict CP and motor function in very preterm or low birth weight neonates with moderate sensitivity and specificity. The ability to predict other long-term outcomes such as neurocognitive and behavioral impairments is limited. Also, other white matter related tests as BA and DEHSI demonstrated limited to no prognostic value.

In the last decade, the use of MRI as a screening tool for very preterm and low-birth-weight neonates has been a topic of major interest and several reviews have been published on its use. Most of these reviews are narrative (describing practical issues like sedation for MRI and/or different types of MRI techniques) or examined the impact of preterm birth and brain abnormalities.
on long-term development through the use of MRI. Although none of them systematically reported test accuracy of MRI for prediction of developmental outcome, most of these reviews, however, recommended the use of term MRI in clinical practice. To our best knowledge, our study is the first that systematically reviews the prognostic accuracy of different MRI findings on various long-term developmental outcomes.

**Clinical implications**

The data in our meta-analysis suggest that presence of moderate/severe WMA has higher positive likelihood ratio, and absence of any WMA has a higher negative likelihood ratio than any other test that we now use for preterm infants (e.g., cranial ultrasound or neurological examination). The prognostic accuracy of WMA finding on MRI therefore supports the use of MRI for preterm infants. However, whether this alters clinical management is a different question. Answering this question was beyond the scope of our meta-analysis. In our opinion however, showing potential prognostic accuracy of a test does not directly justify its clinical use as a standard test. The usefulness of this tool for clinical decision-making requires the presence of possible treatment or specified follow-up strategies following the results of the MRI. At present, there is no specific treatment available addressing the needs of infants with abnormal white matter on MRI. However, the use of term MRI results may give focus to specific follow-up programs (i.e., offering a screening tool for developmental disorders at an earlier age) or improve selection of neonates for early intervention programs (i.e., physiotherapy or speech therapy). Also, available MRI results may help parents of prematurely born infants to better prepare for the future.

On the other hand, after screening all very preterm born or low-birth-weight neonates with a term MRI, there is no other test available with better accuracy. Therefore the possible harm due to false positive and false negative results must be taken into consideration. The value of being timely informed (value of information) must be weighed against the possibility of unnecessary concern for adverse outcome. For example, based on the results of this meta-analysis, we can expect that the finding of moderate to severe WMA in a very preterm born child will increase the probability of developing CP from the known prevalence of 7% in this population to 37% (Table 1). This raises
the question if this increase in probability will change practice for both the clinician and patient. More specifically: will the clinician offer a different follow-up program when the risk of developing CP is 37% instead of 7%? And will the negative posttest probability of 2.5% (i.e., 2.5% will still develop CP after a normal MRI test result) justify a denial of follow-up to those with normal MRI?

Our meta-analysis also shows that adverse outcomes, such as neurocognitive and behavioral impairments, could not be predicted by term MRI abnormalities. Compromised white matter may result in more ‘subtle’ impairments in such areas of the child’s long-term function. The limited prognostic value of WMA for these specific outcomes also suggests that despite MRI abnormalities, whether or not a child develops neurocognitive and behavioral impairments, is also dependent on other factors. Such other factors may include the presence of a stimulating home and/or school environment, educational level of the parents and therapy use.23,24

Other considerations relevant to deciding on the use of MRI for the prediction of developmental outcomes are the substantial health care costs associated with its use. In many neonatal units, MRI technology is unavailable or its use is severely restricted. Also, expert neuroradiologists are needed for proper interpretation of the MRI results. In view of its potential prognostic capacity, it is therefore still debatable whether performing a standard term MRI is cost-effective.

Limitations
This meta-analysis has some limitations that need to be considered. Although a considerable number of studies were identified on the subject, only a limited number of data points were available for each specific combination of MRI findings and neonatal outcome. Although even the results of only two studies can be pooled, the limited number of data points and often limited sample size per study imply limited power (hence wide confidence intervals).25

Also the presence of heterogeneity may raise the question whether pooling of results is justified in our study. In prognostic meta-analysis two possible reasons for heterogeneity of the data are known i.e., clinical heterogeneity, due to differences in features of the cohorts, and heterogeneity due to threshold
effect. We estimate a smaller impact of the clinical heterogeneity as all cohorts included consecutive and comparable populations (although inadequate and inconsistent reporting of possible confounders in the studies, e.g., use of medication, birth weight and presence of neonatal complications during admission, made it impossible to correct for potential confounders in our meta-analysis). Heterogeneity due to threshold effect is a common occurrence in many diagnostic test systematic reviews and probably explaining most of the heterogeneity in our meta-analysis.9 The threshold effect in MRI tests is explained by the relative subjectivity of interpretations of MRI results e.g., one lesion on the MRI might be seen as abnormal for one radiologist, but not by another. Also the use of different scoring systems, and differences in background of the evaluators (neonatologists or radiologist) contribute to this type of heterogeneity. For this review, heterogeneity due to different scoring systems is probably the case in studies describing ‘brain abnormalities’. These studies not only include WMA as one of the MRI findings but also a composite of other MRI findings (i.e., IVH and/or increased ventricle size). However, since this heterogeneous definition of ‘brain abnormality’ reflects common practice, we included these diverging MRI findings.

