



Original Research

Open Access

Neurodevelopmental Outcomes of Extremely Preterm Infants at Two Years Corrected Age

Anderson I^{1*}, Ingram C² and Anderson R³

¹Department of Applied Sciences, Wellington Institute of Neonatology, New Zealand Scopus Author ID: 57422986634
Email: isla.anderson@wellington.edu
ORCID: <https://orcid.org/5844-0083-9450-3787>

²Department of Clinical Research, Auckland Center for Neonatal Research, New Zealand Scopus Author ID: 57888410180 Email: chloe.ingram@auckland.edu
ORCID: <https://orcid.org/9304-6881-3651-7497>

³Department of Clinical Research, Melbourne Institute of Child Health, Australia Scopus Author ID: 57194871208
Email: ruby.anderson@melbourne.edu
ORCID: <https://orcid.org/0066-2256-1845-9983>

***Correspondence to:** Isla Anderson, Department of Applied Sciences, Wellington Institute of Neonatology, New Zealand Scopus Author ID: 57422986634 Email: isla.anderson@wellington.edu

Received: 26 January 2026; **Accepted:** 13 April 2026; **Published:** 9 July 2026

Copyright: © 2026 IA, et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Extremely preterm birth, before 28 completed weeks of gestation, remains the leading contributor to early childhood neurodevelopmental impairment among survivors of neonatal intensive care, and two years corrected age has become the field's standard timepoint for a first formal assessment of cognitive, language, and motor outcome. This article synthesizes the literature on neurodevelopmental outcomes at this timepoint, drawing on the major international cohort studies -- EPICure, EPIPAGE-2, the Swedish EXPRESS study, and the Dutch EPI-DAF study among others -- that have defined the field's empirical baseline, alongside the assessment-instrument literature, the neonatal brain-injury literature, and the intervention trials with the clearest evidence of shifting two-year outcome. A recurring and troubling finding across these cohorts is that survival at the most immature gestations has improved substantially over the past three decades while the proportion of survivors with severe disability has, on the best available like-for-like evidence, remained largely unchanged. We give particular weight to how much of the reported variation in outcome rates across cohorts and calendar time reflects genuine clinical change versus changes in the assessment instruments themselves, to the prognostic gradient carried by intraventricular hemorrhage severity, and to the small number of interventions, chiefly antenatal magnesium sulfate and postnatal caffeine, with randomized evidence of a genuine two-year benefit. Read together, the evidence supports two-year assessment as a genuinely informative but non-final snapshot, leaving the disconnect between improving survival and stagnant severe-disability rates as the field's most pressing open question.

Keywords: extremely preterm infants; neurodevelopmental outcome; corrected age; intraventricular hemorrhage; Bayley Scales

1. Introduction

Two decades ago, a baby born at 24 weeks' gestation was, in most parts of the world, a rare and largely unsurvivable event outside a handful of specialist centers. Today, active resuscitation and intensive care routinely extend survival to infants born at the very edge of viability, and a clinician counseling parents in the delivery room is as likely to be asked about a two-year-old's expected vocabulary and gait as about the odds of the infant leaving the neonatal intensive care unit at all. That shift in the question being asked is itself a marker of progress, but it has also exposed how much less settled the neurodevelopmental side of the ledger is compared with the survival side. Mortality curves for extremely preterm infants — conventionally defined as birth before 28 completed weeks of gestation — have moved steadily in a favorable direction across successive birth cohorts, but the corresponding curves for moderate-to-severe neurodevelopmental impairment at early follow-up have been far more stubborn, and in some national cohorts have barely moved at all over intervals during which survival to discharge rose substantially [1].

