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

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

Affiliations:

[1] National Perinatal Epidemiology Unit (NPEU), Nuffield Department of Population Health, University of Oxford, Old Road Campus, Headington, Oxford, OX3 7LF (Email: reem.malouf@npeu.ox.ac.uk, jenny.kurinczuk@npeu.ox.ac.uk).

[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 (Email: j.k.morris@qmul.ac.uk).

[3] Institute of Women's Health, University College London, Medical School Building, 74 Huntley Street, London, WC1E 6AU (Email: n.marlow@ucl.ac.uk).

Corresponding author:

Louise Linsell
National Perinatal Epidemiology Unit (NPEU)
Nuffield Department of Population Health
University of Oxford
Old Road Campus
Headington
Oxford OX3 7LF
Email: louise.linsell@npeu.ox.ac.uk
Tel: 01865 617922

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

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

Abstract (248 words)

Background: The prediction of long term outcome in surviving infants born very preterm (VPT: ≤32 weeks) or with very low birth weight (VLBW: ≤1250g) is necessary to guide clinical management, provide information to parents and to help target and evaluate interventions. We sought to review and critically assess the methods and reporting of studies that have developed a risk prediction model in this population, and provide guidance for future research in this area

Methods: A systematic review was conducted using Medline, Embase and Pyscinfo databases to identify studies published between 1st January 1990 and 1st June 2014 reporting multivariate prediction models for neurodevelopment in VPT/VLBW children. Seventy-eight studies reporting 222 risk factor models for neurodevelopmental outcome were identified.

Results: The optimal study design, prospective follow up of all live births in a geographically defined region, was used in 25 (32%) of studies. Selection and screening of candidate risk factors was generally quite poorly reported and inappropriate techniques were often used in model building process. The reporting of attrition and missing data and its impact on results was also poor, with only 24 (31%) of studies reported the number of children included in the final model. Twenty-one (27%) of studies did not fully report the estimates of all risk predictors retained in the final model.

Conclusions: 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 and validate risk prediction models in this population.

Introduction

Prematurity and its associated neonatal morbidities have a pervasive effect on neurodevelopment, leading to: cerebral palsy, visual and auditory deficits, impairments in global and executive cognitive function, learning disabilities 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, are used routinely in clinical practice or adopted for use in research studies or policy evaluation. One reason for this is concern about the design, analysis and reporting of such studies, and the lack of formal model validation using robust statistical methods. Another reason is that most studies do not present the model in a format that is easily assessable and simple to use with a clear set of instructions.

The aim of this article is to review the conduct and reporting of 78 studies that were recently included in a systematic review of risk predictors for neurodevelopment outcome in children born very preterm (VPT: ≤32 weeks) or with very low birth weight (VLBW: ≤1250g).[2] *(2 in press - references to be added)*. Studies were included if the aim (or one of the aims) was to develop a multivariate (>2 variables) risk prediction model to predict 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 the risk of bias assessment of the 78 studies included in the review. The findings are then discussed within the framework of recommended approaches for analysis and reporting advocated by experts advising on risk prediction modelling in the medical and statistical literature.[3-8] Despite the recent publication of reporting guidelines for studies developing and validating risk prediction models,[9] there is no equivalent central resource that provides guidance for the design and analysis of such studies.

Methods

The methods for the systematic review of risk predictors for poor neurodevelopment in VPT/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] *(2 in press - references to be added)*.

Search strategy

Three electronic search strategies were devised in the Medline, Embase and Psycinfo databases (Boxes S1-3) using the National Institutes of Health Medical Subject Headings (NIH MeSH). The searches identified any journal articles published from 1st January 1990 to 1st June 2014 reporting a multivariate 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 1st January 1990, (3) study population was ≤32 weeks gestational age (GA) or with birth weight (BW) ≤1250g and not a highly select group (based on other clinical criteria), and (4) one objective was to perform a multivariate risk factor analysis (>2 variables) of a neurodevelopmental outcome assessed after 18 months of age.

