

Prognostic factors for behavioral problems and psychiatric disorders in children born very preterm or very low birth weight: a systematic review

Running title: Risk factors for psychiatric disorders in VPT children

Authors:

Louise Linsell¹, MSc

Reem Malouf¹, MDI

Samantha Johnson², PhD

Joan Morris³, PhD

Jennifer J Kurinczuk¹, MD

Neil Marlow⁴, MD

Affiliations:

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

² Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester LE1 6TP.

³ 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.

⁴ Institute of Women's Health, University College London, Medical School Building, 74 Huntley Street, London, WC1E 6AU.

Corresponding author and reprint address: Louise Linsell, National Perinatal Epidemiology Unit, 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, Fax: 01865 289701].

Conflicts of Interest and Source of Funding : The sponsor, funder and authors have no conflicts of interest to disclose. The research reported in this manuscript was funded by a NIHR Doctoral Research Fellowship Award, NIHR-DRF-2012-05-206.

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.

Abstract

Objective: Risk factors associated with adverse behavioral outcomes in very preterm (VPT) or very low birth weight (VLBW) infants are poorly understood. The aim of this paper is to identify prognostic factors for behavioral problems and psychiatric disorders in children born ≤ 32 weeks gestational age or with birth weight ≤ 1250 g.

Methods: A systematic review was conducted using Medline, Embase and Pyscinfo databases to identify studies published between 01/01/1990 and 01/06/2014 reporting multivariable prediction models for behavioral problems or psychiatric disorders in VPT/VLBW children. Fifteen studies were identified and two independent reviewers extracted key information on study design, outcome definition, risk factor selection, model development, reporting, and conducted a risk of bias assessment.

Results: The 15 studies included reported risk factor analyses for the following domains: general behavioral problems (n=8), any psychiatric disorder (n=2), autism spectrum symptoms/disorders (n=5), and attention deficit/hyperactivity disorder (n=1). Findings were inconclusive due to the: small number of studies in each domain, heterogeneity in outcome measures, lack of overlap in the risk factors examined and differences in strategies for dealing with children with neurological impairments.

Conclusion: There is a lack of evidence concerning risk factors for behavior problems and psychiatric disorders among VPT/VLBW survivors. This review has identified the need for further research examining the etiology of disorders of psychological development in the VPT/VLBW population in order to refine risk prediction and identify targets for intervention. Large, well-conducted studies that use standard diagnostic evaluations to assess psychiatric disorders throughout childhood and adolescence are required.

Key words: Risk factors, child psychiatry, behavior and emotional disorders, autistic spectrum disorder, attention deficit hyperactivity disorder, preterm infants, very low birth weight, systematic review.

Introduction

Advances in obstetric and neonatal care have led to a steady increase in the survival rate of preterm children,^{1,2} but this has also been accompanied by an increase in the prevalence of long term sequelae such as neurodevelopmental impairment and psychiatric disorders. Studies using behavioral screening questionnaires have shown that children born very preterm (VPT; ≤ 32 weeks gestation) and with very low birth weight (VLBW; ≤ 1250 g) are at increased risk of social, emotional and attention problems and internalising problems (anxiety/depression) compared with term-born controls.³ Clinically significant behavior problems on screening questionnaires have been reported in 13 to 46% of VPT/VLBW children.⁴ However screening tools are designed to have a high rate of sensitivity, in order to identify children who are at risk of developing a psychiatric disorder and for whom further assessment would be beneficial⁵⁻⁷ and thus the rates of diagnosed disorders is typically lower. Studies using diagnostic evaluations have reported an excess of attention deficit/hyperactivity disorders (ADHD), autism spectrum disorders (ASD) and psychiatric disorders in general compared to term-born controls.^{8,9} A recent review of clinical cohort studies reported that the prevalence of DSM-IV-TR¹⁰ based ADHD diagnoses ranged between 16 to 19% in VPT/VLBW children, with an increase in odds of 2 to 3 compared to term-born peers.⁴ ASD are less common, with a median prevalence of 0.6% in the general population, but two studies have reported that 3.6% of extremely low birth weight children (ELBW; ≤ 1000 g)¹¹ and 8% of extremely preterm (EPT; ≤ 28 weeks gestation) children,¹² respectively, met diagnostic criteria when assessed between 8-11 years. Behavioral problems in VPT/VLBW children have been shown to persist into adolescence,^{13,14} and there is evidence that the risk of being diagnosed with psychiatric disorders in adulthood increases with decreasing gestational age (GA).^{15,16}

The pattern of behavioral problems observed in VPT/VLBW children has been shown to be similar across different countries, despite cultural differences and disparity in neonatal care, implicating some underlying biological mechanism.¹⁷ It has been suggested that a “preterm behavioral

phenotype” may exist, characterised by socio-communicative and emotional problems and inattention.⁴ The mechanisms underlying this neurobehavioral profile are unclear, though several explanations have been proposed.¹⁸ The VPT/VLBW newborn brain is extremely vulnerable and clinical and environmental factors that disturb a critical period of brain development that normally takes place in utero may be highly influential. Exposure to prolonged hospitalisation and therapeutic interventions may disrupt normal neurodevelopment, even in the absence of focal brain injury. This is compounded by the stressful environment of a busy neonatal intensive care unit (NICU) with a high noise level, constant bright lighting, multiple monitoring devices and reduced opportunity for parent-infant interaction. Early exposure to such a sustained level of stress may have an adverse impact on brain development, akin to that observed in adults.¹⁹ Later environmental influences in early infancy and childhood, such as parental mental health, caregiving style, or limited contact with peers and family due to prolonged periods of hospitalisation/illness may impede the development of coping strategies, emotional regulation, attachment and other social skills,²⁰ all of which are more likely to occur following VPT/VLBW birth.

Early identification of behavioural problems in VPT/VLBW infants may prevent the development of psychiatric disorders later in life, however the risk factors associated with adverse behavioral outcomes in this population are poorly understood. The aim of this paper is to perform a systematic review of articles reporting multivariable outcome prediction models for behavioral problems and psychiatric disorders in the VPT/VLBW population, in order to identify robust predictors of outcome. This paper is part of a wider comprehensive systematic review of risk factors for poor neurodevelopmental outcomes in VPT/VLBW survivors, conducted to consolidate the evidence on risk to inform future prognostic research.

Methods

The methods for the overall systematic review have previously been published in a review protocol (<http://www.crd.york.ac.uk/PROSPERO/>), registration number CRD42014006943 (see Supplemental Digital Content 1).