Two years corrected age — that is, two years from the expected date of delivery rather than from actual birth — has become the *de facto* standard window for a first formal neurodevelopmental assessment of infants born extremely preterm, and for good reason: it is early enough to inform ongoing early-intervention referrals and family planning, late enough that gross motor, cognitive, and language domains have differentiated sufficiently for standardized testing to be meaningful, and it is the endpoint mandated or strongly encouraged by most national perinatal networks and clinical trial protocols, which is what has made it possible to compare results across cohorts, countries, and decades in the first place. The landmark EPICure study of infants born in the UK and Ireland in 1995, and its 2006 successor cohort, anchored a great deal of the subsequent literature to this timepoint, and successor programs in France (EPIPAGE-2), Sweden (EXPRESS), the Netherlands (EPI-DAF), and the United States (the NICHD Neonatal Research Network) have since built a substantial, internationally comparable evidence base around the same two-year window [2,3,4,5].

This article synthesizes that evidence base. It is not a report of new data collection; it is an attempt to pull

together what the cohort literature, the neonatal-morbidity literature, and the intervention literature collectively say about neurodevelopmental outcomes of extremely preterm infants specifically at two years corrected age — what proportion of survivors have impairment and of what severity, which neonatal exposures and brain injuries carry the strongest prognostic weight, how much the assessment tools themselves shape the numbers that get reported, and where genuine open questions remain rather than settled answers dressed up as consensus. We organize the discussion around five themes: the major international cohort studies that have defined the field's empirical baseline; the assessment instruments used to generate a two-year outcome, and their well-documented limitations; the neonatal brain injuries and comorbidities that drive prognostic heterogeneity within the extremely preterm population; the perinatal and neonatal interventions with the clearest evidence of shifting the two-year outcome distribution; and the open methodological and clinical challenges that, in our reading, keep this from being a fully solved problem despite three decades of dedicated cohort science.

It is worth being explicit about scope and its limits from the outset. Two-year outcomes are not destiny — a meaningful fraction of children classified as impaired at two years normalize by school age, and a meaningful fraction of children who look unimpaired at two years go on to manifest attention, executive-function, or learning difficulties only once academic demands increase [6]. We touch on that tension where the literature bears on it directly, but a full treatment of school-age and adolescent outcomes following extremely preterm birth is outside what this article attempts. Our focus stays deliberately on the two-year corrected-age window because that is where the largest, most internationally replicated body of comparable evidence exists, and because it remains the timepoint at which the great majority of clinical decisions about ongoing surveillance and early intervention actually get made.

2. Landmark Cohort Studies of Extremely Preterm Survivors

The single most influential dataset in this field remains the EPICure study, which enrolled essentially all infants born at 20 to 25 completed weeks of gestation in the UK and Ireland during 1995. Wood and colleagues' original report on this cohort, assessed at 30 months corrected

age, established what became the field's reference figures for the era: substantial rates of both cognitive delay and cerebral palsy among survivors born at the lowest gestations, with impairment rates that fell, but did not disappear, as gestational age increased even within the narrow 22-to-25-week band [2]. The same cohort was followed to six years of age, at which point Marlow and colleagues reported that a large majority of surviving children had some degree of cognitive, motor, or sensory impairment relative to term-born peers, a finding that was influential in shifting counseling practice away from treating two-year outcomes as a ceiling and toward treating them as a floor that some children would fall further below as cognitive demands increased with age [7].

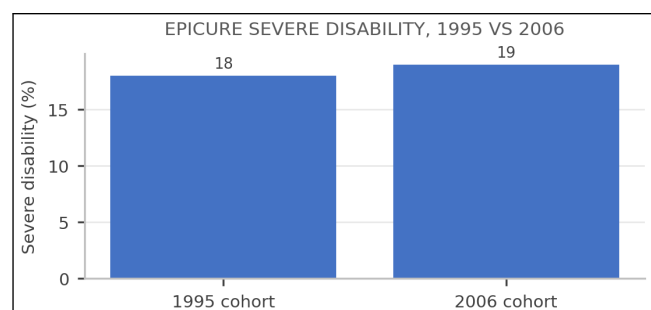


Figure 1: Severe neurodevelopmental disability among survivors born at 22-25 weeks' gestation in the EPICure 1995 and 2006 cohorts, applying the original 1995 disability definitions to both cohorts so the comparison is not confounded by assessment-tool changes -- survival to discharge rose substantially between the two epochs, but the severe-disability rate among survivors was essentially unchanged [8].