Furthermore, the quality of the included studies varied. In general the majority of the studies were of good quality, although the lack of reporting of blinding of the MRI test at follow-up assessment in almost 50% of the papers is a point of concern. However, in view of the limited number of included studies, subgroup analyses by excluding low quality studies is unlikely to resolve this question, as it would merely lead to broader confidence intervals.26 As with all reviews, this systematic review is susceptible to publication bias. Especially cohort studies that did not show any predictive value of MRI have a lower chance of being published. The effect of publication bias may have resulted to overestimation of the predictive value of MRI in our meta-analysis.

Recommendations for clinical care and further research

There is solid evidence that very preterm birth and low birth weight has negative consequences on motor, neurocognitive, and behavioral functioning.1,27,28 Preterm birth is also associated with variable degrees of brain injury and reduced brain volumes.18,29 A multitude of possible confounding factors play a role in the developmental outcomes of these fragile infants. Although MRI results can
add valuable information on the prediction of long-term development, this information is in our opinion too marginal to use it on its own. A next step to consider is the performance of an Individual Patient Data (IPD) analysis gathering the results from the individual level. First, this will enhance correction of confounders of the different cohort studies. Second, this extensive data-analyses technique may be used to develop a prognostic model, in which the presence of WMA on MRI can be combined with other biomarkers known to influence long-term development such as gender, neonatal history, clinical symptoms as infection, poor nutrition, use of steroids, low birth weight, socio-demographic factors, other imaging techniques as ultrasonography, or other promising MRI techniques that might show moderate prognostic accuracy in the near future (e.g., MR spectroscopy, diffusion tensor imaging (DTI), and neurite orientation dispersion and density imaging (NODDI)). A model statistically combining various relevant prognostic factors likely increases the accuracy to predict outcomes, and may therefore be a more valuable tool for clinical use than MRI on its own.

CONCLUSIONS

This meta-analysis shows that the presence of moderate/severe WMA on MRI around term equivalent age can predict CP and motor function in very preterm or low birth weight neonates with moderate sensitivity and specificity. The ability to predict other long-term outcomes such as neurocognitive and behavioral impairments is limited. Before considering the use of this test as a standard test in clinical practice we encourage the continued use of routine MRI in a research setting to generate further evidence on its prognostic capacity together with other prognostic factors.
REFERENCES


Chapter 5

Additional file 1. Full search in Central, Medline, Embase and PsycInfo database

CENTRAL (cochrane library)
#1 mri near/3 (neonat* or a term or term)
#2 (magnetic or resonance or imaging or spectroscopy or tensor or diffusion) near/5 (neonat* or newborn* or a term or term)
#3 volumetric mr imaging or fractional anisotropy or fluid attenuated inversion recovery or flair or apparent diffusion coefficient or diffuse excessive high signal intensity or dehsi or (diffusion near/4 imaging)
#4 (#1 or #2 or #3)
#5 MeSH descriptor: [Infant, Premature] explode all trees
#6 MeSH descriptor: [Intensive Care, Neonatal] explode all trees
#7 (birth weight near/4 week*) or (gestational age near/4 32 week*) or lower gestational age
#8 preterm* or prematur* or elbw or vlbw or low birth weight* or small for date or nicu
#9 #5 or #6 or #7 or #8
#10 #4 and #9