Search strategy

Three electronic search strategies were devised in the Medline, Embase and Psycinfo databases (see Boxes S1-S3, Supplemental Digital Content 2) 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 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 1st January 1990, (3) study population was ≤ 32 weeks GA or with birth weight ≤ 1250 g 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 assessed after 18 months of age.

All study designs were included and 1990 was chosen as a cut-off date for year of birth because surfactant therapy was adopted routinely into clinical care in many countries around this time. This was a transition from the “pre-surfactant” era of high mortality and morbidity to the “surfactant era” of improved survival and prognosis.^{21, 22} There were also improvements in the use of assisted ventilation, prophylactic infection control and antenatal steroid therapy around this time. The birth

weight cut-off of $\leq 1250\text{g}$ was chosen to exclude the subset of more mature but extremely growth restricted children included in the typical $\leq 1500\text{g}$ VLBW cohort which can cause heterogeneity and lead to confounding bias when examining the relationship between risk factors and outcome.²³

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 this 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 in relation to 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.^{24, 25}

Data extraction

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 inclusion criteria were applied (screen 2). The two screens were performed by the first author (LL) in the first instance, 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 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 and risk of bias assessment on a customised MS Access 2010 database. Every single item entered was manually cross-checked for

discrepancies at a face-to-face meeting. These were discussed and resolved or referred to the rest of the author review team if agreement could not be reached.

The following data items were extracted: study design, participant setting, centre selection, study location, year of birth, gestational age, birth weight, age at assessment, selection criteria of study population, sample size, completeness of data at follow-up, details of outcomes assessed, number of candidate risk factors assessed, variable selection, treatment of continuous variables, adjustment for confounders, method of analysis, model assumptions checked, missing data analysis, presentation of multivariable model, details of risk factors included in final model, strength of association, statistical validation and clinical validation. If critical information was missing or unclear the corresponding author was contacted once by email for clarification.

Risk of bias assessment

Overwhelming evidence shows that the conduct and reporting of published articles describing the development or validation prediction models are poor,²⁶ which has led to the development of quality assessment tools specific for these types of study. In this review, the quality of studies was assessed according to a modified version of the QUIPS tool, which is a standardised set of criteria recommended for use in reviews of prognosis²⁷ (see Table S1, Supplemental Digital Content 3). The tool focuses on six areas of potential bias pertinent to studies of prognosis: study participation, study attrition, prognostic factor measurement, outcome measurement, confounding measurement and account, and statistical analysis. Studies were graded as [yes/partly/no] for each domain and classified as having a low-moderate risk of bias if they were graded as [yes] or [partly] in all six bias domains and moderate-high risk of bias otherwise.

Data synthesis and reporting

Results were presented in accordance with the PRISMA guidelines.²⁸ Risk factors that were statistically significant ($p < 0.05$) in the final model were reported for each study. In studies that reported multiple models, for example for different disorders, subscales of a global score or further sensitivity analyses, all models are referenced in the results tables for completeness, but only the significant risk factors from the main models are presented. Studies were grouped according to type of outcome studied; general behavioral problems, psychiatric disorder, ASD and ADHD, and according to age of assessment; early childhood (< 5 years) and middle childhood (≥ 5 years). Assessments in early infancy can be unreliable and based on more general behavioral screeners, whereas assessments in later childhood tend to have higher specificity, particularly if based on strict diagnostic criteria, hence risk factors may differ.

Results

The searches for the comprehensive systematic review retrieved 44,500 articles for the comprehensive review of risk factors for neurodevelopmental outcomes, and after removing duplicates, the first screen on title and abstract was performed on 32,283 articles (Figure 1). For 29,999 the title or abstract clearly indicated that the topic of the article was not relevant to the review question or did not satisfy one of the inclusion criteria. The remaining 2,284 articles were screened on full text, applying the full set of eligibility criteria. Eligibility was unclear in 136 (6%) and these were reviewed by the second independent reviewer (RM), or the author was contacted (where uncertainty was due to missing information). After applying the eligibility criteria, 91 articles (from 48 cohort populations¹) containing multivariable risk factor analyses were eligible for inclusion. Following panel review, 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

¹ 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.

objective; the remaining 78 articles were included in the data extraction for the comprehensive systematic review. No further articles were identified in the hand-search of bibliographies. This review paper summarizes the results of the 15 studies (from 9 cohort populations) reporting risk factor analyses for a behavioral or psychiatric (defined as a diagnosis appearing in DSM-IV-TR) outcome.^{9, 11, 12, 29-40} Two articles containing behavioral outcomes were excluded due to cohort overlap.^{41, 42} The remaining 63 of the 78 studies did not contain behavioral or psychiatric outcomes.

Study characteristics

The main study design was prospective cohort (n=14), and there was one randomized controlled trial (RCT).³¹ Of the 14 prospective cohorts, seven were ascertained from all live births in a geographically defined region,^{9, 12, 29, 32, 34, 35, 40} five were recruited from a single centre NICU,^{11, 30, 33, 36, 37} and three from multiple NICUs.^{31, 38, 39} Studies were conducted in seven countries: United States (n=4), UK or England (n=4), Australia (n=2), France (n=2) and one study each from Germany, Netherlands and New Zealand. The median sample size was 219 (range 75 to 1228) and two studies had more than 1000 participants.^{32, 34} Five studies were restricted to extremely preterm children; <27 weeks^{35, 38, 40} and <26 weeks,^{9, 12} and two studies excluded multiple births.^{32, 34} The risk of bias assessment classified four studies as low-moderate risk of bias and 11 studies as moderate-high risk of bias (Figure 2).

The 15 studies included in the review comprised 47 risk factor analyses for behavioral or psychiatric outcomes. Some studies reported a model for a global score and also models for each subdomain, whilst others reported additional models adjusting for concurrent factors such as cognition and language. The median number of candidate risk factors considered at the outset in each study was 16 (range 7 to 42). For the initial screening of candidates to be entered into the final model, five studies included them all and seven included those with a p-value below a set threshold after initial screening. The most popular method of model building after initial screening was to include all

factors screened (n=6 studies) and stepwise selection (n=5 studies). Only six of the 15 studies reported the number of participants included in the final model presented. One study assessed model discrimination using the area under the receiver operating curve,³⁸ but apart from that no studies performed any type of statistical or clinical validation.