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 and the results from this review are presented in accordance with the PRISMA guidelines.[10] 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. Most studies used the same techniques to develop their models and reported them the same way, so selecting just one model from each article to summarise model building, reporting and validation should be a good reflection of practices adopted in the literature.

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[a]) containing multivariate 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

[a] Studies based in any centre participating in the National Institute of Child Health and Human Development Neonatal Research Network (NICHD NRN) follow-up programme were classified as belonging to the same cohort.

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, were included in the data extraction.[11-88] Twenty eight articles reported a risk prediction model for motor outcomes, 31 for cognitive outcomes, 15 for behavioral problems or psychiatric disorders, three for visual outcomes, and 27 for composite outcomes based on a combination of the other neurodevelopmental domains.

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 randomised controlled trial (RCT) populations,[52,53,56,61,70] one case-control study[54] and one cross-sectional study.[32] Of the 71 prospective cohorts, 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 (NICU) admissions. Forty-five percent (n=35) of studies recruited from a single centre and 55% (n=43) from multiple centres. Seventy-nine percent (n=62) of studies collected risk factor and outcome data prospectively, but 17% (n=13) extracted data retrospectively from hospital records or a database, and in 27% (n=21) of studies the outcome data was collected as part of a routine follow-up visit. Overall, 32% (n=25) of studies were based on a cohort study of all live births in geographically defined region and followed up prospectively.

The majority of studies were conducted in Europe (59%), followed by North America (22%) and Australia and New Zealand (15%). 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), but 27% (n=21) studies followed up children to between 3-5 years of age and 21% (n=17) to between 6-12 years. Six studies had an age range that spanned over more than one of these categories. The median sample size was

219 {IQR: 141 to 461} and 11 studies had more than 1000 participants.[12,13,15,17-19,25,26,55,70,83] Only one study mentioned the issue of power and referred the number of events per variable in the modelling process.[40]

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] and in seven studies it was unclear what the initial list of factors were. 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 standardised 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 or a widely adopted convention was used, such as a threshold of two standard deviations below the mean for intelligence quotient score to denote moderate to severe cognitive impairment, 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 multivariate model, 28% (n=22) only included candidate factors with a p-value below a set threshold in a univariate test of association with outcome and 8% (n=6) screened variables using multivariate 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 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 assessed and 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 most 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 multivariate model, 27% (n=21) of studies only did so for selected variables, or failed to report them at all.

Only four studies[13,33,43,73] attempted to assess the performance of the final model. 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[43] produced a decision tree/algorithm based on their model and another[13] produced a web based tool to predict either death or neurodevelopmental impairment (NDI) - a composite of cerebral palsy, deafness, blindness, motor or cognitive delay - for use in clinical practice (see 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%). The calculator produced can predict the probability of death, death or NDI, and NDI alone. The models that predict death and death/NDI combined have been externally validated in independent populations,[89,90] though the model for NDI only (included in this review) has not been externally validated.

Discussion

This review has highlighted some strengths and weakness in the research methodology and reporting of risk prediction models for neurodevelopmental outcome in the VPT/VLBW population in the published literature. The data extracted on study design and conduct, model development, validation and reporting has identified a number of issues which should be addressed in the design and analysis of future studies. For each of the main issues, recommendations are made based on the latest literature in the field, and discussed in relation to the findings of the review.

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 centres in a geographically defined region, where the sample represents the source population and exposure to risk factors can be measured prior to the occurrence of an outcome.[3] However only 32% of studies in the review fulfilled the optimal study design and many studies were conducted in a single centre NICU. Studies based in a single centre, while convenient, have less generalisability due to centre differences in service provision, referral and management practices and the socio-economic/ethnic characteristics of the local population. Also, NICU 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 centres. Data from RCTs can be used to study prognosis, provided the treatment allocation is included as a predictor variable, though such studies may have reduced generalisabillity 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 to three 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 very small or immature newborns should be based on GA rather than BW criterion.[91] This is because the confounding effect of fetal growth status and maturity could lead to a distorted relationship between potential risk factors and outcome.