A second EPICure cohort, recruiting infants born at the same gestational range in England in 2006, allows a direct decade-on comparison that is unusually informative precisely because it holds geography and case definition constant while survival practice changed substantially in the interim. Moore and colleagues' comparison of the 1995 and 2006 cohorts found that survival to discharge among infants born at 22 to 25 weeks' gestation rose considerably between the two epochs, yet the proportion of survivors with severe neurodevelopmental disability — assessed using the original 1995 study's disability definitions so the two cohorts would be directly comparable — was essentially unchanged: 18% in the 1995 cohort versus 19% in the 2006 cohort [8]. That finding, more than almost any single number in this literature, captures the central tension that runs through the rest of this article: improved survival at the most immature gestations has not, on the evidence of this particular natural experiment, been accompanied by a proportional

improvement in the severe-disability rate among those who do survive. More infants are surviving; among survivors, the proportion who go on to have severe impairment has not obviously fallen.

Comparable national cohorts elsewhere tell a broadly consistent, if not identical, story. The Swedish EXPRESS study, a prospective, population-based cohort of infants born before 27 weeks' gestation, reported neurodevelopmental outcome at 2.5 years using the Bayley-III and found significantly lower cognitive, language, and motor scores among extremely preterm survivors relative to term-born comparison children, with impairment concentrated, as in the UK cohorts, at the lowest gestational ages [4,9]. In France, the EPIPAGE-2 cohort study assessed neurodevelopmental outcome at two years for a broader gestational range, 22 to 34 weeks, allowing the extremely preterm subgroup to be directly benchmarked against very and moderately preterm subgroups within the same national dataset and using the same assessment protocol, which sharpened the field's understanding of just how much of the overall preterm-birth disability burden is concentrated at the lowest gestational ages rather than distributed evenly across the preterm spectrum [3]. More recently, the Dutch EPI-DAF study examined outcomes at two years after national guidance lowered the threshold for offering active treatment from 25 to 24 completed weeks, providing an unusually direct test of what happens to the two-year outcome distribution when a health system deliberately extends active management to a more immature population rather than passively observing gestational-age trends in an unchanged treatment policy [5]. In the United States, Younge and colleagues examined survival and two-year neurodevelopmental outcome specifically among periviable infants, born at 22 to 24 weeks' gestation, across a multicenter NICHD network cohort spanning two decades, and found that both survival and survival without profound neurodevelopmental impairment rose across the study period even at this most immature end of the gestational-age range, a somewhat more optimistic finding than the flat EPICure severe-disability comparison and one that highlights how much the answer to "has outcome improved over time" can depend on which specific gestational-age band, country, and disability threshold a given analysis focuses on [10].

Reading these cohorts together rather than singly

clarifies two things. First, despite differences in national perinatal organization, resuscitation thresholds, and assessment tooling, the broad shape of the two-year outcome distribution for infants born under 26 weeks is remarkably consistent across high-income health systems: a substantial minority with moderate-to-severe impairment, a further substantial minority with mild impairment or scores in a borderline range, and a minority — smaller at the lowest gestations, larger as gestational age approaches 26 to 27 weeks — with scores in the typical range for age. Second, and more troublingly, none of these cohorts has produced clean evidence that the severe end of that distribution is shrinking over calendar time at a rate proportional to the rise in survival, a pattern Cheong and colleagues' review of temporal trends across multiple national datasets summarizes bluntly: survival has improved substantially since the 1990s, but there is no clear, consistent trend of improved neurodevelopment at the population level among survivors over that same interval, with the possible exception of a modest decline in cerebral palsy rates specifically [1].