Risk factors for general behavioral problems

Eight studies contained a risk factor analysis for general behavioral problems (Table 1); five studies assessed outcome under 5 years of age²⁹⁻³³ and three studies over 5 years.³⁴⁻³⁶ Six of the studies excluded and/or adjusted for neurodevelopmental delay or cognitive impairment.^{29, 31-35} All studies used validated, parent report behavioral screening questionnaires, the most common being the Strengths and Difficulties Questionnaire (SDQ)⁴³ (n=4).³²⁻³⁵ The Total Difficulties Score consists of four (5-item) subscales: conduct problems, inattention-hyperactivity, emotional symptoms and peer problems. One study²⁹ used the Child Behavior Checklist (CBCL),⁴⁴ which is a 99-item questionnaire with six syndrome scales that are combined to give an overall Total Problem score: anxious/depressed, withdrawn, aggressive, destructive, sleep problems and somatic behavior. The 169-item Infant-Toddler Symptom Checklist (ITSEA)⁴⁵ and its brief 42-item version (BITSEA)⁴⁶ were used by two studies.^{30, 31} Both checklists include items measuring internalising and externalising problems, dysregulation and socio-emotional competence.. Both the SDQ and ITSEA/BITSEA have been shown to be highly correlated with the CBCL.^{46, 47} One study used the Vineland Adaptive Behavior Scales Screener (VABSS)⁴⁸ which measures adaptive functioning in the domains of communication, socialisation and daily living skills.

There was only one low-moderate risk of bias study (which also had a sample size >1000) among this group of eight studies examining general behavioral problems.³² Factors that were found to be significant predictors for behavioral problems at age 3 years in this study were: hospitalisation after neonatal discharge, lower maternal age, lower level of maternal education, and

neurodevelopmental delay/poor health status measured at the time of assessment. The later study in the same cohort at age 5 years³⁴ had similar findings. All eight studies entered some indicator of socio-economic deprivation into the final model, for example education, income, social risk, and five found at least one of these factors significantly related to behavioral problems.^{30-33, 35} In the six studies that adjusted for neurodevelopmental delay or general cognitive ability at the time of assessment, five studies found a significant association between these factors and poorer behavioral outcomes.³¹⁻³⁵ Two studies reported that female sex and one study reported that male sex was significantly associated with behavioral problems, but five of the studies did not find sex significant in the final model. Overall, there was not enough overlap in the risk factors identified in this small group of studies to provide any conclusive evidence about prognostic factors for general behavioral problems.

Risk factors for psychiatric disorders

Seven studies reported risk factor analyses for psychiatric disorders (Table 2); two for any DSM-IV-TR¹⁰ diagnosis,^{9, 37} five for ASD symptoms or diagnoses^{11, 12, 38-40} and one for ADHD¹¹ (this study also reported a model for ASD).

Both studies examining the risk of any psychiatric disorder used the Development And Well Being Assessment (DAWBA),⁴⁹ which is a structured psychiatric evaluation administered to parents and teachers. In both studies, a DSM-IV-TR diagnosis was assigned by two blinded clinical psychologists aided by the DAWBA computer scoring algorithm for common childhood diagnoses, such as ADHD, ASD, emotional and conduct disorders. The DAWBA has good concurrent validity when compared with clinical diagnoses.⁴⁹ In the moderate-high risk of bias study conducted at age 7 years,³⁷ the prevalence of any DSM-IV-TR diagnosis was 24% in VPT children, which was similar to 23% prevalence rate reported by the low-moderate risk of bias study conducted at age 10-12 years in EPT children.⁹ The first study screened 11 candidate risk factors in a univariate analysis and retained 4

significant factors in the final model: brain abnormality at term, female sex, social-emotional problems at 5 years (SDQ) and higher familial social risk at 7 years.² The second study entered 34 candidate risk factors into a multivariate forward stepwise regression model and retained 5 significant factors in the final model: necrotizing enterocolitis, internalising behavior problems at 2.5 years (CBCL), pervasive attentional and conduct problems at 6 years (SDQ) and serious neurodevelopmental disability at 6 years. These findings suggest that behavioral problems identified by screening tests in infancy and early childhood may help to identify children at risk of developing a psychiatric disorder in later childhood.

The five studies examining risk factors for ASD were divided into those that assessed ASD symptoms using dimensional measures^{12, 39}, the rate of positive screens using screening tools^{11, 38, 40} and diagnoses made using a diagnostic evaluation¹² (Table 2). The two studies reporting risk factor analyses for ASD symptoms were not comparable with respect to age of assessment, outcome measure used, gestational age group, exclusion criteria, risk of bias, and had no significant risk factors in common.^{12, 39} However, similar to the findings for general behavioral problems (Table 1) and any psychiatric disorders (Table 2), markers of social deprivation and language development,³⁹ and earlier cognitive and behavioral assessments¹² were reported to be significantly associated with ASD symptoms later in childhood.

Of the three studies that presented risk factor analyses for a positive ASD screen, two were conducted in early childhood.^{38, 40} One was a low-moderate risk of bias study³⁸ that defined cases as children with at least one positive screen from three different screening tests at 18 months (Pervasive Developmental Disorders Screening Test-II⁵⁰ and two items adapted from the Autism Diagnostic Observation Scales (ADOS)⁵¹). The second was a moderate-high risk of bias study⁴⁰ that used the 23-item Modified Checklist for Autism in Toddlers (M-CHAT)⁵² at 2 years. The third study,

² Familial social risk was based on a composite measure of six social risk factors: family structure, education of primary caregiver, occupation and employment status of primary income earner, language spoken at home and maternal age at birth

also at moderate-high risk of bias, was conducted in later childhood at 8 years¹¹ using the 12 items related to Autistic Disorder from the Parent Child Symptom Inventory (CSI-4).⁵³ Amongst these studies the prevalence of a positive ASD screen varied greatly; 20%, 41% and 2% respectively likely reflecting differences in population denominators and screening tools. Bronchopulmonary dysplasia and male sex were significant risk factors in two out of three of these studies, but there were no other significant risk factors in common. The low-moderate risk of bias ASD screening study³⁸ presented two additional models adjusting for language, cognition and social-emotional behavioral problems, at the same age of assessment, all of which were significant.

Only one low-moderate risk of bias study conducted at age 10-12 years assigned ASD diagnosis based on standard diagnostic DSM-IV-TR criteria, using the DAWBA.¹² The prevalence of an ASD diagnosis was 8% (n=16 cases): 13 (6.5%) with autistic disorder and 3 (1.5%) with pervasive developmental disorder not otherwise specified. The risk prediction model was based on any ASD diagnosis and after entering 42 candidate variables sequentially into a multivariate stepwise model, only two factors remained significant; cognitive impairment and pervasive peer problems at age 6 years (SDQ).

The only study that presented a risk factor analysis for a positive screen for ADHD¹¹ did not report any significant risk factors for either hyperactive, inattentive or the combined type of ADHD. The prevalence of a positive screen for ADHD was reported to be 17% (n=37) in this study.

