Risk factor models for neurodevelopmental outcome in children born very preterm or with very low birth weight: a systematic review of methodology and reporting

Short title: Review of risk factor models for VPT/VLBW children

Authors:

Louise Linsell[1], BSc MSc

Reem Malouf[1], MSc MD

Joan Morris[2], MA MSc PhD

Jennifer J Kurinczuk[1], BSc MBChB MSc MD

Neil Marlow[3], BA MBBS MRCP MD

Address correspondence to (including all queries and proofs):

Louise Linsell (Dr. is not an appropriate honorific)
National Perinatal Epidemiology Unit (NPEU)
Nuffield Department of Population Health,
University of Oxford
Old Road Campus
Headington
Oxford
OX3 7LF
United Kingdom

Email: louise.linsell@npeu.ox.ac.uk
Tel: +44 (0)1865 617922
Fax: +44 (0)1865 289701

Abstract

The prediction of long term outcome in surviving infants born very preterm (VPT) or with very low birth weight (VLBW) is necessary to guide clinical management, provide information to parents and to help target and evaluate interventions. A large literature reports risk factor models for neurodevelopmental outcome in VPT/VLBW children yet few, if any, have been developed for use in routine clinical practice or adopted for use in research studies or policy evaluation. We sought to systematically review the methods and reporting of studies that have developed a multivariable risk factor model for neurodevelopment in surviving VPT/VLBW children. We searched the Medline, Embase and Pyscinfo databases from 01/01/1990 to 01/06/2014 and identified 78 studies reporting 222 risk factor models (protocol registration number CRD42014006943). Most studies presented risk factor analyses that were not intended to be used for prediction, confirming that there is a dearth of specifically designed prognostic modelling studies for long-term outcome in surviving VPT/VLBW children. We highlight the strengths and weaknesses of the research methodology and reporting to date, and provide recommendations for the design and analysis of future studies seeking to study risk prediction or to develop prognostic models for VPT/VLBW children.

Key words: Data reporting, development, preterm infants, prognosis, research methodology, risk factors, systematic review, very low birth weight.

Abbreviations

BW	birth weight
GA	gestational age
VLBW	very low birth weight
VPT	very preterm

Prematurity and its associated neonatal morbidities have a pervasive effect on neurodevelopment, leading to: cerebral palsy, visual and auditory deficits, impairments in motor and cognitive function and behavioral problems.(1) The early identification of factors that mediate long term outcome is necessary to guide the clinical management of children born preterm, provide information to parents and to help develop, target and evaluate interventions. A large literature reports risk factor models for neurodevelopmental outcome in very preterm children yet few, if any, have been developed for use in routine clinical practice or adopted for use in research studies or policy evaluation. The aim of this article is to describe and review the conduct and reporting of these studies over the last two decades, and provide methodological recommendations for future research in this area.

We reviewed and summarised the methods and reporting in 78 studies that were recently included in a systematic review of risk predictors for neurodevelopment outcome in surviving children born ≤32 weeks gestation or with birth weight ≤1,250g.(2-4) Studies were included if the aim (or one of the aims) was to perform a multivariable (>2 variables) risk factor analysis for neurodevelopmental outcome in this population, in one or more of the following domains: motor, cognition, hearing, vision or behavior. This article reports the main elements of study design, model development, reporting and validation (if performed) of the 78 studies included in the review. The findings are then discussed within the framework of recommended approaches for risk factor analysis and reporting advocated by experts advising on prediction modelling in the medical and statistical literature. Despite the recent publication of reporting guidelines for studies developing and validating risk prediction models,(5) there is no equivalent central resource that provides guidance for the

design and analysis of studies seeking to perform a risk factor analysis or develop a prognostic model.

METHODS

The methods for the systematic review of risk predictors for poor neurodevelopment in very preterm (VPT) and very low birth weight (VLBW) children have previously been published in a protocol (http://www.crd.york.ac.uk/PROSPERO/), registration number CRD42014006943, and in three review articles published for motor, cognitive and behavioral outcomes.(2-4) They are described in brief below.

Search strategy

Three electronic search strategies were devised in the Medline, Embase and Psycinfo databases (available with the protocol) using the National Institutes of Health Medical Subject Headings (Web Appendix 1). The searches identified journal articles published from 01/01/1990 to 01/06/2014 reporting a multivariable risk prediction model for a neurodevelopmental outcome assessed after the age of 18 months in VPT/VLBW children. No language restrictions were made. The bibliographies of all articles included for data extraction were hand-searched for further eligible articles.

Eligibility criteria

Articles were included in the review if they satisfied the following eligibility criteria: (1) contained original data, (2) study population was born after 01/01/1990, (3) study population was ≤32 weeks gestational age (GA) or with birth weight (BW) ≤1,250g and not a highly select group (based on other clinical criteria), and (4) one objective was to perform a

multivariable risk factor analysis (>2 variables) of a neurodevelopmental outcome in survivors assessed after 18 months of age.

A BW cut-off of ≤1250g was used, instead of the official definition of VLBW (≤1500g), to exclude the subset of more mature but extremely growth restricted children included in the typical ≤1500g VLBW cohort which can cause heterogeneity and lead to confounding bias when examining the relationship between risk factors and outcome.(6) Explanatory prognostic factor studies which investigate the causal pathway between a single prognostic factor and an outcome (ideally adjusted for confounders) and estimate effect size are not included in the review. In these types of study, other risk factors are included based on the change in the regression coefficient of the prognostic factor under study, whereas in multivariable outcome prediction models risk factors are included in the model based on their predictive ability of the outcome. Current guidelines recommend not combining these two distinct types of study as their objectives and model building strategies differ which, when synthesised, could lead to biased results.(7, 8)

Data extraction and reporting

All articles identified by the search strategies were screened on title and abstract for definite exclusions and duplicates (screen 1). For the remaining articles, the full text was retrieved and the eligibility criteria were applied (screen 2). The two screens were initially performed by the first author (LL), but if there was uncertainty about the eligibility of an article, it was screened independently by the second author (RM). If a decision could not be reached it was referred to the rest of the author review team (JK, NM and JM). Non-English articles included in the review were fully translated. Multiple articles based on the same

cohort of children underwent a panel review (LL, RM and NM). Those reporting the same outcome domain (cognitive, motor, behavior, hearing, vision) at the same age of assessment (<5 years and ≥5 years) were assessed on relevance to the review, and only one article was selected for data extraction. For all articles eligible for inclusion, both reviewers (LL and RM) independently completed a full data extraction form on a customised MS Access 2010 database. These were cross-validated for discrepancies, and referred to the rest of the author review team if agreement could not be reached. If critical information was missing or unclear the corresponding author was contacted once by email for clarification.