Sample size and events per variable

It is generally recommended that there should be a minimum of 10 outcome events per predictor variable (EPV) when using logistic regression models for predictive modelling, and models with an EPV less than five should be interpreted with caution.[92] As EPV declines below 10, simulation studies have shown an increase in bias and variability, unreliable confidence interval coverage and problems with model convergence.[6] Overfitting is a particular problem when small samples are used for predictive modelling, which can lead to the performance of the model being overoptimistic in the dataset from which it was developed.[93] The median sample size of studies included in the review was 219 and the median number of predictors 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%. Only one study commented on the number of events per variable.[40] Despite the difficulties in calculating a suitable sample size for studies of prognosis, and risk factor analysis often being a secondary aim when planning a cohort study, considerations of power and sample size still need to be discussed in relation to any additional analysis conducted.

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 standardisation, for example in the diagnostic criteria used by studies to define cerebral palsy, which are available.[94] 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 grading 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 and 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.[8] 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 categorisation 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.[95] Using cut-points that are data-driven, such as the sample median, are problematic when the model is transported to a different study population.[96] 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 categorisation 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.[8]

Selection of risk predictors

There is no overall consensus on the best strategy to select variables for inclusion in a prognostic model, however some approaches are not recommended. This includes screening candidate factors using univariate tests of association with outcome, as the correlation with other risk factors is not controlled for. This can result in the rejection of important predictors that only become prognostic after adjustment for other factors.[6,97] Twenty-eight percent (n=22) of studies adopted this approach and in 13% (n=10) the screening strategy was unclear. A further 43% (n=33) studies included all candidate variables, but 25 of these studies 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.[93,98,99] 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.[7] 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 the many environmental factors that can affect development following discharge. Six

studies[13,31,46,48,53,86] in the review adopted a sequential multivariate 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, the representativeness of analysis population should also be assessed, however the majority of studies did not even report the number of participants included in the final model after excluding those with missing outcome or risk factor data.

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). Once a risk prediction model has been derived, 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). Only four studies[16,36,46,76] in the review assessed the performance of the model, and in three this was limited to a single measure of discrimination. Only one study[13] went further and used the split-sample method of internal validation and produced a tool available online that could be used in clinical practice.

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.[103] 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 prognostic modelling research, though specific reporting guidelines for this field have only just emerged.[9]

Conclusion

This systematic review of 78 published articles has confirmed that there is a dearth of properly designed and well-conducted prognostic modelling studies for neurodevelopmental outcome in surviving VPT/VLBW children. Inappropriate techniques were often used to develop the risk prediction models included and few studies assessed model performance or attempted any model validation. In addition, poor reporting of the final models, model development, study attrition and missing data were widespread. 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 and validate risk prediction models in this population.

Acknowledgements

Funding source: NIHR Doctoral Research Fellowship, NIHR-DRF-2012-05-206. 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 study sponsor, funder and authors have no conflicts of interest to disclose and no honorariums, grants, or other forms of payment were given to anyone to produce the manuscripts. The sponsor and the funding organization has had no involvement in the (1) design and conduct of the study; (2) collection, management, analysis, and interpretation of the data; (3) preparation, review or approval of the manuscript; and (4) the decision to submit the paper for publication.

Duplicate/previous publication: The methods have previously been published in a review protocol (http://www.crd.york.ac.uk/PROSPERO/), registration number CRD42014006943.

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.

Acknowledgements: 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. No compensation was received for this contribution.