MEDLINE
1. (birth weight adj4 week*).tw.
2. ((gestational age adj4 32 week*) or lower gestational age).tw.
3. exp infant, low birth weight/ or infant, premature/ or neonatal intensive care/
4. (preterm* or prematur* or elbw or vlbw or low birth weight* or small for date or nicu).tw.
5. or/1-4
6. (mri adj4 (neonat* or newborn* or a term or a-term or term)).tw.
7. (magnetic or resonance or imaging) adj5 (neonat* or newborn* or a term or term)).tw.
8. (volumetric mr imaging or fluid attenuated inversion recovery or flair or apparent diffusion coefficient or fractional anisotropy or diffuse excessive high signal intensity or dehsi or (diffusion adj4 imaging)).tw.
9. 6 or 7 or 8
10. (cohort or prospective or retrospective or longitudinal or prognosis or risk or case control or long term or longterm).tw.
11. exp cohort studies/ or exp prognosis/ or exp risk/ or case control studies/
12. 10 or 11
13. exp mental disorders diagnosed in childhood/
14. exp Nervous System Diseases/
15. exp mortality/
16. (corpus callosum or cerebrospinal fluid or white matter or grey matter or ((brain or cerebell*) adj2 (volum* or abnormalit* or atroph*)|)).tw.
17. (periventricular leu*omalacia or intraventricular h?emorrhage or cerebrospinal fluid or cerebellum).tw.
18. exp Cognition Disorders/ or cognition.tw.
19. (seizure* or epileps* or cerebral pals* or (learning adj3 disorder*) or deafness or blindness or (vision adj3 disorder*) or (thearing adj3 disorder*) or visuospatial memory)).tw.
20. exp Intelligence Tests/
21. exp intelligence/
22. (intelligen* or stanford-binet or wechsler or bayley scal* or iq).tw.
23. exp Education, Special/
24. (outcome or neurological sequelae).mp.
25. (mental development index or psychomotor development index or social emotional development or movement assessment or executive function or neurodevelopment* or motor impairment or cognitive impairment or language skills or language development or language delay).tw.
26. or/13-25
27. 5 and 9 and 26
28. 5 and 9 and 12
29. 27 or 28
30. animal/ not (human/ and animal/)
31. 29 not 30
32. limit 31 to yr="1980 -Current"

142
MRI for prediction of developmental outcomes in prematures

**EMBASE** (1980- and weekly alerts)
1. (birth weight adj4 week*).tw.
2. (gestational age adj4 32 week*) or lower gestational age).tw.
3. exp low birth weight/ or exp prematurity/ or neonatal intensive care.mp.
4. (preterm* or prematur* or elbw or vlbw or low birth weight* or small for date or nicu).tw.
5. or/1-4
6. (volumetric mr imaging or fractional anisotropy or fluid attenuated inversion recovery or flair or apparent diffusion coefficient or diffuse excessive high signal intensity or dehsi or (diffusion adj4 imaging)).tw.
7. (mri adj4 (neonat* or newborn* or a term or a-term or term)).tw.
8. (magnetic or resonance or imaging) adj5 (neonat* or newborn* or a term or term)).tw.
9. 6 or 7 or 8
10. (cohort or prospective or retrospective or longitudinal or prognosis or risk or case control).tw.
11. cohort analyse/ or follow up/ or prospective study/ or retrospective study/ or exp prognosis/ or exp risk/ or case control study/
12. 10 or 11
13. exp mental disease/
14. exp Neurologic disease/
15. exp mortality/
16. (seizure* or epilepsy* or cerebral pals* or white matter or grey matter or ((brain or cerebell*) adj2 (volum* or abnormality* or atrophy*))).tw.
17. (periventricular leu*omalacia or intraventricular h?emorrhage or cerebrospinal fluid or cerebellum).tw.
18. cognitive defect/ or cognition.tw.
19. ((learning adj3 disorder*) or deafness or blindness or (vision adj3 disorder*) or ((hearing adj3 disorder*) or visuospatial memory)).tw.
20. exp Intelligence Test/
21. exp intelligence/
22. (intelligence* or stanford-binet or wechsler or bayley scal* or iq).tw.
23. exp special education/
24. (outcome or neurological sequelae).mp.
25. (mental development or psychomotor development or social emotional development or movement assessment or executive function or neurodevelopment* or motor impairment or cognitive impairment or language skills or language development or language delay).tw.
26. or/13-25
27. 5 and 9 and 26
28. 5 and 9 and 12
29. 27 or 28
30. (animal/ or nonhuman/) not (human/ or ((animal/ or nonhuman/) and human/))
31. 29 not 30

**PsycInfo**
1. (birth weight adj4 week*).tw.id.
2. (gestational age adj4 32 week*) or lower gestational age).tw.
3. exp birth weight/ or premature birth/ or neonatal intensive care.mp.id.
4. (preterm* or prematur* or elbw or vlbw or low birth weight* or small for date or nicu).tw.id.
5. or/1-4
6. exp magnetic Resonance Imaging/ or magnetic resonance spectroscopy.tw.id.
7. (mri or ((magnetic or cerebell*) adj3 imaging)).tw.id.
8. 6 or 7
9. neonatal period/ or newborn*.tw.id. or neonat*.tw.id.
10. 9 and 8
11. (mri adj4 (neonat* or newborn* or a term or a-term or term)).tw.id.
12. ((magnetic or resonance or imaging or spectroscopy or tensor or diffusion) adj5 (neonat* or newborn* or a term or term)).tw.id.
13. (volumetric mr imaging or fractional anisotropy or fluid attenuated inversion recovery or flair or apparent diffusion coefficient or diffuse excessive high signal intensity or dehsi or (diffusion adj4 imaging)).tw.
14. 11 or 12 or 13
15. 10 or 14
Chapter 5