Summary statistics were provided for each item of data extracted (listed in Tables 1-3) and the results from this review are presented in accordance with the PRISMA guidelines (Web Appendix 2).(9) If more than one model was presented in an article, the first model reported was selected to summarise model characteristics to avoid the over-representation of studies presenting multiple models.

RESULTS

The search strategy retrieved 44,500 articles (Figure 1), and 2,284 articles were screened on full text, applying the full set of eligibility criteria. Ninety one articles (from 48 cohort populations) containing multivariable risk factor analyses were eligible for inclusion. Following panel review (LL, RM and NM), a further 13 articles were excluded as they reported the same outcome domain at the same age of assessment in the same cohort as another article with a more relevant objective. The remaining 78 articles, reporting 222 risk prediction model for a neurodevelopmental outcome in surviving VPT/VLBW children, were included in the data extraction.(10-87)

Study design

The design characteristics of the 78 studies that were included for data extraction are shown in Table 1. The main study design was cohort (91%) and there were five randomized controlled trial populations,(51, 52, 55, 60, 69) one case-control study(53) and one cross-sectional study.(31) Of the 71 cohorts followed up prospectively, half (n=36) were ascertained from all live births in a geographically defined region and half (n=35) were recruited from neonatal intensive care unit admissions. Forty-five percent (n=35) of studies recruited from a single center and 79% (n=62) of studies extracted risk factor and outcome data prospectively. Overall, 32% (n=25) of studies were based on a cohort study of all live births in geographically defined region and followed up prospectively.

Fifty-five percent of (n=43) studies defined the study cohort using GA only and 21% (n=16) used BW only, the remaining studies used some combination of the two. The most common age of assessment was 18-24 months (45%, n=34) and no studies followed up beyond the age of 12 years. The median sample size was 219 {IQR: 141 to 461} and 11 studies had more than 1,000 participants.(11, 12, 14, 16-18, 24, 25, 54, 69, 82) Only one study mentioned the issue of power and referred to the number of events per variable in the modelling process.(39)

Model development

The median number of risk factor models presented per study was two and 25% of studies presented four or more models. A summary of the model building techniques used is presented in Table 2 (for the first model presented in the article). The median number of candidate risk factors considered at the outset was 17 [range: 3 to 51]. Seventy two percent

(n=56) of studies provided a rationale for their choice of candidate factors, or used a comprehensive list with wide coverage (≥20 factors), but 28% (n=22) of studies with less than 20 candidate factors gave no rationale at all. The derivation and coding of outcomes and risk factors were described clearly in 85% (n=66) of studies and outcomes were generally measured using comprehensive, well-validated tests or assessed using standardized diagnostic criteria, though blinding to previous medical history at assessment was frequently not reported. Sixty-nine percent (n=54) of studies categorised some or all of the risk factors or that were measured on a continuous scale. In many cases there was a clinical rationale, but the use of arbitrary thresholds for no reason was also widespread practice.

Forty-three percent (n=33) of studies entered all of the candidate variables into the multivariable model, 28% (n=22) only included candidate factors with a *P*-value below a set threshold in a univariable test of association with outcome and 8% (n=6) screened variables using multivariable analysis. The most popular method of model building after initial screening (or no screening) was to simply include all factors (39%, n=30) regardless of statistical significance, and stepwise selection (35%, n=27). The median number of risk factors included in the final prediction model was 5 {IQR: 4 to 9}.

Reporting and model validation

A summary of reporting and model validation (if performed) is presented in Table 3. Generally the reporting of study attrition and missing data was quite poor, making it difficult to assess how representative the children analysed were of the original study sample recruited. A comparison of baseline characteristics between those lost to follow-up and

those assessed was not performed by 30% (n=21) studies. Furthermore only 25% (n=19) of studies presented the amount of missing data for each variable included in the final model and only 31% (n=24) reported the number of infants included in the final model. Although 52 (67%) studies reported the point estimates, confidence interval or standard errors, and the *P*-value or significance level for all model coefficients included in the final multivariable model, the remaining studies only did so for selected variables or failed to report them at all.

Four studies(12, 32, 42, 72) assessed the performance of the final model using statistical validation techniques. All four reported the area under the receiver operating curve which measures the discriminatory ability of the model to differentiate between individuals by severity of outcome. One study(42) produced a decision tree/algorithm based on their model and another(12) produced a web based tool to predict either death or neurodevelopmental impairment for use in clinical practice (https://neonatal.rti.org/OTEstimator/). In this study the model was developed by randomly splitting the dataset into a development set (70%) and tested the model in a validation set (30%).

DISCUSSION

This review has highlighted some strengths and weakness in the research methodology and reporting of multivariable risk factor models for neurodevelopmental outcome in the VPT/VLBW population in the published literature. Recommendations for future studies, based on the latest literature in the field, are discussed in relation to the findings of the

review. These guidelines are intended as an aid for researchers planning risk factor studies, and are not intended as a prescribed checklist.