3. Measuring Neurodevelopment at Two Years: Tools and Their Limits

Almost every number cited in the preceding section depends on a specific instrument, and the choice, edition, and administration of that instrument matters more to the reported outcome rate than is always appreciated by readers encountering a single cohort's headline figure in isolation. The Bayley Scales of Infant and Toddler Development, now in widespread use in its third and fourth editions, is the dominant tool for generating a standardized cognitive, language, and motor score at two years corrected age across essentially all of the cohorts discussed above. The transition from the second edition (Bayley-II) to the third (Bayley-III), which occurred across roughly the 2006-to-2012 period in most networks, was not a neutral methodological update: Moore and colleagues directly compared scores from the two editions administered to the same extremely preterm children and found that the Bayley-III systematically yields higher scores, and correspondingly lower rates of classified developmental delay, than the Bayley-II would have produced in the same children [11]. This has an uncomfortable implication for exactly the kind of decade-on comparison discussed in the previous section: some fraction of any apparent improvement in cognitive outcome rates across cohorts

assessed with different Bayley editions reflects a change in measurement, not a change in the children being measured, and studies that hold the assessment instrument constant across cohorts — as Moore and colleagues did by reapplying the original 1995 disability criteria to the 2006 cohort — are for this reason considerably more trustworthy for tracking genuine temporal change than studies that simply compare the headline disability rate each cohort reported using whatever tool was standard at the time [8].

Two-year assessment also has a more fundamental limitation that no instrument revision fixes: it is a snapshot of a rapidly developing nervous system at a point when several of the cognitive and behavioral domains most affected by extreme prematurity — attention regulation, executive function, specific learning skills — have not yet become testable in any reliable, standardized way. Rysavy and colleagues, examining the relationship between neurodevelopmental impairment classification and concurrent functional outcomes in extremely preterm children, underscore that the two-year impairment label, while clinically useful and strongly prognostic on average, is an imperfect proxy for how a given child will actually be functioning by the time school entry places genuinely new cognitive demands on them [6]. This does not argue against using two-year assessment — no later, more informative timepoint is available early enough to guide the early-intervention referrals and family counseling that actually happen in the first two years of life — but it does argue for describing what two-year assessment can and cannot tell a family with more precision than the phrase "normal outcome" typically conveys.

A related and underappreciated source of variability across cohorts is examiner training and site standardization. Multi-site, multinational cohort comparisons are only as comparable as the fidelity with which each site administers and scores the same instrument, and differences in examiner certification, testing environment, and even which corrected-age window a site uses to schedule the assessment can measurably shift a site's or a country's reported outcome rate independent of any genuine difference in the underlying population. This is one of the more mundane but persistent threats to validity running through the entire international-comparison literature, and it is one

reason the field places disproportionate weight on studies, like the EPICure decade comparison, that hold assessment methodology fixed across the groups being compared [8].

It is also worth noting what the Bayley-family instruments are, and are not, designed to detect. They generate composite cognitive, language, and motor scores calibrated against a normative population of term-born children, and a score in the impaired range signals that a child's performance on a structured set of age-referenced tasks fell meaningfully below that normative expectation. What a two-year Bayley assessment cannot do, by design, is distinguish reliably between a transient developmental lag that a child will close by school age and a persistent difficulty that will still be present, in a different and more academically consequential form, years later. Cohort studies that have gone on to reassess the same children at school age consistently find both false negatives, children who tested in the typical range at two years but were later identified with specific learning, attention, or executive-function difficulties, and false positives in the other direction, children whose early scores looked concerning but who caught up substantially by the time formal schooling began [6]. Interpreting a two-year Bayley result in isolation, without this context, risks conveying a false sense of either reassurance or alarm to families receiving the result in a clinical follow-up appointment.