16. 5 and 15
17. (22* or 23* or 28* or 32*).cc.
18. (white matter or grey matter or cerebellum or ((brain or cerrebel*) adj2 (volum* or abnormalit* or atroph*)) or corpus callosum or periventricular leu*omalcia or intraventricular h?emorrhage or cerebrospinal fluid).mp,id.
19. 5 and 18
20. 5 and (8 or 13) and 18
21. 16 or 20
22. 17 and 21
23. limit 22 to yr="1980 -Current"

Additional file 2. Modified version of QUADAS-2 assessment tool

Scoring for methodologic quality: Risk of Bias and Applicability Judgement

Based on QUADAS-2

Rater:  
Author:  
Date of publication:  
Reference Manager Number:  

Patient Selection
1. Describe methods of patient selection

2. Describe included patients (previous testing, presentation, intended use of MRI and setting)

3. Was a consecutive or random sample of patients enrolled?  
4. Did the study avoid inappropriate exclusion?  
5. Could the selection of patients have introduced bias?  
6. Are there concerns that the included patients do not match the review question?

MRI
7. Describe how the MRI was conducted and interpreted

8. Were the MRI results interpreted without knowledge of the results of the test at follow-up (criterium test)?  
9. If a threshold was used (on MRI findings) was it prespecified?  
10. Could the conduct or interpretation of the MRI have introduced bias?  
11. Are there concerns that the MRI, its conduct or its interpretation differ from the review question?

Test at follow-up (Criterium test)
12. Describe the test at follow-up and how it was conducted and interpreted

13. Is the test at follow-up likely to correctly classify the target condition?  
14. Were the test at follow-up results interpreted without knowledge of the results of the MRI?  
15. Could the test at follow-up, its conduct, or its interpretation have introduced bias?  
16. Are there concerns that the target condition as defined by the follow-up does not match the review question?
Flow and Timing
17. Describe any patients who did not receive the MRI or test-at follow-up (loss-to-follow-up) or who were excluded from the 2x2 table (refer to flow diagram)

18. Describe the interval and any interventions between MRI and the test at follow-up

19. Did all patients receive a test at follow-up? yes/no/unclear
   a. Percentage of loss-to follow up
20. Did all patients receive the same test at follow-up? yes/no/unclear
21. Were all patients included in the analysis? yes/no/unclear
22. Could the patients flow have introduced bias? high/low/unclear

Additional file 3. Standardized data extraction form

Standardized data extraction form
Rater:
Author:
Date of publication:
Reference Manager number:

Characteristics of the study
Participants
Total number of cases Total: MRI: Follow-up:

Number of preterms <32 weeks GA male female

Number of cases by GA

Important Baseline Characteristics
- Mean GA:
- Mean birth weight:
- Asphyxia
- Sepsis/NEC
- IUGR
- Interventions
- Other:

MRI performance
Gestational age of MRI performance

Used MRI technique

Prognostic factor (ex. White matter lesions/haemorrhage)

Used Classification for prognostic factor

Used cut-off values/range

Which part of the brain is imaged?
Chapter 5

Cerebrum / Cerebellum
Study design
Prospective/Retrospective/Unclear
Single/Multi centre

Test at follow-up
Age at follow-up
Corrected?

Type of tests used
Cut-off points tests used

Which areas do these tests cover?
☐ Neurologic performance
☐ Behaviour
☐ Cognitive
☐ Somatic

Mortality numbers during follow-up?

Results of the study
Main finding (related to prognostic factor)

2x2 table

Table 1: Used cut-off points MRI………………………………………………………………………………………………………

<table>
<thead>
<tr>
<th>Used cut-off points Follow-up</th>
<th>Abnormal outcome (Follow-up)</th>
<th>Normal outcome (Follow-up)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test abnormal (MRI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test normal (MRI)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Used cut-off points MRI………………………………………………………………………………………………………

<table>
<thead>
<tr>
<th>Used cut-off points Follow-up</th>
<th>Abnormal outcome (Follow-up)</th>
<th>Normal outcome (Follow-up)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test abnormal (MRI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test normal (MRI)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EXTRA 2x2 Tables