Study design

Selection of participants. The optimal study design for prognostic research in the VPT/VLBW population is a prospective cohort of all live births in all centers in a geographically defined region. The strength of this design is that the sample represents the source population and exposure to risk factors can be measured prior to the occurrence of an outcome.(88) Studies based in a single center, while convenient, have less generalizability due to center differences in service provision, referral and management practices and the socio-economic/ethnic characteristics of the local population. Also, neonatal intensive care unit populations are less representative as they tend to have a different case-mix compared with the general VPT/VLBW population, containing a higher proportion of infants with poorer outcomes due to referrals of higher-risk infants from lower-level centers. Data from randomized controlled trials can be used to study prognosis, provided the treatment allocation is included as a predictor variable, though such studies may have reduced generalizability because of restricted trial eligibility criteria.

There was a lack of consistency in the GA and BW criteria used nationally and internationally to define cohorts, with some studies using prematurity or BW alone and others using a combination of both. Before the routine use of ultrasound, cohorts were generally defined by BW due to the unreliable measurement of GA. Although there has been a shift from using BW defined cohorts in the last two decades, even studies conducted more recently used varying criteria. The typical VLBW cohort can be fairly heterogeneous group due to the

inclusion of the more mature but extremely growth restricted infants, and it is recommended that epidemiologic studies of immature newborns should be based on GA rather than BW criterion.(6)

Sample size and events per variable. It is generally recommended that there should be a minimum of 10 outcome events per predictor variable when using logistic regression models for risk factor analysis, and models with less than five events per variable should be interpreted with caution.(89) As the number of events per variable declines below 10, simulation studies have shown an increase in bias and variability, unreliable confidence interval coverage and problems with model convergence.(90) Overfitting is a particular problem when small samples are used for risk prediction modelling, which can lead to the performance of the model being overoptimistic in the dataset from which it was developed.(91) The median sample size of studies included in the review was 219 and the median number of risk factors in the final models was five, which means that, on average, studies lacked the power to detect associations in any outcome with an event rate of less than 20%. It is difficult to calculate a suitable sample size for risk prediction studies, and risk factor analysis is often a secondary aim when planning a cohort study, however, it is advisable not stray too far below the recommended number of events per variable.

Model development

Definitions of outcomes and risk factors. The measurement and assessment of outcome were generally clearly defined and robust in the studies included in this review, though it would be helpful to have more international agreement and standardization, for example in the diagnostic criteria used by studies to define cerebral palsy, which are available.(92)

There was more consistency across studies in the tests used to measure motor and cognitive outcomes than those used in the other outcome domains, however studies did not always use the same cut-points for impairment and some used continuous scores which made it difficult to compare findings. Few studies reported whether assessments were blinded to previous medical history. For risk factor variables, some conditions such as bronchopulmonary dysplasia and intraventricular haemorrhage were defined quite consistently across studies, whereas other factors such as sepsis and necrotizing enterocolitis varied in definition or were not clearly defined at all. It is recommended that outcomes should be assessed prospectively using comprehensive, well-validated tests or using standard diagnostic criteria with a strict protocol, blinded to previous medical history. The definitions of outcomes and risk factors should be described clearly in sufficient detail if the model is to be correctly interpreted or applied by other researchers.

Coding and modelling of continuous variables. The factors retained in the final model, and the value of their coefficients are strongly influenced by the coding and methods used to model continuous and categorical variables.(93) Many studies included in the review categorised some or all of the continuous risk predictors used in the modelling, often without a clear rationale. The arbitrary categorization of continuous predictors should be avoided as this results in a loss of information and statistical power. It also results in the classification of individuals who are close but on either side of the chosen cut-point as having very different levels of risk.(94) Using cut-points that are data-driven, such as the sample median, are problematic if a model is transported to a different study population.(95) Assessing whether the relationship between a continuous predictor and an outcome is linear or nonlinear is important; yet few studies reported doing this. If a

nonlinear relationship exists, then categorization may be a reasonable option or fitting a nonlinear term, for example a quadratic term. The use of splines and fractional polynomials could be also be considered in larger samples.(93)

Selection of risk predictors. There is no overall consensus on the best strategy to select variables for inclusion in a risk factor model, however some approaches are not recommended. This includes screening candidate factors using univariable tests of association with outcome, as the correlation with other risk factors is not controlled for. This can result in the rejection of important risk factors that only become predictive after adjustment for other factors.(90, 96) Nearly half the studies included all candidate variables, but many started with fewer than 20 variables at the outset, so it is likely that some preselection process was applied but not reported. It is important to report any procedure used to reduce the number of candidate variables in sufficient detail so that the degree of coverage (of factors considered) can be assessed.

Automated variable selection processes, such as forward selection, backward elimination or stepwise approaches, are cautioned against as they are data-driven and ignore clinical plausibility. This can lead to poorly performing models with biased regression coefficients.(91, 97, 98) Some important risk factors that are well-established in the literature may not always be statistically significant in a particular dataset, but it is advisable to include these in the development process. The preferred approach is to start with the full model and eliminate factors (taking both statistical significance and clinical relevance into consideration) as this avoids overfitting and selection bias and provides correct estimates of standard error.(99) The level of significance has a major effect on the number of variables

retained; a 1% level will almost always result in a more parsimonious model than a 5% level. However starting with the full model can be problematic when there are a large number of candidate factors, as is usually the case when predicting neurodevelopmental outcome in the VPT/VLBW population, due to the vast amount of data usually collected during the neonatal period and following discharge. Six studies(12, 30, 45, 47, 52, 85) in the review adopted a sequential multivariable approach to prediction, fitting candidate factors in stages according to the time frame in which they occurred or according to themes, such as clinical and socio-demographic. This approach seems reasonable, given that prognosis in these children is a complex and dynamic process, with environmental factors potentially superseding the influence of early biological factors as the child grows up.

Reporting and model validation

Attrition and missing data. One area that could easily be improved is the reporting of study ascertainment and attrition. While this was excellent in some studies, it was lacking in many others which made it difficult to determine how representative the study populations were. In prospective cohorts where children are followed up at multiple time points, the numbers and reasons for exclusions and dropouts at each stage should be clearly reported in a flow diagram. In retrospective studies, the number of infants assessed for inclusion should be reported, in addition to the number selected for inclusion, which is important for assessing the risk of selection bias. There is evidence of selective drop-out in studies of preterm children, with those lost to follow-up more likely to be severely impaired or have mothers with a lower level of education,(100-102) therefore it is important to report and assess the potential impact of drop-out on the results. While the majority of studies reported that a comparison of children lost versus not lost to follow-up had been conducted, the data

relating to this was not always fully reported in the article (or provided as supplemental material). If further participants are excluded from the final model due to missing data, then this number should be reported and the representativeness of analysis population should be assessed.