Discussion

Summary of findings

The eight studies reporting risk factor analysis for general behavioral problems (Table 1) in children born VPT/VLBW all had a moderate to high risk of bias, with one exception,³² and the screening tools used were fairly heterogeneous with different sub-domains assessed. The modelling of outcome

scores also varied with some studies reporting the proportion of children scoring above the cut-off for a positive screen and others analysing continuous scores. The studies also differed in the way children with neurodevelopmental delay or disability were handled in the design and analysis; some studies excluded them completely, some adjusted for motor and/or cognitive impairment and some adopted both or neither strategy. There was also a lack of commonality in the risk factors studied for prognosis, therefore it was difficult to synthesise the results and reach any meaningful conclusion. The only factors that appeared to be consistent predictors of general behavioral problems were markers of socio-economic deprivation and neurodevelopmental or cognitive delay, but apart from these there was no clear evidence about the prognostic value of any other risk factors studied.

Two studies examined the risk of developing a DSM-IV-TR psychiatric disorder in later childhood^{9, 37} and reported that social-emotional and behavioral problems identified by screening questionnaires in infancy or early childhood were predictive of later disorders. This finding is supported by other studies that have examined the predictive validity of screening tests and the stability of diagnoses over time.^{54, 55} Early screening tests are known to identify a large number of false-positives, particularly in impaired populations with high rates of neurologic and cognitive impairment,⁵ so the positive predictive value for later psychiatric diagnoses may be low. There is also a lack of evidence about how sensitive these general screening tests are for predicting specific types of DSM-IV-TR disorder. However, there is evidence of enhanced specificity in prediction in VPT/EPT populations for both general disorders and specific behavioral outcomes.⁵⁶ Given the stability in neurodevelopmental and behavioral outcomes in VPT/VLBW children, early screening may thus have greater predictive validity and clinical utility in preterm populations.

The only factors consistently associated with ASD symptoms, positive screen or diagnosis, were cognitive or language impairment, and poor performance on a behavioral screening test earlier in childhood. Aside from this, no clear evidence emerged for any other risk factors. The number of

cases with an ASD diagnosis was very small in the study conducted in later childhood using DSM-IV-TR based diagnostic criteria,¹² so a lack of power means that results should be interpreted with caution. Only one study presented a risk factor analysis for a positive screen for ADHD and no significant factors were identified.¹¹

Explanation of findings

An explanation for the inconclusive findings, beside the small number of studies examining each type of disorder and the lack of commonality in the candidate risk factors studied, is the several different strategies used for dealing with confounding due to neurologic and cognitive impairment. In some studies the whole cohort was included, representing the whole spectrum of disability in the VPT/VLBW population. In other studies, children with neurological and/or cognitive impairment were excluded from the modelling process to identify risk factors in a more homogeneous population and to eliminate the noise created by additional impairments. Other studies attempted to achieve this by adjusting for these factors in the analysis. The risk factors for a psychiatric disorder in the absence of any impairment may be very different to those factors which are prognostic for behavioral difficulties accompanying profound impairment. Therefore the strategy for dealing with motor, neurosensory and cognitive impairment in risk factor analyses, in terms of exclusion and/or adjustment, will crucially affect the findings. Adjustment for cognitive impairment is particularly problematic as it is frequently associated with psychiatric conditions in the term population; adjustment in a VPT/VLBW population where cognitive delay is more common and part of the preterm phenotype might lead to overcorrection.⁵⁷

VPT/VLBW children with motor or cognitive impairment have been reported to be at higher risk of developing behavioral and emotional problems, compared to VPT/VLBW children with no impairments.⁵⁸ It is possible that the challenge of living with a profound impairment could induce feelings of anxiety, insecurity and detachment which then manifests as a behavioral problem.

However, the high rate of problems in children with neurodevelopmental impairments may also be related to measurement issues. In a cohort of 2 year old EPT children, Kuban et al⁵⁹ reported that increased odds of a positive screen for autism using the M-CHAT among those unable to sit or stand was 23-fold, 8-fold in those with a major vision or hearing impairment and 13-fold in those with severe cognitive impairment, compared to EPT children without such impairments. Moore et al⁴⁰ reported similar findings in a cohort of EPT children at 2 years; 16.5% of children without disability screened positive on the M-CHAT compared to 96% with severe motor impairment, 56% with cognitive impairment and all children with a significant vision or hearing impairment. However, such findings should be interpreted with caution, as many items on the M-CHAT rely on an intact motor, hearing and vision function which leads to an inflated false-positive rate among children with impairment(s) of these functions. Indeed, a recent study has shown that screening for autism using the M-CHAT questionnaire was especially confounded in a preterm population.⁶⁰ However, the rate of positive screens were still 3-fold higher among unimpaired EPT children compared to unselected populations,⁵⁹ hence neurological impairment cannot be the sole explanation for the differences observed. Even so, it is difficult to disentangle the etiology of neurobehavioral disorders in the context of the neurological sequelae that follow VPT birth.

Strengths and limitations

We used a broad search filter with no language restriction in order to capture all studies with exploratory risk factor analyses, which is recommended in this type of review.⁶¹ 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. The study cohorts spanned a 20-year period and represent diverse international populations with differing methods of ascertainment and clinical practices which may also explain the inconclusive results. Also, studies did not all consider the same sets of candidate factors. Some prognostic factors for behavioral problems are likely to be interrelated, therefore we focused our systematic review on studies in which multivariable prediction

models were used as these take account of any multicollinearity between variables during the development process. One challenge in this review was the lack of independence between observations, arising from studies based on the same cohort population or single studies reporting more than one model. We selected studies for inclusion before data synthesis was conducted using standard rules, although it was difficult apply a strict set of criteria for each case. There was no evidence that GA was a predictor of behavioral or psychiatric problems, despite recent studies demonstrating a gradient of risk of poor neurodevelopment with decreasing GA across the full GA spectrum.^{62, 63} However, this review included only a restricted range of children born VPT/VLBW. A significant association with GA may be more likely to be observed if children born across the full spectrum of GA were studied.

Recommendations

This systematic review points to the need for further well-conducted research investigating risk factors for psychiatric disorders and behavioral problems in the VPT/VLBW population. Such conditions are common following VPT birth and can have an adverse impact on the lives of children and their families. As such, the identification of predictive factors is important for understanding etiological mechanisms and for developing appropriate screening, intervention and treatment strategies. Studies with larger sample sizes and greater power are needed for studying childhood psychiatric disorders in this population, particularly for less common conditions such as ASD or ADHD. Longer term follow-up with outcome evaluations beyond 18-24 months is also needed, as the risk for psychiatric disorders cannot be reliably assessed at this age and because of the natural course of some disorders which may onset later in childhood. Furthermore, prognosis is likely to be a dynamic process with social and environmental factors potentially superseding the influence of early clinical and biological factors as the child grows up. This review included studies using both GA and birth weight criteria to define the study population, but future studies evaluating behavioral problems and psychiatric disorders in preterm infants should use cohorts defined solely by GA. This

avoids the distorted birth weight distribution created in the study population when GA is paired with a birth weight criterion.