Table 3: Used cut-off points MRI………………………………………………………………………………………………………

<table>
<thead>
<tr>
<th>Used cut-off points Follow-up</th>
<th>Abnormal outcome (Follow-up)</th>
<th>Normal outcome (Follow-up)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test abnormal (MRI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test normal (MRI)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Additional file 4. Table S1. Study details of the 20 included studies for meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Country and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Development outcome studied</th>
<th>Used Cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treyvaud et al., 2012</td>
<td>Prospective single center</td>
<td>Australia, 2001-2003</td>
<td>87/79 T1, T2, PD</td>
<td>WMA</td>
<td>Inder and Woodward(^{a,b})</td>
<td>No WMA</td>
<td>Mild or Moderate-severe WMA</td>
<td>2y corrected age</td>
<td>BSID-II MDI</td>
<td>No cut offs</td>
<td>Score &lt;70*</td>
<td>BSID-II PDI No cut offs</td>
</tr>
<tr>
<td>Kidokoro et al., 2011</td>
<td>Prospective single center</td>
<td>Australia, 2001-2003</td>
<td>81/79 T1, T2, PD</td>
<td>DEHSI ADC values and FA</td>
<td>Own classification system</td>
<td>DEHSI grade 0-4</td>
<td>DEHSI grade 1-4</td>
<td>2y corrected age</td>
<td>BSID-II MDI</td>
<td>Score &gt;70</td>
<td>Score &lt;70</td>
<td>BSID PDI No cut offs</td>
</tr>
<tr>
<td>Spittle et al., 2011</td>
<td>Prospective single center</td>
<td>Australia, 2001-2003</td>
<td>97/96 T1, T2, PD</td>
<td>WMA</td>
<td>Inder and Woodward(^{a,b})</td>
<td>2 cut offs: nil and mild WMA</td>
<td>2 cut offs: mild and moderate-severe WMA</td>
<td>5y corrected age</td>
<td>MABC</td>
<td>No signs of CP</td>
<td>Mild to severe &lt;15th centile or moderate to severe &lt;5th centile</td>
<td></td>
</tr>
<tr>
<td>Howard et al., 2011</td>
<td>Prospective single center</td>
<td>Australia, 2001-2003</td>
<td>96/91 T1, T2, PD</td>
<td>WMA</td>
<td>Inder and Woodward(^{a,b})</td>
<td>2 cut offs: nil and mild WMA</td>
<td>2 cut offs: mild and moderate-severe WMA</td>
<td>5y corrected age</td>
<td>KSEALS</td>
<td>Mild below the mean and severe &lt;2SD below the mean</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a,b}\) For specific cut-off values, see original study.
### Additional file 4. Table S1. continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Country and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Developmental outcome studied</th>
<th>Used Cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauchamp et al., 2008</td>
<td>Prospective single center</td>
<td>Australia, 2001-2003</td>
<td>82/74</td>
<td>1.5 Tesla, T1, T2</td>
<td>WMA, hippocampal volumes</td>
<td>Inder and Woodward (ab)</td>
<td>2 cut offs: mild and moderate-severe WMA</td>
<td>2 cut offs: mild and moderate-severe WMA</td>
<td>2y corrected age</td>
<td>BSID-II MDI</td>
<td>Mild between -1 and -2SD and severe &lt; -2SD</td>
<td>Delayed alternation/response task</td>
</tr>
<tr>
<td>Treyvaud et al, 2013</td>
<td>Prospective single center</td>
<td>Australia, 2001-2003</td>
<td>94/83</td>
<td>1.5 Tesla, T1, T2, PD</td>
<td>WMA, GMA, CA</td>
<td>Own classification system</td>
<td>2 cut offs: None and mild global brain abnormality</td>
<td>2 cut offs: moderate and severe global brain abnormality</td>
<td>7y corrected age</td>
<td>DAWBA</td>
<td>No psychiatric disorder</td>
<td>Psychiatric disorder</td>
</tr>
<tr>
<td>Munch et al, 2009</td>
<td>Prospective single center</td>
<td>Finland 2001-2006</td>
<td>102/80</td>
<td>0.23 Tesla and 1.5 Tesla, T1, T2</td>
<td>BA (IVH and WMA)</td>
<td>Own classification system</td>
<td>2 cut offs: normal and minor brain pathology</td>
<td>2 cut offs: minor and major brain pathology</td>
<td>2y corrected age</td>
<td>BSID-II MDI</td>
<td>Score &lt; -2SD (&gt; 70)</td>
<td>Score &gt; -2SD (&lt; 70)</td>
</tr>
<tr>
<td>Valkama et al, 2000</td>
<td>Prospective single center</td>
<td>Finland 1993-1995</td>
<td>27/24</td>
<td>1.0 Tesla, T1, T2, PD</td>
<td>BA (parenchymal lesion)</td>
<td>Own classification system</td>
<td>Parenchymal lesions not present</td>
<td>Parenchymal lesions present</td>
<td>18mo corrected age</td>
<td>Neurological exam: CP</td>
<td>No signs of CP</td>
<td>Signs of CP</td>
</tr>
</tbody>
</table>
### Additional file 4. Table S1. continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Design and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Developmental outcome studied</th>
<th>Used Cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setanen et al., 2013</td>
<td>Prospective single center Finland 2001-2006</td>
<td>122/95</td>
<td>0.23 Tesla (n=125) and 1.5 Tesla T1, T2</td>
<td>BA (MH, WMA and extracerebral space)</td>
<td>Own classification system</td>
<td>2 cut offs: normal findings or ≥1 minor pathologies</td>
<td>2 cut offs: one major pathology or several major pathologies</td>
<td>5y chronological age</td>
<td>WPPSI-R (FSIQ)</td>
<td>Normal intelligence score ≥85</td>
<td>Score 70-84 (&gt;1SD) or score &lt;70 significantly below the normal intelligence</td>
</tr>
</tbody>
</table>