Model reporting and validation. Some studies only reported the results for selected variables in the final model, or only provided *P*-values or the regression coefficients with no confidence intervals. The absolute minimum that should be reported are the regression coefficients with confidence intervals for all factors retained in the final model, but it is also helpful to report the results of important intermediate stages of model development (as supplemental material if space is not permitted). If a study wishes to go further and develop a prognostic model, then its performance can be assessed for calibration (comparing observed and predicted outcomes for groups), accuracy (comparing the observed and predicted outcome for individuals) and discrimination (ability to distinguish between individuals at low and high risk of developing an outcome). These tests can be carried out on the data used to derive the model, but should ideally be carried out on new data, either from the same source (internal validation) or ideally on data from an independent source (external validation) to evaluate the transportability of the model. Methods of internal validation include split-sample (splitting the sample randomly into a "training" set for model development and a "test" set for model validation), cross-validation (developing a model in each set and testing it on the other set) and bootstrapping (resampling with replacement). The latter two resampling methods are more effective than the split-sample approach, which is inefficient and results in a loss of power, and therefore are the preferred approach for internal validation.(88, 103)

Strengths and limitations of the review

The search filter used in this review was intentionally broad at the expense of precision in order to capture all studies reporting risk factor analyses, which resulted in a large number of articles retrieved. This approach is recommended for reviews in fields in which clinical prediction models are largely underdeveloped, rather than a more specific search filter, which would have had a high false negative rate, potentially leading to many articles being missed.(104) No language restrictions were imposed and no further articles were identified in the hand-search of bibliographies of all studies included, so it is unlikely that there were any major omissions. Some studies in the review were published before the proliferation of comprehensive guidelines on the conduct and reporting of research studies (http://www.equator-network.org/), and may not reflect the standards of more current work. Furthermore, many of them preceded the publication of the large literature that now exists on risk prediction modelling. Specific reporting guidelines for this field have only just emerged.(5)

Conclusion

This systematic review of 78 published articles reporting multivariable risk factor models for neurodevelopmental outcome in surviving VPT/VLBW children, has revealed some shortcomings in methodology and reporting that could be improved in future studies, and confirmed that that there is a dearth of properly designed and well-conducted prognostic modelling studies in this field. Modelling long-term outcome in such a heterogeneous population is challenging; often with the existence of multiple impairments within the same individual and with multiple risk factors acting sequentially over time. In the three published

reviews of risk factors for cognitive and motor impairment and behavioural problems that were based on these studies, the evidence for most risk factors was mixed or unclear.(2-4) This may be due to the difficulty of modelling prognosis in this population, but also due to differences in study design, study population, methodological quality and lack of standardization of measures. The findings and recommendations of this critical review should be used as a basis for the design and analysis of future studies seeking to develop multivariate risk factor or prognostic models in this population.

ACKNOWLEDGEMENT SECTION

Author affiliations

[1] National Perinatal Epidemiology Unit (NPEU), Nuffield Department of Population Health, University of Oxford, Old Road Campus, Headington, Oxford, OX3 7LF, United Kingdom.

[2] Queen Mary University of London, Centre for Environmental and Preventive Medicine, Barts and The London School of Medicine and Dentistry, Charterhouse Square, London, EC1M 6BQ, United Kingdom.

[3] Institute of Women's Health, University College London, Medical School Building, 74 Huntley Street, London, WC1E 6AU, United Kingdom.

Author contributions

The corresponding author LL contributed substantially to the design and conduct of the study, the collection, management and analysis of the data, and the preparation, review and approval of the manuscript. RM contributed substantially to the collection, management and analysis of the data and the preparation, review and approval of the manuscript. JK, JM and NM contributed substantially to the design and conduct of the study and the preparation, review and approval of the manuscript. LL wrote the first draft of the manuscript and each author listed has seen, critically revised and approved the submission of this version of the manuscript and takes full responsibility for the manuscript.

Funding

NIHR Doctoral Research Fellowship, NIHR-DRF-2012-05-206.

Acknowledgments

We thank Nia Wyn Roberts (MSc Econ), Outreach Librarian, Bodleian Health Care Libraries, Knowledge Centre, ORC Research Building, Old Road Campus, Headington, Oxford OX3 7DQ for her input and expertise during the search phase of the review.

Disclaimer

This article presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflict of interest

The sponsor, funder and authors have no conflicts of interest to disclose.