The risk of bias assessment identified a number of improvements that could be made to the design, conduct and reporting of future studies which should be made in accordance to the recent TRIPOD guidelines on the transparent reporting of prognostic research.²⁶ We recommend as standard the reporting of attrition and missing data, the use of standard diagnostic evaluations to assess outcome, the evaluation of a broad range of biologic and social risk factors over time and a clear statement and rationale as to the the inclusion or exclusion of children with cognitive or neurologic impairment. Future studies should also go beyond the scope of fitting risk factor models and test the robustness of their performance over time and in other independent cohorts using methods of statistical validation.

References

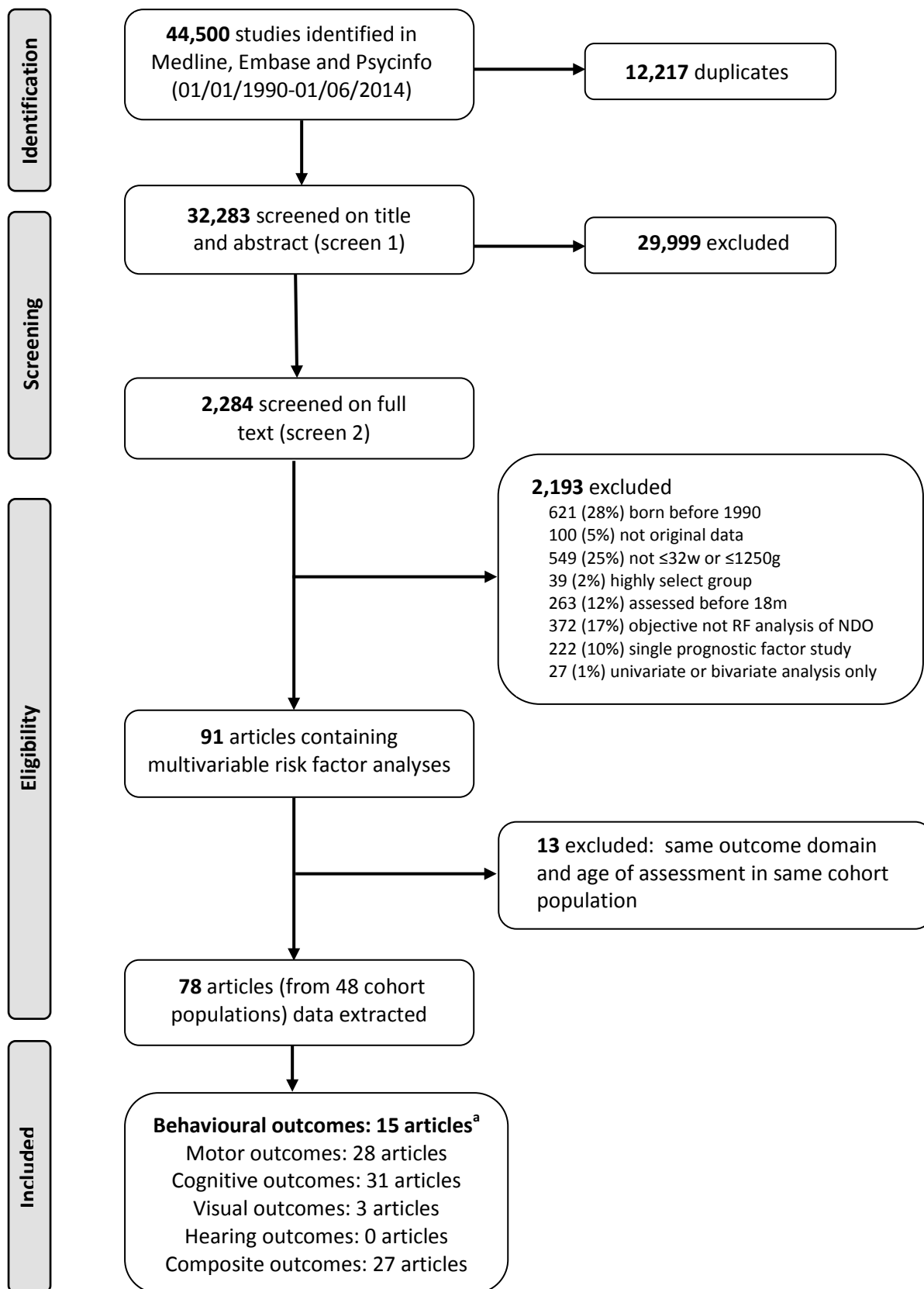
1. Tucker J, McGuire W. Epidemiology of preterm birth. *British Medical Journal* 2004; **329**: 675-8.
2. Goldenberg RL, Culhane JF, Iams JD, et al. Epidemiology and causes of preterm birth. *The Lancet* 2008; **371**: 75-84.
3. Farooqi A, Hagglof B, Sedin G, et al. Mental health and social competencies of 10- to 12-year-old children born at 23 to 25 weeks of gestation in the 1990s: A Swedish national prospective follow-up study. *Pediatrics* 2007; **120**(1): 118-33.
4. Johnson S, Marlow N. Preterm birth and childhood psychiatric disorders. *Pediatric Research* 2011; **69**(5 PART 2): 11R-8R.
5. Johnson S, Hollis C, Hennessy E, et al. Screening for autism in preterm children: diagnostic utility of the Social Communication Questionnaire. *Archives of Disease in Childhood* 2011; **96**(1): 73-7.
6. Glascoe FP. Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews* 2005; **11**: 173-9.
7. Limperopoulos C, Bassan H, Sullivan NR, et al. Positive screening for autism in ex-preterm infants: prevalence and risk factors. *Pediatrics* 2008; **121**(4): 758-65.
8. Bhutta AT, Cleves MA, Casey PH, et al. Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *JAMA* 2002; **288**(6): 728-37.
9. 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): 453-63.e1.
10. Association AP. Diagnostic and Statistical Manual of Mental Disorders. 4th edition, text revision (4th ed.) ed. Washington, DC: American Psychiatric Association; 2000.
11. 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): 122-30.
12. Johnson S, Hollis C, Kochhar P, et al. Autism Spectrum Disorders in Extremely Preterm Children. *Journal of Pediatrics* 2010; **156**(4): 525-31.e2.
13. Indredavik MS, Vik T, Heyerdahl S, et al. Psychiatric symptoms and disorders in adolescents with low birth weight. *Arch Dis Child Fetal Neonatal Ed* 2004; **89**: F445-F50.
14. Saigal S, Pinelli J, Hoult L, et al. Psychopathology and social competencies of adolescents who were extremely low birth weight. *Pediatrics* 2003; **111**(5): 969-75.
15. Moster D, Terje R, Markestad T. Long-term medical and social consequences of preterm birth. *NEJM* 2008; **359**: 262-73.
16. Lindstrom K, Lindblad F, Hjern A. Psychiatric morbidity in adolescents and young adults born preterm: a Swedish national cohort study. *pediatrics* 2009; **123**: e47-e53.
17. Hille ETM, den Ouden AL, Saigal S, et al. Behavioural problems in children who weigh 1000g or less at birth in four countries. *The Lancet* 2001; **357**: 1641-43.
18. Perlman. Neurobehavioral deficits in premature graduates of intensive care - potential medical and neonatal environmental risk factors. *pediatrics* 2001; **108**: 1339-48.
19. Sapolsky RM. Why stress is bad for your brain. *Science* 1996; **273**: 749-50.
20. Burnett AC. Prevalence of psychiatric diagnoses in preterm and full-term children, adolescents and young adults: a meta-analysis. *Psychological Medicine* 2011; **41**: 2463-74.
21. Baron IS, Rey-Casserly C. Extremely preterm birth outcome: a review of four decades of cognitive research. *Neuropsychological Review* 2010; **20**: 430-52.
22. Hack M, Fanaroff AA. Outcomes of children of extremely low birthweight and gestational age in the 1990's. *Early Human Development* 1999; **53**: 193-218.