References

1. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008; **371**(9608): 261-9.
2. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Prognostic Factors for Poor Cognitive Development in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review. *Jama, Pediatr* 2015: 1-11.
3. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *British Medical Journal* 2009; **338**: 1317-20.
4. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *British Medical Journal* 2009; **338**: 1373-7.
5. Steyerberg EW. Clinical prediction models. New York: Springer; 2008.
6. 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**: 361-87.
7. Harrell FE. Regression modelling strategies: with application to linear models, logistic regression and survival analysis. New York: Springer-Verlag; 2001.
8. 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.
9. Collins G, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of clinical epidemiology* 2015; **68**: 134-43.
10. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 2009; **339**: b2535.
11. Aarnoudse-Moens CSH, Weisglas-Kuperus N, Duivenvoorden HJ, Oosterlaan J, Van Goudoever JB. Neonatal and parental predictors of executive function in very preterm children. *Acta Paediatrica, International Journal of Paediatrics* 2013; **102**(3): 282-6.
12. Adams-Chapman I, Bann CM, Vaucher YE, Stoll BJ, Eunice Kennedy Shriver National Institute of Child H, Human Development Neonatal Research N. Association between feeding difficulties and language delay in preterm infants using Bayley Scales of Infant Development-Third Edition. *Journal of Pediatrics* 2013; **163**(3): 680-5.e1-3.
13. Ambalavanan N, Carlo WA, Tyson JE, Langer JC, Walsh MC, Parikh NA, et al. Outcome trajectories in extremely preterm infants. *Pediatrics* 2012; **130**(1): e115-e25.
14. Andrews WW, Cliver SP, Biasini F, Peralta-Carcelen AM, Rector R, Alriksson-Schmidt AI, 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.e1-.e11.
15. Arnaud C, Daubisse-Marliac L, White-Koning M, Pierrat V, Larroque B, Grandjean H, 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): 1053-61.
16. Ayoubi JM, Audibert F, Boithias C, Zupan V, Taylor S, Bosson JL, 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): 124-31.
17. Beaino G, Khoshnood B, Kaminski M, Marret S, Pierrat V, Vieux R, 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): 370-8.

18. Beaino G, Khoshnood B, Kaminski M, Pierrat V, Marret S, Matis J, 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-e25.
19. Bolisetty S, Dhawan A, Abdel-Latif M, Bajuk B, Stack J, Lui K, et al. Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants. *Pediatrics* 2014; **133**(1): 55-62.
20. Brevaut-Malaty V, Busuttil M, Einaudi MA, Monnier AS, D'Ercole C, Gire C. 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): 13-8.
21. Charkaluk ML, Truffert P, Fily A, Ancel PY, Pierrat V. 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): 684-9.
22. 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-F3.
23. Davis N, Ford G, Anderson P, Doyle L. 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): 325-30.
24. de Waal CG, Weisglas-Kuperus N, van Goudoever JB, Walther FJ. Mortality, neonatal morbidity and two year follow-up of extremely preterm infants born in the netherlands in 2007. *PLoS ONE* 2012; **7**(7): e41302.
25. Delobel-Ayoub M, Arnaud C, White-Koning M, Casper C, Pierrat V, Garel M, et al. Behavioral problems and cognitive performance at 5 years of age after very preterm birth: The EPIPAGE study. *Pediatrics* 2009; **123**(6): 1485-92.
26. Delobel-Ayoub M, Kaminski M, Marret S, Burguet A, Marchand L, N'Guyen S, et al. Behavioral outcome at 3 years of age in very preterm infants: the EPIPAGE study. *Pediatrics* 2006; **117**(6): 1996-2005.
27. Dewey D, Creighton DE, Heath JA, Wilson BN, Anseeuw-Deeks D, Crawford SG, et al. Assessment of developmental coordination disorder in children born with extremely low birth weights. *Developmental Neuropsychology* 2011; **36**(1): 42-56.
28. Doyle LW. Outcome at 5 years of age of children 23 to 27 weeks' gestation: Refining the prognosis. *Pediatrics* 2001; **108**(1): 134-41.
29. Emsley HCA, Wardle SP, Sims DG, Chiswick ML, D'Souza SW. 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.
30. Farooqi A, Hagglof B, Sedin G, Serenius F. Impact at age 11 years of major neonatal morbidities in children born extremely preterm. *Pediatrics* 2011; **127**(5): e1247-e57.
31. Fily A, Pierrat V, Delporte V, Breart G, Truffert P. 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): 357-66.
32. Ford RM, Neulinger K, O'Callaghan M, Mohay H, Gray P, Shum D. 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): 632-44.
33. Franz AR, Pohlandt F, Bode H, Mihatsch WA, Sander S, Kron M, 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): e101-9.
34. Gianni ML, Picciolini O, Vegni C, Gardon L, Fumagalli M, Mosca F. Twelve-month neurofunctional assessment and cognitive performance at 36 months of age in extremely low birth weight infants. *Pediatrics* 2007; **120**(5): 1012-9.