4. Neonatal Brain Injury and Its Developmental Consequences

If gestational age at birth sets the broad statistical envelope of two-year risk, the specific pattern of neonatal brain injury a given infant sustains does much of the work of explaining why two infants born at the identical gestational age can have such different two-year trajectories. Volpe's influential reframing of preterm brain injury as an "encephalopathy of prematurity" — a combination of destructive lesions (intraventricular hemorrhage, periventricular white matter injury) and secondary disturbances of ongoing brain maturation affecting white matter, subplate neurons, thalamus, basal ganglia, and cerebellum — has shaped how the field now interprets neuroimaging findings in the extremely preterm population: not as a binary presence or absence of a discrete lesion, but as a graded continuum of destructive-and-dysmaturational

injury whose severity tracks reasonably well, though imperfectly, with two-year functional outcome [12].

Table 1: Representative quantitative findings on neonatal brain injury severity and a major neuroprotective intervention, drawn from the meta-analytic and trial literature synthesized in this article.

Exposure / Comparison	Source	Outcome	Reported Effect
Mild IVH vs. no IVH	Zhou et al. [13]	Neurodevelopmental impairment	OR 1.20 (95% CI 1.08-1.34)
Severe IVH vs. mild IVH	Zhou et al. [13]	Neurodevelopmental impairment	OR 1.78 (95% CI 1.44-2.20)
Severe IVH vs. mild IVH	Zhou et al. [13]	Cerebral palsy	OR 3.28 (95% CI 2.56-4.20)
Antenatal magnesium sulfate vs. placebo	Crowther et al. [15]	Cerebral palsy (neuroprotective-intent analysis)	RR 0.68 (95% CI 0.53-0.87), NNT -42

Germinal matrix-intraventricular hemorrhage (IVH) remains the single most widely used bedside prognostic marker, precisely because cranial ultrasound is universally available, inexpensive, and can be performed serially at the bedside without moving a critically ill infant. A recent meta-analysis by Zhou and colleagues, pooling outcome data from preterm and low-birth-weight infants with IVH across multiple studies, quantified the severity gradient with unusual precision: relative to infants with no IVH, mild IVH was associated with a significantly higher odds of neurodevelopmental impairment (odds ratio 1.20), while the step from mild to severe IVH carried a substantially larger additional odds increase (odds ratio 1.78) [13]. The same analysis found an even steeper gradient specifically for cerebral palsy — odds ratio 1.93 for mild versus no IVH, rising to 3.28 for severe versus mild IVH — and a comparably steep gradient for visual impairment, underscoring that IVH severity is not a uniform risk multiplier across developmental domains but interacts differently with motor, cognitive, and sensory pathways depending on where in the periventricular white matter and adjacent structures the hemorrhage and any associated parenchymal injury are concentrated [13].

Advances in structural neuroimaging have, over the past decade, begun to supplement this ultrasound-based staging with quantitative measures that capture the dysmaturational half of Volpe's encephalopathy-of-prematurity framework rather than only its destructive half. Jang and colleagues' recent

work applying multimodal structural MRI with local connectivity metrics to predict two-year neurodevelopmental outcome in preterm infants is representative of this direction: rather than classifying injury by discrete lesion grade alone, quantitative structural and connectivity features extracted from term-equivalent-age MRI are used as continuous predictors of subsequent Bayley performance, an approach that in principle captures gradations of white-matter and network maturation that a binary IVH grade or a qualitative white-matter-injury score cannot [14]. This line of work remains predominantly a research tool rather than routine bedside practice — the imaging, processing, and modeling pipelines involved are not yet standardized or validated across the range of scanner hardware and gestational ages a typical neonatal intensive care unit encounters — but it represents a plausible path toward earlier and more individualized two-year outcome prediction than lesion-grade staging alone can offer.