23. 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): 604-13.
24. 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).
25. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLOS Medicine* 2013; **10**(2).
26. 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**: 134-43.
27. Hayden JA, Van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013; **158**: 280-6.
28. 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.
29. 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.
30. 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): 909-18.
31. 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): 887-92.
32. 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.
33. Jones KM, Champion PR, Woodward LJ. Social competence of preschool children born very preterm. *Early Human Development* 2013; **89**(10): 795-802.
34. 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): 1485-92.
35. 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): 112-9.
36. 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): 459-69.
37. 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): 772-9.
38. 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): 535-41.
39. 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.
40. 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): 514-20.
41. 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): 684-9.
42. 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): 214-33.
43. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology & Psychiatry* 1997; **1997**(38): 581-6.

44. Achenbach TM. Manual for the Child Behavior Checklist/2-3 and 1992 profile. Burlington: University of Vermont, Department of Psychiatry; 1992.
45. Briggs-Gowan MJ, Carter AS. The Infant Toddler Social and Emotional Assessment (ITSEA). Pre-publication version. Boston, MA: Yale University; 2000.
46. Briggs-Gowan MJ, Carter AS, Irwin JR, et al. The Brief Infant Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology* 2004; **29**: 143-55.
47. Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology* 1999; **27**(1): 17-24.
48. Sparrow SS, Carter AS, Cicchetti DV. Vineland Screener: Overview, reliability, validity, administration and scoring. New Haven, CT: Yale University; 2000.
49. Goodman R, Ford T, Richards H, et al. The Development and Well-Being Assessment: description and initial valuation of an integrated assessment of child and adolescent psychopathology. *Journal of child psychology and psychiatry* 2000; **41**: 645-55.
50. Siegel B. Pervasive Developmental Disorders Screening Test-II. San Antonio, TX: PsychCorp; 2004.
51. Lord C, Rutter M, DiLavore PCea. Autism Diagnostic Observation Schedule Manual. Los Angeles, CA: Western Psychological Services; 2001.
52. Robins DL, Dumont-Mathieu TM. Early screening for autism spectrum disorders: update on Modified Checklist for Autism in Toddlers and other measures. *Journal of Developmental and Behavioral Pediatrics* 2006; **27** (Suppl 2): S111-9.
53. Gadow KD, Sprafkin J. Child Symptom Inventory-4 Norms Manual. Stony Brook, NY: Checkmate Plus Ltd.; 1997.
54. Treyvaud K, Doyle LW, Lee KJ, et al. Social-emotional difficulties in very preterm and term 2 year olds predict specific social-emotional problems at the age of 5 years. *Journal of Pediatric Psychology* 2012; **37**(7): 779-85.
55. Gray RF, Indurkha A, McCormick MC. Prevalence, stability, and predictors of clinically significant behavior problems in low birth weight children at 3, 5, and 8 years of age. *Pediatrics* 2004; **114**(3): 736-43.
56. Johnson S, Marlow N. Growing up after extremely preterm birth: Lifespan mental health outcomes. *Seminars in Fetal and Neonatal Medicine* 2014; **19**: 97-104.
57. Dennis M, Francis DJ, Cirino PT, et al. Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society* 2009; **15**: 331-43.
58. Doyle LW, Bowman E, Callanan C, et al. Postnatal corticosteroids and sensorineural outcome at 5 years of age. *Journal of Paediatrics and Child Health* 2000; **36**(3): 256-61.
59. Kuban KC, O'Shea MT, Allred EN, et al. Positive screening on the Modified Checklist for Autism in Toddlers (M-CHAT) in extremely low gestational age newborns. *Journal of Pediatrics* 2009; **154**: 535-40.
60. Guy A, Seaton SE, Boyle EM, et al. Infants born late/moderately preterm are at increased risk for a positive autism screen at 2 years of age. *Journal of pediatrics* 2015; **166**(2): 269-75.
61. 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).
62. MacKay DF, Smith GCS, Dobbie R, et al. Gestational age at delivery and special educational need: retrospective cohort of 407,503 schoolchildren. *PLOS Medicine* 2010; **7**(6).
63. Poulsen G, Wolke D, Kurinczuk JJ, et al. Gestational age and cognitive ability in early childhood: a population-based cohort study. *Paediatric and perinatal epidemiology* 2013; **27**: 371-9.
64. Bayley N. Bayley Scales of Infant Development-II. 2nd ed: Psychological Corporation; San Antonio, TX; 1993.

65. Allison C, Baron-Cohen S, Wheelwright S, et al. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18-24 months of age: preliminary report *Journal of Autism and Developmental Disorders* 2008; **38**: 1414-25.
66. Rutter M, Bailey A, Lord C. The Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2003.

Figure 1: Flow diagram



^a Reviewed in this article.