- **Hearing**
  - Severe hearing impairment (amplification or hearing impairment >40dB)

- **Neurological exam:**
  - No signs of CP
  - Signs of CP

- **NDI**
  - FSIQ <85, CP, Severe hearing impairment (amplification or >40dB) or severe visual impairment (visual acuity <0.3 or blindness)
<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Country and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Developmental outcome studied</th>
<th>Used Cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gianni et al., 2007</td>
<td>Prospective single center</td>
<td>Italy 1996-2001</td>
<td>71/70</td>
<td>Conventional. Not further specified</td>
<td>BA (ventriculomegaly, cystic and noncystic PVL, focal parenchymal brain lesions)</td>
<td>No classification described</td>
<td>Normal: presence of major brain lesions</td>
<td>GMDS GQ&gt;88</td>
<td>36mo chronological age</td>
<td>Normal Abnormal: presence of major brain lesions</td>
<td>GQ &gt; 88 (1SD), Severe delay GQ &lt; 70</td>
<td></td>
</tr>
<tr>
<td>Iwata et al., 2011</td>
<td>Prospective single center</td>
<td>Japan 1995-2001</td>
<td>46/30</td>
<td>0.5 Tesla T1, T2, FLAIR</td>
<td>DEHSI, WMA on FLAIR imaging, GMA</td>
<td>Inder and Woodward abc</td>
<td>Normal grey matter score</td>
<td>WISC IQ &gt; 85</td>
<td>9y chronological age</td>
<td>Normal Abnormal grey matter score</td>
<td>Mild IQ &lt; 85 or Moderate IQ &lt; 70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neurological exam: CP</td>
<td>WISC IQ &gt; 85</td>
<td>Mild IQ &lt; 85 or Moderate IQ &lt; 70</td>
<td>Presence of hypertonicity, hyperreflexia, dystonia and spasticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Parental interview: social emotional problems</td>
<td>WISC IQ &gt; 85</td>
<td>Mild IQ &lt; 85 or Moderate IQ &lt; 70</td>
<td>Need for special assistance at school</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Additional file 4. Table S1. continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Country and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Developmental outcome studied</th>
<th>Used Cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeon et al, 2012</td>
<td>Prospective single center</td>
<td>Korea 2004-2008</td>
<td>59/67</td>
<td>3 Tesla, T1, T2, FLAIR</td>
<td>DEHSI, WMA, IVH</td>
<td>Nanba and Papile et al²</td>
<td>DEHSI: yes, WMA: cystic encephalomalacia, punctate lesions, loss of volume of white matter or corpus callosum, ventricular dilatation and myelination delay</td>
<td>DEHSI: no, WMA: no</td>
<td>18-24mo corrected age</td>
<td>BSID-II MDI Score &lt;1SD (&gt;85)</td>
<td>Mild between -1 and -2SD or severe &gt;-2SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BSID-II PDI Score &lt;1SD (&gt;85)</td>
<td>Mild between -1 or -2SD and severe &gt;-2SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neurological exam: CP</td>
<td>No signs of CP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neuro-sensory impairment (hearing and vision)</td>
<td>Normal vision and hearing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neuro-sensory impairment (hearing and vision)</td>
<td>Visual defect (requirement corrective lenses, surgery or strabismus or blindness) or hearing defect (hearing loss &gt;30dB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Country and Cohort</td>
