Predictive value of the Bayley Scales of Infant Development on development of very preterm/very low birth weight children: a meta-analysis
Luttikhuizen Dos Santos ES, de Kieviet JF, Konigs M, van Elburg RM, Oosterlaan J

CRD summary
The authors concluded that the predictive value of the Bayley scales of Infant Development for later development of very preterm and very low birth weight infants was limited. The evidence was based on correlations which can lead to overstated results so the results may not be reliable but the authors' conclusions reflect the limited evidence and seem appropriate.

Authors' objectives
To assess the predictive value of the Bayley Scales of Infant Development for the neurodevelopmental progress of very preterm and very low birth weight infants in the first three years of life.

Searching
The Cochrane Library, PubMed, PsycINFO and CINAHL were searched for peer-reviewed studies published in English before March 2013. Search terms were reported. Reference lists of selected studies were searched manually.

Study selection
Eligible studies assessed the predictive value of Bayley Scales for Infant Development (I, II, or III) on later functioning in children born very preterm (gestational age ≤32 weeks) and/or very low birth weight (≤1,500g). Bayley Scales I and II include the Mental Developmental Index and Psychomotor Developmental Index; the Bayley II scale comprises five scales including a cognitive, language and motor composite. Eligible studies had to report or provide sufficient outcome data to calculate statistical outcomes and had to compare with assessment later in life using another standardised test that assessed any aspect of development.

The mean age of included infants at first assessment ranged from four to 36 months; mean age at follow-up ranged from three to 14 years. Where reported, mean birth weight ranged between 811g and 1,221g. Four studies included pre-term infants; other studies were of low birth weight infants. Some studies included children with cerebral palsy. Included studies measured cognitive, motor and language functioning in infants. Various different measures were used at follow-up for the different function outcomes (as reported in the review).

The authors did not state how many reviewers screened studies for inclusion.

Assessment of study quality
Two reviewers independently assessed the quality of the included studies using the Newcastle-Ottawa Scale. Studies could score between zero and 9 points (higher scores indicated better quality). Any discrepancies between reviewers were resolved by consensus.

Data extraction
Outcome data (including Pearson's correlation coefficients, Spearman's correlation coefficients, odds ratios and variance) were extracted from individual studies. Where studies reported multiple time point measures, the mean of these correlations was calculated. Sensitivity and specificity were extracted where these results were reported.

The authors did not state how many reviewers extracted data.

Methods of synthesis
Where four or more studies reported a predictive outcome, a fixed-effect (or random-effects model, where there was evidence of statistical heterogeneity) was used to pool correlations (r) and their 95% confidence intervals. The magnitude of effect sizes was reported according to Cohen's definitions (0.10 indicated small, 0.20 medium and 0.50 large correlations). Where fewer than four studies reported on outcomes, findings were described narratively. Results were reported separately for cognitive, motor and language function and by developmental index.
Statistical heterogeneity was assessed using Cochran's Q statistic. Meta-regression was performed to assess potential moderating effects of study characteristics (including age at assessment, age at follow-up, interval between assessments, data of birth, gender and study quality) where meta-analysis contained at least 10 studies. Sensitivity analyses were performed to assess the effect of different versions of the Bayley Scales and the inclusion of children with cerebral palsy on effect sizes.

Publication bias was assessed using funnel plots, Rosenthal's fail-safe N and Egger's test.

**Results of the review**

Sixteen studies (17 study arms; 1,792 infants) were included in the review; one study reported results separately for boys and girls. Fifteen studies used a prospective design and one used a retrospective design. Studies scored between six and eight out of nine on quality.

**Cognitive function:** There was a large positive relationship between the Mental Developmental Index and cognitive function in the first three years of life in infants ($r = 0.61$, 95% CI 0.57 to 0.64; 14 studies). There was no evidence of statistical heterogeneity. There was no evidence of publication bias. Meta-regression analysis did not identify any significant moderating effects of study characteristics. Two studies reported sensitivity (37% and 91%) and specificity (97% and 71%) at the two standard deviation threshold. Results for one standard deviation were also reported.

Findings were mixed for the relationship between the Bayley motor scale and later cognitive function (five study arms; results were reported in the review).

**Motor function:** There was a medium positive relationship between the Bayley motor scale and later motor function ($r = 0.34$, 95% CI 0.26 to 0.42; five studies) with no evidence of statistical heterogeneity. Removing one study that included children with cerebral palsy did not significantly alter the findings. Two studies reported sensitivity (21% and 80%) and specificity (100% in both) at the two standard deviation threshold. Results for one standard deviation were also reported.

Findings were mixed for the relationship between Motor Developmental Index and later motor function (three studies; results were reported in the review).

**Language function:** The relationships between Motor Developmental Index (three study arms) or the Bayley motor scale and later language function both showed mixed findings (as reported in the review).

There was no evidence of publication bias for cognitive or motor function.

**Authors' conclusions**

The predictive value of the Bayley scales of Infant Development for later development of very preterm and very low birth weight infants is limited. Further research is needed to enhance the prediction of developmental delay.

**CRD commentary**

The review question was clearly stated and was supported by appropriate inclusion criteria. Electronic databases were searched for relevant data. The search was restricted by language and publication status so potentially relevant data may have been missed. Study quality was assessed; overall scores were reported for individual studies but were not discussed further. Studies were reported to be of moderate to high quality. Study quality was performed in duplicate but it was unclear whether this was also the case for study selection and data extraction so reviewer error and bias could not be ruled out.

Studies were synthesised using a meta-analysis of correlation coefficients which are not generally considered to be good summaries of predictive value and so the results may be misleading. There was heterogeneity in the follow-up measures, which the authors acknowledged and suggested may have reduced the predictive value of the Bayley scales. The authors went some way to account for and investigate differences between studies using appropriate statistical methods. The authors acknowledged some further limitations of the evidence including small sample sizes and the small number of studies that assessed motor and language function. They also acknowledged the limited reporting on sensitivity and specificity, which would have been more informative than correlation statistics. The results that were reported for sensitivity and specificity suggested limited predictive value of the Bayley scales.
There was potential for bias in the review process and the evidence base was limited, particularly as the main findings were based on correlations which can lead to overstated results. The results may not be reliable but the authors’ conclusions reflect the limited evidence and seem appropriate.

**Implications of the review for practice and research**

**Practice:** The authors stated that clinicians and researchers using the Bayley Scale of Infant Development should be aware of the restricted predictive value of using this one behavioural assessment scale.

**Research:** The authors stated that more research was needed to assess to what extent multiple known risk factors for development enhance the ability to make predictions about the developmental course of very preterm and very low birth weight children. The authors recommended that future research should report on sensitivity and specificity.

**Funding**
Not stated.

**Bibliographic details**

**PubMedID**
23597678

**DOI**
10.1016/j.earlhumdev.2013.03.008

**Original Paper URL**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Child Development; Female; Humans; Infant, Extremely Premature /physiology; Infant, Newborn; Infant, Very Low Birth Weight /physiology; Male; Predictive Value of Tests

**AccessionNumber**
12013022660

**Date bibliographic record published**
07/05/2013

**Date abstract record published**
24/09/2013

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.