REFERENCES

1. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet. 2008;371(9608):261-269.
2. Linsell L, Malouf R, Morris J, et al. Prognostic Factors for Poor Cognitive Development in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review. Jama, Pediatr. 2015;169(12):1162-1172.
3. Linsell L, Malouf R, Morris J, et al. Prognostic Factors for Cerebral Palsy and Poor Motor Development in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review [Available online ahead of print Feb 10, 2016]. Developmental Medicine and Child Neurology. 2016:doi: 10.1111/dmcn.12972.
4. Linsell L, Malouf R, Johnson S, et al. Prognostic Factors for Behavioural Problems and Psychiatric Disorders in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review. Journal of Developmental and Behavioral Pediatrics. 2016;37(1):88-102.
5. Collins G, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of clinical epidemiology. 2015;68(2):112-121.
6. Arnold CC, Kramar MS, Hobbs CA, et al. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. American Journal of Epidemiology. 1991;134(6):604-613.
7. Van den Berg T, Heymans MW, Leone SS, et al. Overview of data-synthesis in systematic reviews of studies on outcome prediction models. BMC Medical research methodology. 2013;13(42).
8. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. PLOS Medicine. 2013;10(2):e1001380. .
9. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. British Medical Journal. 2009;339:b2535.
10. Aarnoudse-Moens CSH, Weisglas-Kuperus N, Duivenvoorden HJ, et al. Neonatal and parental predictors of executive function in very preterm children. Acta Paediatrica, International Journal of Paediatrics. 2013;102(3):282-286.
11. Adams-Chapman I, Bann CM, Vaucher YE, et al. Association between feeding difficulties and language delay in preterm infants using Bayley Scales of Infant Development-Third Edition. Journal of Pediatrics. 2013;163(3):680-685.e681-683.
12. Ambalavanan N, Carlo WA, Tyson JE, et al. Outcome trajectories in extremely preterm infants. Pediatrics. 2012;130(1):e115-e125.
13. Andrews WW, Cliver SP, Biasini F, et al. Early preterm birth: association between in utero exposure to acute inflammation and severe neurodevelopmental disability at 6 years of age. American Journal of Obstetrics and Gynecology. 2008;198(4):466.e461-466.e411.
14. Arnaud C, Daubisse-Marliac L, White-Koning M, et al. Prevalence and associated factors of minor neuromotor dysfunctions at age 5 years in prematurely born children: The EPIPAGE study. Archives of Pediatrics and Adolescent Medicine. 2007;161(11):1053-1061.
15. Ayoubi JM, Audibert F, Boithias C, et al. Perinatal factors affecting survival and survival without disability of extreme premature infants at two years of age. European journal of obstetrics, gynecology, and reproductive biology. 2002;105(2):124-131.
16. Beaino G, Khoshnood B, Kaminski M, et al. Predictors of the risk of cognitive deficiency in very preterm infants: The EPIPAGE prospective cohort. Acta Paediatrica, International Journal of Paediatrics. 2011;100(3):370-378.
17. Beaino G, Khoshnood B, Kaminski M, et al. Predictors of cerebral palsy in very preterm infants: The EPIPAGE prospective population-based cohort study. Developmental Medicine and Child Neurology. 2010;52(6):e119-e125.
18. Bolisetty S, Dhawan A, Abdel-Latif M, et al. Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants. Pediatrics. 2014;133(1):55-62.

19. Brevaut-Malaty V, Busuttil M, Einaudi MA, et al. Longitudinal follow-up of a cohort of 350 singleton infants born at less than 32 weeks of amenorrhea: neurocognitive screening, academic outcome, and perinatal factors. European Journal of Obstetrics Gynecology and Reproductive Biology. 2010;150(1):13-18.
20. Charkaluk ML, Truffert P, Fily A, et al. Neurodevelopment of children born very preterm and free of severe disabilities: The Nord-Pas de Calais Epipage cohort study. Acta Paediatrica, International Journal of Paediatrics. 2010;99(5):684-689.
21. Cooke RWI. Perinatal and postnatal factors in very preterm infants and sequent cognitive and motor abilities. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2005;90(1):F60-F63.
22. Davis N, Ford G, Anderson P, et al. Developmental coordination disorder at 8 years of age in a regional cohort of extremely-lowbirthweight or very preterm infants. Developmental Medicine & Child Neurology. 2007;49(5):325-330.
23. de Waal CG, Weisglas-Kuperus N, van Goudoever JB, et al. Mortality, neonatal morbidity and two year follow-up of extremely preterm infants born in the netherlands in 2007. PLoS ONE. 2012;7(7):e41302.
24. Delobel-Ayoub M, Arnaud C, White-Koning M, et al. Behavioral problems and cognitive performance at 5 years of age after very preterm birth: The EPIPAGE study. Pediatrics. 2009;123(6):1485-1492.
25. Delobel-Ayoub M, Kaminski M, Marret S, et al. Behavioral outcome at 3 years of age in very preterm infants: the EPIPAGE study. Pediatrics. 2006;117(6):1996-2005.
26. Dewey D, Creighton DE, Heath JA, et al. Assessment of developmental coordination disorder in children born with extremely low birth weights. Developmental Neuropsychology. 2011;36(1):42-56.
27. Doyle LW. Outcome at 5 years of age of children 23 to 27 weeks' gestation: Refining the prognosis. Pediatrics. 2001;108(1):134-141.
28. Emsley HCA, Wardle SP, Sims DG, et al. Increased survival and deteriorating developmental outcome in 23 to 25 week old gestation infants, 1990-4 compared with 1984-9. Archives of Disease in Childhood: Fetal and Neonatal Edition. 1998;78(2):F99-F104.
29. Farooqi A, Hagglof B, Sedin G, et al. Impact at age 11 years of major neonatal morbidities in children born extremely preterm. Pediatrics. 2011;127(5):e1247-e1257.
30. Fily A, Pierrat V, Delporte V, et al. Factors associated with neurodevelopmental outcome at 2 years after very preterm birth: the population-based Nord-Pas-de-Calais EPIPAGE cohort. Pediatrics. 2006;117(2):357-366.
31. Ford RM, Neulinger K, O'Callaghan M, et al. Executive function in 7-9-year-old children born extremely preterm or with extremely low birth weight: Effects of biomedical history, age at assessment, and socioeconomic status. Archives of Clinical Neuropsychology. 2011;26(7):632-644.
32. Franz AR, Pohlandt F, Bode H, et al. Intrauterine, early neonatal, and postdischarge growth and neurodevelopmental outcome at 5.4 years in extremely preterm infants after intensive neonatal nutritional support. Pediatrics. 2009;123(1):e101-109.
33. Gianni ML, Picciolini O, Vegni C, et al. Twelve-month neurofunctional assessment and cognitive performance at 36 months of age in extremely low birth weight infants. Pediatrics. 2007;120(5):1012-1019.
34. Goyen TA, Lui K. Developmental coordination disorder in "apparently normal" schoolchildren born extremely preterm. Archives of Disease in Childhood. 2009;94(4):298-302.
35. Hack M, Taylor HG, Schluchter M, et al. Behavioral outcomes of extremely low birth weight children at age 8 years. Journal of developmental and behavioral pediatrics : JDBP. 2009;30(2):122-130.
36. Hansen BM, Hoff B, Uldall P, et al. Perinatal risk factors of adverse outcome in very preterm children: A role of initial treatment of respiratory insufficiency? Acta Paediatrica, International Journal of Paediatrics. 2004;93(2):185-189.