5. Perinatal and Neonatal Interventions Shaping Outcome

Unlike most of the risk factors discussed above, which are prognostic markers rather than modifiable exposures, a small number of interventions in this literature have randomized-trial evidence of actually shifting the two-year outcome distribution, and they deserve to be treated as a distinct category. The clearest example is antenatal magnesium sulfate administered to women at imminent risk of preterm birth specifically for fetal neuroprotection rather than for its older obstetric indications. Crowther and colleagues' individual-participant-data meta-analysis, pooling data across the major neuroprotection trials, found a strong protective effect against cerebral palsy in surviving children: a relative risk of 0.68 (95% confidence interval 0.54 to 0.87), corresponding to a number needed to treat of approximately 46 women to prevent one case of cerebral palsy among their children [15]. This is among the strongest pieces of level-one evidence anywhere in the extremely preterm outcomes literature for a single, low-cost, already widely available intervention meaningfully improving a two-year motor outcome, and it has driven the intervention's rapid incorporation into obstetric guidelines across most high-income health systems over the past decade.

Postnatal caffeine citrate therapy, originally introduced

simply to treat apnea of prematurity, turned out on longer follow-up to carry neurodevelopmental benefit that went beyond its original indication. The Caffeine for Apnea of Prematurity (CAP) trial, a large multicenter randomized trial, first established caffeine's short-term respiratory benefit — reduced rates of bronchopulmonary dysplasia and earlier discontinuation of positive airway pressure — in its initial report [16], and its planned 18-to-21-month follow-up went on to demonstrate a reduction in the composite outcome of death or survival with neurodevelopmental disability in the caffeine-treated group relative to placebo [17]. An unusually long subsequent follow-up of the same randomized cohort to age 11 years found that the trial's early academic-performance and motor advantages had, by that point, substantially attenuated, with the caffeine and placebo groups no longer clearly distinguishable on most academic and behavioral measures at school age [18]. Read together, the three CAP trial reports offer a genuinely instructive case study for the rest of this literature: an intervention with a real, randomized, two-year neurodevelopmental benefit whose long-term durability turned out to be considerably less than the early follow-up alone would have suggested, a pattern that should temper how confidently any two-year finding — favorable or unfavorable — is generalized to school-age or later outcomes without dedicated longer-term follow-up data.

Beyond these two pharmacological and pharmaco-obstetric interventions, a substantial parallel literature has developed around care-delivery interventions — kangaroo (skin-to-skin) care, individualized developmental care programs, post-discharge parenting and stimulation interventions — that are lower-cost, more broadly implementable across resource settings, and in some cases show measurable benefit on early neurodevelopmental trajectories, though generally with smaller and more heterogeneous effect sizes across trials than the magnesium sulfate and caffeine evidence base, and with outcome measures that are not always harmonized closely enough across trials to support the kind of individual-participant-data pooling that has been possible for magnesium sulfate specifically. We return to the more general question of how much of the residual, non-pharmacological variance in two-year outcome is realistically modifiable through care practice, as distinct

from being set by the severity of neonatal illness itself, in the discussion of open challenges below.

The EPI-DAF study is a useful bridge between the cohort-comparison literature discussed in the previous section and this section's focus on interventions, because it is, in effect, a natural experiment in a policy-level intervention rather than a randomized trial of a specific drug or device. By examining outcomes at two years after Dutch national guidance deliberately lowered the threshold for offering active perinatal management from 25 to 24 completed weeks, the study allows the field to ask what happens to the population-level two-year outcome distribution when a health system's counseling and resuscitation practice, rather than any single neonatal treatment, is the thing that changes [5]. This kind of policy-level natural experiment is harder to interpret cleanly than a randomized trial, because it changes which infants are born into the eligible cohort in the first place rather than randomizing a fixed population to two treatment arms, but it is arguably a more externally valid test of what actually happens when a whole country's practice shifts, and it belongs in the same conversation as the drug- and device-level trials rather than being read as a separate, purely descriptive cohort exercise.