35. Goyen TA, Lui K. Developmental coordination disorder in "apparently normal" schoolchildren born extremely preterm. *Archives of Disease in Childhood* 2009; **94**(4): 298-302.
36. Hack M, Taylor HG, Schluchter M, Andreias L, Drotar D, Klein N. Behavioral outcomes of extremely low birth weight children at age 8 years. *Journal of developmental and behavioral pediatrics : JDBP* 2009; **30**(2): 122-30.
37. Hansen BM, Hoff B, Uldall P, Greisen G. 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): 185-9.
38. Helderman JB, O'Shea TM, Kuban KC, Allred EN, Hecht JL, Dammann O, et al. Antenatal antecedents of cognitive impairment at 24 months in extremely low gestational age newborns. *Pediatrics* 2012; **129**(3): 494-502.
39. Hok-Wikstrand M, Hard AL, Niklasson A, Hellstrom A. 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): 658-64.
40. Holmstrom GE, Kallen K, Hellstrom A, Jakobsson PG, Serenius F, Stjernqvist K, 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): 182-9.
41. Howard K, Roberts G, Lim J, Lee KJ, Barre N, Treyvaud K, 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): 239-49.
42. Iijima S, Arai H, Ozawa Y, Kawase Y, Uga N. 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.
43. Janssen AJWM, der Sanden MWGNv, Akkermans RP, Tissingh J, Oostendorp RAB, Kollee LAA. A model to predict motor performance in preterm infants at 5 years. *Early Human Development* 2009; **85**(9): 599-604.
44. Janssen AJWM, Nijhuis-van Der Sanden MWG, Akkermans RP, Oostendorp RAB, Kollee LAA. 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): 926-31.
45. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N. 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): 453-63.e1.
46. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N. Autism Spectrum Disorders in Extremely Preterm Children. *Journal of Pediatrics* 2010; **156**(4): 525-31.e2.
47. Johnson S, Wolke D, Hennessy E, Marlow N. Educational outcomes in extremely preterm children: Neuropsychological correlates and predictors of attainment. *Developmental Neuropsychology* 2011; **36**(1): 74-95.
48. Jones KM, Champion PR, Woodward LJ. Social competence of preschool children born very preterm. *Early Human Development* 2013; **89**(10): 795-802.
49. Kiechl-Kohlendorfer U, Ralser E, Pupp Peglow U, Pehboeck-Walser N, Fussenegger B. 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.
50. Leversen KT, Sommerfelt K, Ronnestad A, Kaaresen PI, Farstad T, Skranes J, et al. Predicting neurosensory disabilities at two years of age in a national cohort of extremely premature infants. *Early Human Development* 2010; **86**(9): 581-6.
51. Leversen KT, Sommerfelt K, Ronnestad A, Kaaresen PI, Farstad T, Skranes J, et al. Prediction of neurodevelopmental and sensory outcome at 5 years in norwegian children born extremely preterm. *Pediatrics* 2011; **127**(3): e630-e8.