37. Helderman JB, O'Shea TM, Kuban KC, et al. Antenatal antecedents of cognitive impairment at 24 months in extremely low gestational age newborns. Pediatrics. 2012;129(3):494-502.
38. Hok-Wikstrand M, Hard AL, Niklasson A, et al. Early postnatal growth variables are related to morphologic and functional ophthalmologic outcome in children born preterm. Acta Paediatrica, International Journal of Paediatrics. 2010;99(5):658-664.
39. Holmstrom GE, Kallen K, Hellstrom A, et al. Ophthalmologic outcome at 30 months' corrected age of a prospective Swedish cohort of children born before 27 weeks of gestation: the extremely preterm infants in sweden study. JAMA Ophthalmol. 2014;132(2):182-189.
40. Howard K, Roberts G, Lim J, et al. Biological and environmental factors as predictors of language skills in very preterm children at 5 years of age. Journal of Developmental and Behavioral Pediatrics. 2011;32(3):239-249.
41. Iijima S, Arai H, Ozawa Y, et al. Clinical patterns in extremely preterm (22 to 24 weeks of gestation) infants in relation to survival time and prognosis. American Journal of Perinatology. 2009;26(6):399-406.
42. Janssen AJWM, der Sanden MWGNv, Akkermans RP, et al. A model to predict motor performance in preterm infants at 5 years. Early Human Development. 2009;85(9):599-604.
43. Janssen AJWM, Nijhuis-van Der Sanden MWG, Akkermans RP, et al. Influence of behaviour and risk factors on motor performance in preterm infants at age 2 to 3 years. Developmental Medicine and Child Neurology. 2008;50(12):926-931.
44. Johnson S, Hollis C, Kochhar P, et al. Psychiatric Disorders in Extremely Preterm Children: Longitudinal Finding at Age 11 Years in the EPICure Study. Journal of the American Academy of Child and Adolescent Psychiatry. 2010;49(5):453-463.e451.
45. Johnson S, Hollis C, Kochhar P, et al. Autism Spectrum Disorders in Extremely Preterm Children. Journal of Pediatrics. 2010;156(4):525-531.e522.
46. Johnson S, Wolke D, Hennessy E, et al. Educational outcomes in extremely preterm children: Neuropsychological correlates and predictors of attainment. Developmental Neuropsychology. 2011;36(1):74-95.
47. Jones KM, Champion PR, Woodward LJ. Social competence of preschool children born very preterm. Early Human Development. 2013;89(10):795-802.
48. Kiechl-Kohlendorfer U, Ralser E, Pupp Peglow U, et al. Early risk predictors for impaired numerical skills in 5-year-old children born before 32weeks of gestation. Acta Paediatrica, International Journal of Paediatrics. 2013;102(1):66-71.
49. Leversen KT, Sommerfelt K, Ronnestad A, et al. Predicting neurosensory disabilities at two years of age in a national cohort of extremely premature infants. Early Human Development. 2010;86(9):581-586.
50. Leversen KT, Sommerfelt K, Ronnestad A, et al. Prediction of neurodevelopmental and sensory outcome at 5 years in norwegian children born extremely preterm. Pediatrics. 2011;127(3):e630-e638.
51. Lowe J, MacLean PC, Shaffer ML, et al. Early working memory in children born with extremely low birth weight: Assessed by object permanence. Journal of Child Neurology. 2009;24(4):410-415.
52. Marston L, Peacock JL, Calvert SA, et al. Factors affecting vocabulary acquisition at age 2 in children born between 23 and 28 weeks' gestation. Developmental Medicine and Child Neurology. 2007;49(8):591-596.
53. Matsuda Y, Kouno S, Hiroyama Y, et al. Intrauterine infection, magnesium sulfate exposure and cerebral palsy in infants born between 26 and 30 weeks of gestation. European Journal of Obstetrics, Gynecology, & Reproductive Biology. 2000;91(2):159-164.
54. Mercier CE, Dunn MS, Ferrelli KR, et al. Neurodevelopmental outcome of extremely low birth weight infants from the vermont oxford network: 1998-2003 and the Vermont Oxford Network ELBW infant follow-up study group. Neonatology. 2010;97(4):329-338.