Figure 2: Risk of bias assessment of the 15 behavioral studies included in the review

	Risk of bias domain						Low-moderate risk of bias
	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Analysis and reporting	
Delobel-Ayoub (2006)	😊	😊	😊	😊	😊	😊	✓
Delobel-Ayoub (2009)	😊	😞	😊	😊	😊	😊	
Hack (2009)	😞	😊	😞	😊	😊	😞	
Johnson (2010a)	😊	😊	😊	😊	😊	😊	✓
Johnson (2010b)	😊	😊	😊	😊	😊	😊	✓
Jones (2013)	😞	😊	😊	😊	😊	😊	
Moore (2012)	😊	😞	😊	😊	😊	😊	
Peralta-Carcelen (2013)	😞	😞	😊	😊	😊	😊	
Spittle (2009)	😞	😊	😊	😊	😊	😊	
Stahlmann (2009)	😊	😊	😊	😊	😊	😞	
Stephens (2012)	😊	😊	😊	😊	😊	😊	✓
Stoelhorst (2003)	😊	😊	😊	😊	😊	😞	
Taylor (2006)	😞	😊	😊	😊	😊	😞	
Treyvaud (2013)	😞	😊	😊	😊	😊	😊	
Wong (2014)	😊	😊	😞	😊	😊	😊	

Risk of bias: 😊 Low 😊 Moderate 😞 High

Table 1: Summary of studies reporting risk factor analyses for general behavioral problems in children born very preterm or with very low birth weight

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ birth weight (grams)	Design and participants	Number (%) of survivors assessed ^a	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Early childhood <5 years								
Stoelhorst (2003) ^{29, b}	Netherlands 1996-1997	2	<32w	PC of all live births in 3 Dutch health regions comprising 9% of the population.	160 (68%)	Total Problem score, Internalising and Externalising scale from the CBCL. Parent report.	Adjusted for: neurological abnormalities at 2 yrs.	Total problem: SGA. Internalising: SGA. Externalising: None significant.
Spittle (2009) ³⁰	Australia 2001-2003	2	<30w or <1250g	PC study of Infants admitted to a single centre NICU and enrolled in Victorian Infant Brain Studies (Melbourne).	188 (84%)	Externalising, Internalising, Dysregulation and Competence domains from ITSEA. Parent report.	None.	Internalising: Higher social risk. ^c Externalising: None significant. Dysregulation: None significant. Competence: Lower BW, PN steroids, female sex, moderate-severe WMA.
Peralta-Carcelen (2013) ³¹	United States 1999-2001	2.5	<1000g	Infants admitted to the NICU of 15 centres participating in the multi-centre NICHD NRN routine FUP and enrolled in a glutamine supplementation RCT.	696 (60%)	Total Competence (≤15th vs. >15th centile) and Total Problem score (≥75th vs. <75th centile) from BITSEA. Parent report.	Excluded: blind, deaf, syndrome (n=30). Adjusted for: CP, abnormal neurological exam, MDI<70 and PDI<70 from BSID-II at 2.5 yrs.	Total Competence: Hispanic or non-white ethnicity, MDI<70 and PDI<70 from BSID-II at 2.5 yrs. Total Problem: Female sex, lower household income, MDI<70 and PDI<70 from BSID-II at 2.5 yrs.
Delobel-Ayoub (2006) ³²	France 1997	3	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study). Excluded multiples.	1228 (69%)	Total Difficulties score from SDQ (≤10th vs. >10th centile of control group). Parent report.	Excluded: blind, deaf, severe CP (n=63). Adjusted for: neurodevelopmental delay and health status at 3 yrs.	Hospitalisation in last yr, lower maternal age, lower maternal education, neurodevelopmental delay and poor health status at 3 yrs.
Jones (2013) ³³ [E] 2230	New Zealand 1998-2000	4	<33w	PC of infants admitted to a single centre NICU (Christchurch).	105 (98%)	Social competence composite score. ^d Parent report.	Excluded from all: blind (n=1). Model 1: Risk factors to term. Model 2: Adjusted for family functioning and parenting 2-4 yrs. Model 3: Adjusted for factors in model 2 plus IQ at 4 yrs.	Model 1: Family SES, male sex, severity of neonatal WMA. Model 2: Male sex, higher maternal anxiety at 2-4 yrs, negative and intrusive parenting at 4 yrs. Model 3: GA<28w, higher maternal anxiety at 2-4 yrs, negative and intrusive parenting at 4 yrs, lower IQ at 4 yrs.

Table 1: Summary of studies reporting risk factor analyses for general behavioral problems in children born very preterm or with very low birth weight (continued)

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ birth weight (grams)	Design and participants	Number (%) of survivors assessed ^a	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Middle childhood ≥5 years								
Delobel-Ayoub (2009) ³⁴	France 1997	5	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study). Excluded multiples.	1102 (59%)	Total Difficulties score from SDQ (≤10th vs. >10th centile of control group). Parent report.	Excluded: blind, deaf, severe CP (n=63). Adjusted for: IQ and development (parent reported) at 5 yrs.	Hospitalisations in last 5 years, lower maternal age, poor maternal mental well-being in previous month, lower IQ and delayed development at 5yrs
Stahlmann (2009) ³⁵	Germany 1997-1999	7-9	<27w	PC of all live births in all 8 perinatal centres in Schleswig-Holstein.	75 (82%)	Total Difficulties score from SDQ. Parent report.	Adjusted for: IQ at 7-9 yrs.	Lower maternal education, IQ<70 at 7-9 yrs.
Taylor (2006) ^{36 e}	United States 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (86%)	Adaptive Behavior Composite score from the VABSS (cts and <1SD below mean of control group). Parent report.	Each risk factor was fitted separately and adjusted sex, race, parental SES, family stressors and family resources (p-values not reported).	Model 1 (cts score): Longer neonatal hospital stay, outborn. Model 2 (<70 vs. ≥70): Longer neonatal hospital stay, NRI>3.

^a Percentage of survivors assessed for outcome measure specified.

^b 9 models reported in total; Total Problem score, Internalising and Externalising scales and the 6 syndrome scores that comprise the total score.

^c Social risk was based on a composite measure of six social risk factors: family structure, education of primary caregiver, occupation and employment status of primary income earner, language spoken at home and maternal age at birth.

^d Social competence composite score was derived by the authors by summing sub-scale scores across the following measures: SDQ, Behavioral Inventory of Executive Function – Preschool version (BRIEF-P), Emotional Regulation Checklist (ERC), Infant Toddler Symptom Checklist (ITSC), Emotional Regulation subscale from BSID-II, Penn Interactive Peer Play Scale (PIPPS). 3 models were reported; the full model adjusting for IQ is reported in this review.

^e 2 models for Adaptive Behavior Composite and its 3 domains reported; one based on dichotomous outcome and one based on continuous outcome.