52. Lowe J, MacLean PC, Shaffer ML, Watterberg K. Early working memory in children born with extremely low birth weight: Assessed by object permanence. *Journal of Child Neurology* 2009; **24**(4): 410-5.
53. Marston L, Peacock JL, Calvert SA, Greenough A, Marlow N. Factors affecting vocabulary acquisition at age 2 in children born between 23 and 28 weeks' gestation. *Developmental Medicine and Child Neurology* 2007; **49**(8): 591-6.
54. Matsuda Y, Kouno S, Hiroyama Y, Kuraya K, Kamitomo M, Ibara S, 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): 159-64.
55. Mercier CE, Dunn MS, Ferrelli KR, Howard DB, Soll RF. 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): 329-38.
56. Messinger D, Lambert B, Bauer CR, Bann CM, Hamlin-Smith K, Das A. 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): 214-33.
57. Mikkola K, Ritari N, Tommiska V, Salokorpi T, Lehtonen L, Tammela O, 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): 1391-400.
58. Moore T, Johnson S, Hennessy E, Marlow N. Screening for autism in extremely preterm infants: Problems in interpretation. *Developmental Medicine and Child Neurology* 2012; **54**(6): 514-20.
59. 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.
60. Orchinik LJ, Taylor HG, Espy KA, Minich N, Klein N, Sheffield T, et al. Cognitive outcomes for extremely preterm/extremely low birth weight children in kindergarten. *Journal of the International Neuropsychological Society* 2011; **17**(6): 1067-79.
61. Peralta-Carcelen M, Bailey K, Rector R, Gantz M. Behavioral and socioemotional competence problems of extremely low birth weight children. *Journal of Perinatology* 2013; **33**(11): 887-92.
62. 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): 573-80.
63. Potharst ES, Houtzager BA, Van Sonderen L, Tamminga P, Kok JH, Last BF, 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): 240-6.
64. Potharst ES, Van Wassenaer-Leemhuis AG, Houtzager BA, Livesey D, Kok JH, Last BF, et al. Perinatal risk factors for neurocognitive impairments in preschool children born very preterm. *Developmental Medicine and Child Neurology* 2013; **55**(2): 178-84.
65. Sann L, Bourgeois J, Stephant A, Putet G. 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): 500-10.
66. Sann L, Bourgeois J, Stephant A, Putet G. Outcome of 249 infants of less than 29 weeks gestational age. *Archives de Pediatrie* 2001; **8**(3): 250-8.
67. 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): 133-47.
68. Schlapbach LJ, Aebischer M, Adams M, Natalucci G, Bonhoeffer J, Latzin P, et al. Impact of sepsis on neurodevelopmental outcome in a swiss national cohort of extremely premature infants. *Pediatrics* 2011; **128**(2): e348-e57.

69. Schmidhauser J, Caflisch J, Rousson V, Bucher HU, Largo RH, Latal B. Impaired motor performance and movement quality in very-low-birthweight children at 6 years of age. *Developmental Medicine and Child Neurology* 2006; **48**(9): 718-22.
70. Schmidt B, Asztalos EV, Roberts RS, Robertson CM, Sauve RS, Whitfield MF. 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): 1124-9.
71. Spittle AJ, Treyvaud K, Doyle LW, Roberts G, Lee KJ, Inder TE, 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): 909-18.
72. Stahlmann N, Rapp M, Herting E, Thyen U. 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): 112-9.
73. Stephens BE, Bann CM, Watson VE, Sheinkopf SJ, Peralta-Carcelen M, Bodnar A, et al. Screening for Autism spectrum disorders in extremely preterm infants. *Journal of Developmental and Behavioral Pediatrics* 2012; **33**(7): 535-41.
74. Stoelhorst GM, Martens SE, Rijken M, van Zwieten PH, Zwinderman AH, Wit JM, et al. Behaviour at 2 years of age in very preterm infants (gestational age < 32 weeks). *Acta Paediatrica* 2003; **92**(5): 595-601.
75. Stoelhorst GMSJ, Rijken M, Martens SE, Van Zwieten PHT, Feenstra J, Zwinderman AH, 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.
76. Taylor HG, Klein N, Anselmo MG, Minich N, Espy KA, Hack M. Learning problems in kindergarten students with extremely preterm birth. *Archives of Pediatrics and Adolescent Medicine* 2011; **165**(9): 819-25.
77. Taylor HG, Klein N, Drotar D, Schluchter M, Hack M. Consequences and risks of <1000-g birth weight for neuropsychological skills, achievement, and adaptive functioning. *Journal of Developmental and Behavioral Pediatrics* 2006; **27**(6): 459-69.
78. Tommiska V, Heinonen K, Kero P, Pokela ML, Tammela O, Jarvenpaa AL, 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.
79. Toome L, Varendi H, Mannamaa M, Vals MA, Tanavsuu T, Kolk A. Follow-up study of 2-year-olds born at very low gestational age in Estonia. *Acta Paediatrica, International Journal of Paediatrics* 2013; **102**(3): 300-7.
80. Treyvaud K, Ure A, Doyle LW, Lee KJ, Rogers CE, Kidokoro H, et al. Psychiatric outcomes at age seven for very preterm children: rates and predictors. *J Child Psychol Psychiatry* 2013; **54**(7): 772-9.
81. Trittmann JK, Nelin LD, Klebanoff MA. Bronchopulmonary dysplasia and neurodevelopmental outcome in extremely preterm neonates. *European Journal of Pediatrics* 2013; **172**(9): 1173-80.
82. Vincer MJ, Allen AC, Joseph KS, Stinson DA, Scott H, Wood E. Increasing prevalence of cerebral palsy among very preterm infants: a population-based study. *Pediatrics* 2006; **118**(6): e1621-6.
83. Vohr BR, Wright LL, Poole WK, McDonald SA. Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks' gestation between 1993 and 1998. *Pediatrics* 2005; **116**(3): 635-43.
84. Voss W, Jungmann T, Wachtendorf M, Neubauer AP. 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): 569-73.