55. Messinger D, Lambert B, Bauer CR, et al. The relationship between behavior ratings and concurrent and subsequent mental and motor performance in toddlers born at extremely low birth weight. Journal of Early Intervention. 2010;32(3):214-233.
56. Mikkola K, Ritari N, Tommiska V, et al. Neurodevelopmental outcome at 5 years of age of a national cohort of extremely low birth weight infants who were born in 1996-1997. Pediatrics. 2005;116(6):1391-1400.
57. Moore T, Johnson S, Hennessy E, et al. Screening for autism in extremely preterm infants: Problems in interpretation. Developmental Medicine and Child Neurology. 2012;54(6):514-520.
58. Neubauer AP, Voss W, Kattner E. Outcome of extremely low birth weight survivors at school age: The influence of perinatal parameters on neurodevelopment. European Journal of Pediatrics. 2008;167(1):87-95.
59. Orchinik LJ, Taylor HG, Espy KA, et al. Cognitive outcomes for extremely preterm/extremely low birth weight children in kindergarten. Journal of the International Neuropsychological Society. 2011;17(6):1067-1079.
60. Peralta-Carcelen M, Bailey K, Rector R, et al. Behavioral and socioemotional competence problems of extremely low birth weight children. Journal of Perinatology. 2013;33(11):887-892.
61. Pin TW, Eldridge B, Galea MP. Motor trajectories from 4 to 18months corrected age in infants born at less than 30weeks of gestation. Early Human Development. 2010;86(9):573-580.
62. Potharst ES, Houtzager BA, Van Sonderen L, et al. Prediction of cognitive abilities at the age of 5years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. Developmental Medicine and Child Neurology. 2012;54(3):240-246.
63. Potharst ES, Van Wassenaer-Leemhuis AG, Houtzager BA, et al. Perinatal risk factors for neurocognitive impairments in preschool children born very preterm. Developmental Medicine and Child Neurology. 2013;55(2):178-184.
64. Sann L, Bourgeois J, Stephant A, et al. A study of 490 premature babies born before 31 weeks' gestation age. Mortality and development before 2 years. Revue Francaise de Gynecologie et d'Obstetrique. 1999;94(6):500-510.
65. Sann L, Bourgeois J, Stephant A, et al. Outcome of 249 infants of less than 29 weeks gestational age. Archives de Pediatrie. 2001;8(3):250-258.
66. Sansavini A, Guarini A, Savini S. Linguistic and cognitive delays in very preterm infants at 2 years: General or specific delays? Revista de Logopedia, Foniatria y Audiologia. 2011;31(3):133-147.
67. Schlapbach LJ, Aebischer M, Adams M, et al. Impact of sepsis on neurodevelopmental outcome in a swiss national cohort of extremely premature infants. Pediatrics. 2011;128(2):e348-e357.
68. Schmidhauser J, Caflisch J, Rousson V, et al. Impaired motor performance and movement quality in very-low-birthweight children at 6 years of age. Developmental Medicine and Child Neurology. 2006;48(9):718-722.
69. Schmidt B, Asztalos EV, Roberts RS, et al. Impact of bronchopulmonary dysplasia, brain injury, and severe retinopathy on the outcome of extremely low-birthweight infants at 18 months: Results from the trial of indomethacin prophylaxis in preterms. JAMA: Journal of the American Medical Association. 2003;289(9):1124-1129.
70. Spittle AJ, Treyvaud K, Doyle LW, et al. Early Emergence of Behavior and Social-Emotional Problems in Very Preterm Infants. Journal of the American Academy of Child and Adolescent Psychiatry. 2009;48(9):909-918.
71. Stahlmann N, Rapp M, Herting E, et al. Outcome of extremely premature infants at early school age: health-related quality of life and neurosensory, cognitive, and behavioral outcomes in a population-based sample in northern Germany. Neuropediatrics. 2009;40(3):112-119.
72. Stephens BE, Bann CM, Watson VE, et al. Screening for Autism spectrum disorders in extremely preterm infants. Journal of Developmental and Behavioral Pediatrics. 2012;33(7):535-541.
73. Stoelhorst GM, Martens SE, Rijken M, et al. Behaviour at 2 years of age in very preterm infants (gestational age < 32 weeks). Acta Paediatrica. 2003;92(5):595-601.

74. Stoelhorst GMSJ, Rijken M, Martens SE, et al. Developmental outcome at 18 and 24 months of age in very preterm children: A cohort study from 1996 to 1997. Early Human Development. 2003;72(2):83-95.
75. Taylor HG, Klein N, Anselmo MG, et al. Learning problems in kindergarten students with extremely preterm birth. Archives of Pediatrics and Adolescent Medicine. 2011;165(9):819-825.
76. Taylor HG, Klein N, Drotar D, et al. Consequences and risks of <1000-g birth weight for neuropsychological skills, achievement, and adaptive functioning. Journal of Developmental and Behavioral Pediatrics. 2006;27(6):459-469.
77. Tommiska V, Heinonen K, Kero P, et al. A national two year follow up study of extremely low birthweight infants born in 1996-1997. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2003;88(1):F29-F34.
78. Toome L, Varendi H, Mannamaa M, et al. Follow-up study of 2-year-olds born at very low gestational age in Estonia. Acta Paediatrica, International Journal of Paediatrics. 2013;102(3):300-307.
79. Treyvaud K, Ure A, Doyle LW, et al. Psychiatric outcomes at age seven for very preterm children: rates and predictors. J Child Psychol Psychiatry. 2013;54(7):772-779.
80. Trittmann JK, Nelin LD, Klebanoff MA. Bronchopulmonary dysplasia and neurodevelopmental outcome in extremely preterm neonates. European Journal of Pediatrics. 2013;172(9):1173-1180.
81. Vincer MJ, Allen AC, Joseph KS, et al. Increasing prevalence of cerebral palsy among very preterm infants: a population-based study. Pediatrics. 2006;118(6):e1621-1626.
82. Vohr BR, Wright LL, Poole WK, et al. Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks' gestation between 1993 and 1998. Pediatrics. 2005;116(3):635-643.
83. Voss W, Jungmann T, Wachtendorf M, et al. Long-term cognitive outcomes of extremely low-birth-weight infants: The influence of the maternal educational background. Acta Paediatrica, International Journal of Paediatrics. 2012;101(6):569-573.
84. Wong HS, Huertas-Ceballos A, Cowan FM, et al. Evaluation of early childhood social-communication difficulties in children born preterm using the quantitative checklist for autism in toddlers. The Journal of Pediatrics. 2014;164(1):26-33.
85. Wood NS, Costeloe K, Gibson AT, et al. The EPICure study: Associations and entecedents of neurological and developmental disability at the 30 months of age following extremely preterm birth. Archives of Disease in Childhood: Fetal and Neonatal Edition. 2005;90(2):F134-F140.
86. Zanudin A, Burns Y, Gray PH, et al. Perinatal events and motor performance of children born with ELBW and nondisabled. Pediatr. 2013;25(1):30-35.
87. Zanudin A, Gray PH, Burns Y, et al. Perinatal factors in non-disabled ELBW school children and later performance. Journal of Paediatrics and Child Health. 2013;49(1):E62-E67.
88. Moons KG, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: what, why, and how? British Medical Journal. 2009;338:b606.
89. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology. 1996;49(12):1373-1379.
90. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;15(4):361-387.
91. Steyerberg EW, Eijkemans MJC, Harrell FE, et al. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Medical Decision Making. 2001;21:45-56.
92. Surveillance of cerebral palsy in Europe (SCPE): a collaboration of cerebral palsy surveys and registers. Developmental Medicine and Child Neurology. 2000;42(12):816-824.
93. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester UK: John Wiley and Sons Ltd; 2008.