Abbreviations: BITSEA Brief Infant-Toddler Social and Emotional Screening;⁴⁶ BSID Bayley Scales of Infant Development;⁶⁴ BW birth weight; CBCL Child Behavior Checklist;⁴⁴ CP cerebral palsy; FUP follow up; GA gestational age; IQ intelligence quotient; ITSEA Infant-Toddler Social and Emotional Assessment;⁴⁵ MDI Mental Developmental Index from the BSID-II; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; NRI neonatal risk index; PC prospective cohort; PDI Psychomotor Developmental Index from the BSID-II; PN post natal; RCT randomized controlled trial; SES socio-economic status ; SDQ Strengths and Difficulties Questionnaire;⁴³ SGA small for gestational age; VABSS Vineland Adaptive Behavior Scales Screener;⁴⁸ WMA white matter abnormality.

Table 2: Summary of studies reporting risk factor analyses for psychiatric disorders in children born very preterm or with very low birth weight

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ birth weight (grams)	Design and participants	Number (%) of survivors assessed ^a	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Any psychiatric disorder: diagnosis								
Treyvaud (2013) ³⁷	Australia 2001-2003	7	<30w or <1250g	PC study of Infants admitted to a single centre NICU and enrolled in Victorian Infant Brain Studies (Melbourne).	177 (79%)	DAWBA. Parent report. DSM-IV-TR diagnosis assigned using scoring algorithm and clinical judgement of 2 blinded reviewers.	None.	Brain abnormality at term, female sex, social-emotional problems at 5 years (SDQ), higher familial social risk at 7 years. ^b
Johnson (2010a) ⁹	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICURE Study).	219 (71%)	DAWBA. Parent report. DSM-IV-TR diagnosis assigned using scoring algorithm and clinical judgement of 2 blinded reviewers.	None.	NEC, Internalising behavior problems at 2.5 yrs (CBCL), pervasive attentional and conduct problems at 6 yrs (SDQ), serious functional disability at 6 yrs.
Autism spectrum symptoms: dimensional measure								
Wong (2014) ³⁹	England 2010-2012	1.8-2.2	<33w	Infants attending routine FUP in 13 centres (London). Neonatal data extracted retrospectively.	141 (70%)	Q-CHAT score. Parent report.	Excluded: CP or severe neurosensory impairment (n=10). Adjusted for: Language Composite Score from BSID-III at 2 yrs.	Higher deprivation, non-white ethnicity, BSID-III Language Composite Score at 2 yrs.
Johnson (2010b) ¹²	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICURE Study).	219 (71%)	Total score from SCQ. Parent report.	None.	No breast milk, IQ<2SD at 6 yrs, pervasive attentional and peer problems at 6 yrs (SDQ), withdrawn (CBCL) at 2.5 yrs.
Autism spectrum disorder: positive screen								
Stephens (2012) ³⁸	United States 2008-2010	1.5-1.9	<27w	PC of infants admitted to the NICU of 15 centres participating in the multi-centre NICHD NRN routine FUP.	554 (74%)	1+ positive screen on 3 tests: PDDST-II (parent report), Response to Joint Attention and Response to Name (ADOS, direct observation).	Excluded from all: severe CP, blind, deaf (n=31). Model 1: Unadjusted Model 2: Adjusted for cognition and language at 18m. Model 3: Adjusted for cognition, language and behavior at 18m.	Model 1: Lower BW, non-white ethnicity, male sex. Model 2: Male sex, lower Cognitive and Language Composite Score from BSID-III at 18m. Model 3: Lower Cognitive and Language Composite Score from BSID-III at 18m, higher Problem and lower Competence Score from BITSEA at 18m.
Moore (2012) ^{40 c}	England 2006	2	<27w	PC of all live births in England (EPICURE-2 Study).	559 (54%)	Positive M-CHAT screen. Parent report.	Model 1: Full cohort Model 2: Excluded neuro-sensory impairment (n=72). Model 3: Excluded any disability (n=320)	Model 1: Severe BPD, any CUSS abnormality, PN steroids, positive blood culture ≥72 hrs, male sex. Model 2: Positive blood culture ≥72 hrs. Model 3: Any CUSS abnormality, positive blood culture <72hrs and ≥72 hrs, male sex.
Hack (2009) ¹¹	United States 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	219 (97%)	CSI-4 based on DSM-IV-TR diagnostic criteria. Parent report.	Each risk factor was fitted separately and adjusted sex, race, parental SES (p-values not reported).	BPD.

Table 2: Summary of studies reporting risk factor analyses for psychiatric disorders in children born very preterm or with very low birth weight (continued)

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ birth weight (grams)	Design and participants	Number (%) of survivors assessed ^a	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Autism spectrum disorder: diagnosis								
Johnson (2010b) ¹²	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICURE Study).	219 (71%)	DAWBA. Parent report. DSM-IV-TR diagnosis assigned using scoring algorithm and clinical judgement of 2 blinded reviewers.	None.	Cognitive impairment at 6 yrs, pervasive peer problems at 6 yrs (SDQ).
Attention deficit/hyperactivity disorder: positive screen								
Hack (2009) ¹¹	United States 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	219 (97%)	CSI-4 based on DSM-IV-TR diagnostic criteria. Parent report.	Each risk factor was fitted separately and adjusted sex, race, parental SES (p-values not reported).	None significant for hyperactive, inattentive or combined type of ADHD.

^a Percentage of survivors assessed for outcome measure specified.

^b Familial social risk was based on a composite measure of six social risk factors: family structure, education of primary caregiver, occupation and employment status of primary income earner, language spoken at home and maternal age at birth.

^c 5 further models were reported: 3 models using only risk factors known at birth for the full cohort, excluding all disability and excluding neurosensory disability; 2 models with more stringent criteria for a positive screen.

Abbreviations: ADOS Autism Diagnostic Observation Scales;⁵¹ BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development;⁶⁴ BW birth weight; CBCL Child Behavior Checklist;⁴⁴ CP cerebral palsy; CSI-4 Parent Child Symptom Inventory;⁵³ CUSS cranial ultrasound abnormality; DAWBA Development And Wellbeing Assessment;⁴⁹ DSM-IV-TR Diagnostic and Statistical Manual of Mental Disorders;¹⁰ FUP follow up; GA gestational age; M-CHAT Modified Checklist for Autism in Toddlers;⁵² NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; NEC necrotizing enterocolitis; PDDST Pervasive Developmental Disorders Screening Test;⁵⁰ PC prospective cohort; Q-CHAT Quantitative Checklist for Autism in Toddlers;⁶⁵ SDQ Strengths and Difficulties Questionnaire;⁴³ SCQ Social Communication Questionnaire.⁶⁶