<td>Male/Female Ratio</td>
<td>MRI technique</td>
<td>Prognostic factor</td>
<td>Classification system</td>
<td>Defined Normal</td>
<td>Defined abnormal</td>
<td>Follow-up interval</td>
<td>Developmental outcome studied</td>
<td>Used Cut-off value for normal</td>
<td>Abnormal findings</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------</td>
<td>-------------------------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Clarc et al., 2010</td>
<td>Prospective single center</td>
<td>New Zealand 1998-2000</td>
<td>55/52</td>
<td>1.5 Tesla, T1, T2, PD</td>
<td>WMA, GMA, brain volumes</td>
<td>Woodward et al.</td>
<td>WMA and GMA 2 cut offs: nil and mild</td>
<td>2 cut offs: mild and severe WMA, mild and severe GMA</td>
<td>6y corrected age</td>
<td>WPPSI-R verbal memory</td>
<td>Passing the memory task</td>
<td>Failing to pass the task</td>
</tr>
<tr>
<td>Woodward et al., 2006</td>
<td>Prospective multi center</td>
<td>New Zealand 1998-2000 &amp; Australia 2001-2002</td>
<td>78/88</td>
<td>1.5 Tesla, T1, T2, PD</td>
<td>WMA, GMA</td>
<td>Inder et al.</td>
<td>2 cut offs: nil and mild WMA and normal GMA</td>
<td>2 cut offs: mild and moderate-severe WMA and abnormal GMA</td>
<td>2y corrected age</td>
<td>WPPSI-R visuospatial memory</td>
<td>Passing the memory task</td>
<td>Failing to pass the task</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BSID-II MDI Score &lt;1SD (&gt;85)</td>
<td>BSID-II PDI Score &lt;1SD (&gt;85)</td>
<td>Mild between -1 and -2SD or severe &gt;-2SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neurological exam: CP</td>
<td>No signs of CP</td>
<td>Signs of CP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neurosensory impairment (hearing and vision)</td>
<td>Normal vision and hearing</td>
<td>Visual defect (requirement corrective lenses, surgery or strabismus or blindness) or hearing defect (hearing loss &gt;30dB)</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Country and Cohort</td>
<td>Male/Female Ratio</td>
<td>MRI technique</td>
<td>Prognostic factor</td>
<td>Classification system</td>
<td>Defined Normal</td>
<td>Defined abnormal</td>
<td>Follow-up interval</td>
<td>Developmental outcome studied</td>
<td>Used Cut-off value for normal</td>
<td>Abnormal findings</td>
</tr>
</tbody>
</table>
|-----------------------|-----------------------------|-----------------------------|-------------------|---------------|-------------------|----------------------|----------------|-------------------|---------------------|-----------------------------|-------------------------------|-----------------
| Woodward et al., 2012 | Prospective single center   | New Zealand 1998-2000       | 54/52             | 1.5 Tesla, T1, T2, PD | WMA             | Woodward et al<sup>b</sup> | 2 cut offs: nil and mild WMA | 2 cut offs: mild and moderate-severe WMA | 4y corrected age | WPPSI-R mild delay >-1SD or severe delay > 2SD | Mild delay >-1SD or severe delay > 2SD |
| Skiöld et al, 2012    | Prospective single center   | Sweden, 2004-2007           | 48/43             | 1.5 Tesla, T1, T2, PD | WMA, DEHSI      | Inder et al<sup>a</sup> | Normal and mild WMA, DEHSI not defined | Moderate-severe WMA, DEHSI not defined | 30mo corrected age | BSID-III MDI >-2SD (>70) | BSID-III MDI >-2SD (>70) |