6. Comorbidities, Modifiers, and Sources of Heterogeneity

Extremely preterm infants who survive to two years of age rarely do so without at least one major neonatal morbidity along the way, and several of these morbidities carry independent prognostic weight for neurodevelopment beyond what gestational age and brain-injury grade alone predict. Bronchopulmonary dysplasia (BPD), the chronic lung disease of prematurity, is among the most extensively studied of these comorbidities, in part because its severity grading is itself a marker of how much of the neonatal course was spent in a state of chronic hypoxemia, growth restriction, and prolonged ventilatory support — all plausible independent contributors to impaired neurodevelopment beyond the pulmonary system itself. Within the specific population of very low birth weight infants who develop BPD, a systematic review and meta-analysis by Gao and colleagues examined whether postnatal corticosteroid treatment for established BPD affected later cognitive outcome, and found a significant reduction in the risk of cognitive impairment among

steroid-treated infants relative to untreated controls within this BPD-affected population (relative risk 0.83, 95% confidence interval 0.72 to 0.96) [19] — a finding that is informative less as a general statement about BPD's developmental impact than as an illustration of how much heterogeneity in the BPD literature itself hinges on that population's other, co-occurring exposures and treatments, a caution that generalizes to most of the comorbidity literature discussed in this section.

Maternal and pregnancy-level factors also shape the population of extremely preterm survivors in ways that ripple forward into two-year outcome, sometimes through pathways that are still being worked out. Chawla and colleagues, examining in-hospital mortality and morbidity among extremely preterm infants in relation to maternal body mass index, found that survival to discharge was measurably lower among infants born to mothers with obesity compared with those born to mothers of overweight or normal body mass index, even after adjustment for the confounders typically considered in this literature, although survival without major in-hospital morbidity did not differ as sharply across BMI categories [20]. Because in-hospital survival and morbidity are themselves the gateway that determines which infants are even eligible to be assessed at two years, findings like this reshape the two-year outcome literature indirectly, by shifting the composition of the surviving population that later gets counted, rather than only through a direct biological pathway from maternal BMI to the child's own neurodevelopment.

Sex is one of the more consistently reported, if still incompletely explained, sources of outcome heterogeneity within the extremely preterm population. A recent nationwide Korean cohort study by Yang and colleagues, examining gestational-age-specific sex differences in neurological outcome among very low birth weight preterm infants, found that the male disadvantage long reported in this literature — boys born extremely preterm tend to fare worse neurodevelopmentally than girls born at the same gestational age — is not uniform across the gestational-age range but varies in magnitude depending on how immature the infant is at birth, a nuance that a simple pooled male-versus-female comparison across a wide gestational-age band would

obscure [21]. This kind of gestational-age-stratified analysis is a useful methodological model for the field more broadly: many of the modifiers discussed in this section likely interact with gestational age itself rather than acting as simple, additive risk factors, and cohort analyses that report only a pooled effect across the full extremely preterm range risk averaging away exactly the interaction that would be most clinically informative.

None of the individual modifiers reviewed in this section — BPD severity, maternal BMI, infant sex — operates in isolation on a given child's two-year trajectory, and one of the more persistent methodological difficulties in this literature is separating a modifier's direct biological contribution to neurodevelopment from its indirect contribution through which infants survive to be assessable at two years in the first place. An infant's sex, BPD status, and the metabolic environment of the pregnancy that produced them are all entangled, to varying degrees, with survival itself, gestational age at delivery, and the likelihood of severe brain injury occurring in the first place, which means a modifier that looks prognostically important in a simple bivariate comparison can shrink substantially, or in some analyses disappear altogether, once gestational age and injury severity are accounted for as covariates. This is not a reason to discount the modifiers discussed here — several, sex in particular, have survived rigorously adjusted analysis across multiple independent cohorts — but it is a reason to read any single unadjusted comparison in this literature with some caution until a comparably designed, adjusted analysis has reproduced it.