94. Royston P, Altman DG, Sauerbrei W. Dichotomising continuous predictors in multiple regression: a bad idea. Statistics in medicine. 2006;25(1):127-141.
95. Altman DG, Royston P. The cost of dichotomising continuous variables [statistical notes]. British Medical Journal. 2006;332:1080.
96. Sun GW, Shook TL, Kay GL. Innapropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. Journal of clinical epidemiology. 1996;49(8):907-916.
97. Steyerberg EW, Eijkemans MJ, Harrell FE, et al. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Statistics in medicine. 2000;19(8):1059-1079.
98. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. journal of clinical epidemiology. 2004;57(11):1138-1146.
99. Harrell FE. Regression modelling strategies: with application to linear models, logistic regression and survival analysis. New York: Springer-Verlag; 2001.
100. Wolke D, Sohne B, Ohrt B, et al. Follow-up of preterm children: important to document drop-outs [letter]. Lancet. 1995;345(8947):447.
101. Johnson S, Fawke J, Hennessy E, et al. Neurodevelopmental disability through 11 years of age in children born before 26 weeks of gestation. Pediatrics. 2009;124(2):e249-e257.
102. Tin W, Fritz S, Wariyar U, et al. Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty. Arch Dis Child Fetal Neonatal Ed. 1998;79(2):F83-F87.
103. Steyerberg EW. Clinical prediction models. New York: Springer; 2008.
104. Geersing G-J, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS ONE. 2012;7(2).

Figure 1: Flow diagram of study search and selection (1990-2014)

Table 1: Summary of the Study Design of Studies Included in the Systematic Review (1990-2014)

Study design characteristics	Number of studies	% of studies (n=78)
Study design:		
Cohort	71	91
RCT	5	6.4
Case-control	1	1.3
Cross-sectional	1	1.3
Participant selection:		
Live births	36	46
Admissions to NICU	35	45
RCT participants	5	6.4
Selected from a database	1	1.3
Not clear	1	1.3
Center selection:		
Single center	35	45
All centers in a geographical -defined region	30	38
Multicenter (non-random)	13	17
Data extraction:		
Prospective	62	79
Retrospective	13	17
Not clear	3	4
Data collected as part of routine follow-up:		
No	57	73
Yes	21	27
Continent of study:		
Europe	46	59
North America	17	22
Australia and New Zealand	12	15
Asia	2	2.6
International	1	1.3
Start year of recruitment:		
1990-1995	25	32
1996-2000	33	42
2001-2010	20	26
GA and BW inclusion criteria:		
VPT (≤32 weeks)	24	31
EPT (≤29 weeks)	19	24
ELBW (≤1000g)	16	21
Other GA/BW combination	19	24
Exclusion criteria:		
Congenital anomalies	34	44
Severity of disability	22	28
Multiple births	7	9
Age of assessment (years):		
1.5 – 2	34	45
3 – 5	21	27
6 – 12	17	22
Other	6	8
Sample size:[a]	219 {141, 461}	[40, 7632]

[a] Median {IQR} [Range]. BW birth weight; EPT extremely preterm; ELBW extremely low birth weight; GA gestational age; IQR interquartile range; NICU neonatal intensive care unit; RCT randomized controlled trial; VPT very preterm.

Table 2: Summary of Model Development for the First Model Presented in Each Study Included in the Systematic Review (1990-2014)

Model development	Number of studies	% of studies (n=78)
Number of candidate risk factors at outset:[a]	17 {11, 24}	[3, 51]
Unclear	7	9
Rational for candidate risk factors		
Yes	29	37
No, but wide coverage (≥20)	27	35
No	22	28
Treatment of continuous risk factors:		
All left as continuous	17	22
Some categorized	33	42
All categorized	21	27
No continuous risk factors	1	1
Unclear	6	8
Coding of risk factors and outcome clearly described:	66	85
Statistical model		
Linear/logistic regression	74	95
Multinomial/ordinal logistic regression	2	3
Multilevel modelling	1	1
Unclear	1	1
Method of initial screening of candidate risk factors:		
No screening - all included	33	43
P-value from univariable model	22	28
P-value from multivariable model	6	8
Other	7	9
Unclear	10	13
Model building strategy:		
All candidates/screened candidate risk factors included	30	39
Stepwise selection	27	35
Backward selection	9	12
Forward selection	4	5
Other	3	4
Unclear	5	6
Any interaction terms fitted:	2	3
Any non-linear terms fitted:	0	0
Any factors forced into the final multivariable model:	15	19
Multicollinearity of risk factors mentioned/discussed:	11	14
Model fit/assumptions checked:	15	19
Number of risk factors in final multivariable model:[a]	5 {4, 9}	[0, 25]
Unclear	1	1

[a] Median {IQR} [Range]. IQR interquartile range.

Table 3: Summary of Reporting and Model Validation in Studies Included in the Systematic Review (1990-2014)

Model reporting and validation	Number of studies	% of studies (n=78)
Comparison of lost versus not lost to follow-up:		
No	21	30
Yes	48	70
Not applicable – all followed up	9	
Missing data presented for each variable:		
No	51	67
Yes	19	25
For some variables	6	8
Not applicable – none missing	2	
Number of particpants included in final model reported:	24	31
Point estimates of model coefficients reported:		
No	8	10
Yes – all variables	57	73
Significant variables only	7	9
Selected variables only	6	8
Confidence intervals/standard errors of model coefficients reported:		
No	11	14
Yes – all variables	56	72
Significant variables only	5	6
Selected variables only	6	8
P-value/significance level of model coefficients reported:		
No	5	6
Yes – all variables	55	71
Significant variables only	11	14
Selected variables only	7	9
Any statistical validation performed:	4	5
If yes, type of validation		
Discrimination	4	
Split sample	1	
Published in format for clinical use:	2	3