<sup>a</sup> Normal and mild WMA, DEHSI not defined
<sup>b</sup> Normal and mild WMA, DEHSI not defined; WMA, WMA.
<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Country and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Developmental outcome studied</th>
<th>Used Cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>de Bruine et al., 2011</td>
<td>Prospective single center</td>
<td>The Netherlands 2006-2007</td>
<td>68/42</td>
<td>3 Tesla, T1, T2</td>
<td>DEHSI, Punctate WM lesions and ventricular dilatation</td>
<td>Miller et al* DEHSI:no, WM lesions normal: ≤6 lesions, Normal ventricles: &lt;12mm</td>
<td>DEHSI:yes, WM lesions abnormal: &gt;6 lesions, Moderate dilatation ventrile 12-16mm, Severe &gt;16mm</td>
<td>2y corrected age</td>
<td>BSID-II MDI Score &gt;-2SD (&gt;70)</td>
<td>BSID-III Social Emotional Neurological exam: CP Normal or unspecific sings Abnormal (sings of CP)</td>
<td>Mild between 1 and -2SD or severe &gt;-2SD</td>
<td></td>
</tr>
<tr>
<td>Rose et al., 2009</td>
<td>Retrospective single center</td>
<td>USA 1999-2001</td>
<td>41/37</td>
<td>1.5 Tesla, T1, T2, FLAIR, DTI</td>
<td>BA (incl. ventriculomegaly and/or parenchymal abnormality), DWI</td>
<td>Own classification system and ADC values continuous measures of ADC values</td>
<td>18-22mo corrected age</td>
<td>BSID-II MDI Score &gt;-2SD (&lt;70)</td>
<td>BSID-II PDI Score &lt;1SD (&gt;85)</td>
<td>GMFCS Neurological exam: CP</td>
<td>GMFCS score of 1-2, Severe CP score of 4-5 Moderate CP score 2-3, Severe CP score &gt;2-5</td>
<td>Mild between -1 and -2SD or severe &gt;-2SD</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Country and Cohort</td>
<td>Male/Female Ratio</td>
<td>MRI technique</td>
<td>Prognostic factor</td>
<td>Classification system</td>
<td>Defined Normal</td>
<td>Defined abnormal</td>
<td>Follow-up interval</td>
<td>Developmental outcome studied</td>
<td>Used Cut-off value for normal</td>
<td>Abnormal findings</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>------------------</td>
<td>------------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Mirmiran et al., 2004</td>
<td>Prospective single center</td>
<td>USA 1996-1999</td>
<td>31/30</td>
<td>1.5 Tesla, T1, T2, PD</td>
<td>WMA, hemorrhages, mineralisation and ventricular size</td>
<td>own classification system</td>
<td>normal score of C0 and C1</td>
<td>abnormal score of C2 and C3</td>
<td>21-31 mo corrected age</td>
<td>Neurological exam: CP</td>
<td>No CP</td>
<td>mild, moderate and severe CP</td>
</tr>
<tr>
<td>Augustine et al., 2008</td>
<td>Retrospective single center</td>
<td>USA 2001-2003</td>
<td>17/19</td>
<td>1.5 Tesla, T1, T2, FLAIR, MRS</td>
<td>BA (hemorrhage or mineralization, ventriculomegaly, parenchymal abnormality)</td>
<td>Own classification system</td>
<td>normal score of C0 and C1</td>
<td>abnormal score of C2 and C3</td>
<td>18-24 mo corrected age</td>
<td>BSID-II MDI</td>
<td>Score &lt;1SD (&gt;85)</td>
<td>Score &gt;1SD (&lt;85)</td>
</tr>
</tbody>
</table>
### Additional file 4. Table S1. continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Country and Cohort</th>
<th>Male/Female Ratio</th>
<th>MRI technique</th>
<th>Prognostic factor</th>
<th>Classification system</th>
<th>Defined Normal</th>
<th>Defined abnormal</th>
<th>Follow-up interval</th>
<th>Developmental outcome studied</th>
<th>Used cut-off value for normal</th>
<th>Abnormal findings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BSID-II PDI Score &lt;1SD (&lt;85)</td>
<td>Score &gt;1SD (&gt;85)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neurological exam: CP abnormal muscle tone/movement in at least 1 extremity</td>
<td>No CP</td>
<td>Signs of CP</td>
</tr>
</tbody>
</table>

*Data derived after contact with author.

ADC, apparent diffusion coefficient; BA, brain abnormality; BSID-II, Bayley scales of infant development; CA, cerebellar abnormalities; CP, cerebral palsy; CELF-P, clinical evaluation of language fundamental; CVLT-C, California verbal learning test, children's version; DAWBA, developmental and well-being assessment; DEHSI, diffuse excessive high signal intensity; DTI, diffusion tensor imaging; FA, fractional anisotropy; FLAIR, fluid attenuated inversion recovery imaging; FSIQ, full scale intelligence quotient; GAMA, gray matter abnormality; GMDS, Griffiths mental developmental scales; GMFCS, gross motor function classification system; GQ, general quotient; IVH, intraventricular haemorrhage; ITSEA, infant-toddler social and emotional assessment; KSEALS, kaufman survey of early academic and language skills; MABC, motor assessment battery for children; MDI, mental development index; MRS, magnetic resonance spectroscopy; MSML, multisearch multilocation search task; NDI, neurodevelopmental impairment; NEPSY II: neuropsychological assessment; PD, proton density; PDI, psychomotor development index; PVL, periventricular leukomalacia; SDQ, strengths and difficulties questionnaire; TBV, total brain volume; WISC: wechsler intelligence scale for children; WJ-III, woodcock johnson-III test tests of achievement; WMA, white matter abnormality; WPPSI, wechsler preschool and primary scale of intelligence; WPPSI-R: wechsler preschool and primary scale of intelligence-revised;
MRI for prediction of developmental outcomes in premature infants