85. Wong HS, Huertas-Ceballos A, Cowan FM, Modi N. 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.
86. Wood NS, Costeloe K, Gibson AT, Hennessy EM, Marlow N, Wilkinson AR. 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-F40.
87. Zanudin A, Burns Y, Gray PH, Danks M, Poulsen L, Watter P. Perinatal events and motor performance of children born with ELBW and nondisabled. *Pediatr* 2013; **25**(1): 30-5.
88. Zanudin A, Gray PH, Burns Y, Danks M, Watter P, Poulsen L. Perinatal factors in non-disabled ELBW school children and later performance. *Journal of Paediatrics and Child Health* 2013; **49**(1): E62-E7.
89. Boland RA, Davis PG, Dawson JA, Doyle LW. Predicting death or major neurodevelopmental disability in extremely preterm infants born in Australia. *Arch Dis Child Fetal Neonatal Ed* 2013; **98**(3): F201-4.
90. Chong Lee H, Green C, Hintz SR, Tyson JE, Parikh NA, Langer J, et al. Prediction of death for extremely premature infants in a population-based cohort. *Pediatrics* 2010; **126**: e644.
91. Arnold CC, Kramar MS, Hobbs CA, Mclean FH, Usher RH. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. *American Journal of Epidemiology* 1991; **134**(6): 604-13.
92. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology* 1996; **49**(12): 1373-9.
93. Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JDF. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Medical Decision Making* 2001; **21**: 45-56.
94. Surveillance of cerebral palsy in Europe (SCPE): a collaboration of cerebral palsy surveys and registers. *Developmental Medicine and Child Neurology* 2000; **42**: 816-24.
95. Royston P, Altman DG, Sauerbrei W. Dichotomising continuous predictors in multiple regression: a bad idea. *Statistics in medicine* 2006; **25**: 127-41.
96. Altman DG, Royston P. The cost of dichotomising continuous variables. *British Medical Journal* 2006; **332**: 1080.
97. 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**: 907-16.
98. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in medicine* 2000; **19**: 1059-79.
99. 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**: 1138-46.
100. Wolke D, Sohne B, Ohrt B, Riegel K. Follow-up of preterm children: important to document drop-outs. *Lancet* 1995; **345**(8947): 447.
101. Johnson S, Fawke J, Hennessy E, Rowell V, Thomas S, Wolke D, et al. Neurodevelopmental disability through 11 years of age in children born before 26 weeks of gestation. *Pediatrics* 2009; **124**(2): e249-e57.
102. Tin W, Fritz S, Wariyar U, Hey E. 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**: F83-F7.
103. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS ONE* 2012; **7**(2).