7. Open Challenges

Several gaps recur across the literature synthesized here, and naming them plainly is more useful to a reader than leaving them implicit. First, the disconnect between improving survival and stagnant severe-disability rates, most starkly documented in the EPICure decade comparison, remains substantially unexplained [8]. It is not obvious whether this reflects a genuinely fixed biological ceiling on how much intact survival is achievable at the most extreme gestations regardless of care quality, a "more fragile survivors" compositional effect in which improved resuscitation and intensive care are preferentially extending the survival of infants who would previously have died and who happen to carry higher baseline neurodevelopmental risk, or

simply that the specific interventions responsible for the survival gains to date (better respiratory support, improved nutrition, infection control) are not the same interventions that would be needed to move the neurodevelopmental needle, and that neuroprotection-specific measures like antenatal magnesium sulfate have not yet been in widespread use for long enough, or are not powerful enough on their own, to show up in whole-cohort disability statistics. Disentangling these explanations would require exactly the kind of detailed, standardized, multi-decade cohort data that only a handful of research groups worldwide are positioned to collect, and even EPICure's own comparison, valuable as it is, spans only two cohorts a decade apart rather than a genuinely continuous trend line.

Second, measurement instability across the Bayley-II-to-Bayley-III-to-Bayley-4 transitions, and across national and linguistic adaptations of these instruments, continues to make like-for-like comparison across cohorts and calendar time genuinely difficult, and Moore and colleagues' direct demonstration that the same children score differently on the two Bayley editions is a caution the field has, in our reading, not fully internalized when headline disability rates from different eras are compared casually in review articles and clinical counseling materials [11]. Third, the field's growing interest in quantitative neuroimaging-based prediction, exemplified by work like Jang and colleagues' structural-connectivity modeling, is scientifically promising but has not yet been validated at the scale, or with the cross-site standardization, needed to move from a research finding to a bedside prognostic tool that a clinician could responsibly use in family counseling [14]. Fourth, most of the intervention evidence reviewed here — magnesium sulfate and caffeine specifically — was generated in trials conducted predominantly in high-income health systems with resuscitation and intensive-care infrastructure that is far from universally available; the degree to which these interventions' two-year benefits generalize to lower-resource settings, where the population of infants surviving to be eligible for two-year assessment in the first place looks quite different, remains comparatively under-studied. Fifth, and perhaps most consequential for how the two-year literature should be read by clinicians and families alike, the CAP trial's 11-year follow-up demonstrating attenuation of an intervention's early benefit is a

reminder that two-year outcome, for all its practical value as an early counseling and surveillance timepoint, is not a stable endpoint in the sense that a single measurement fully determines a child's longer-term trajectory [18,6]. We do not think any of these gaps undermine the substantial value of two-year assessment as currently practiced — it remains the field's most internationally comparable, most extensively validated early outcome measure, and it demonstrably shapes referral and support decisions that matter to families during exactly the developmental window when early intervention has the best evidence of effectiveness — but they are the honest state of the evidence, and future work claiming a settled, fully explained account of extremely preterm neurodevelopmental outcome should be read with these open questions in mind.

8. Conclusion

The literature reviewed here supports a fairly specific set of claims. Extremely preterm birth carries a substantial, well-replicated burden of neurodevelopmental impairment at two years corrected age, concentrated at the lowest gestational ages and graded closely, though imperfectly, with the severity of neonatal brain injury as staged by cranial ultrasound and increasingly by quantitative structural MRI [2,12,13,14]. Survival at the most immature gestations has improved substantially across the past three decades of neonatal intensive care, but the proportion of survivors with severe disability has, on the clearest available like-for-like evidence, not improved proportionally over the same interval, a finding that should temper any assumption that improving survival and improving population-level neurodevelopmental outcome are the same achievement [8,1]. A small number of interventions — antenatal magnesium sulfate most clearly, postnatal caffeine with an important caveat about durability — have randomized-trial evidence of shifting the two-year outcome distribution in a favorable direction, and represent the field's clearest examples of modifiable, rather than merely prognostic, factors [15,17,18].

None of this argues for abandoning two-year assessment in favor of waiting for a more definitive later timepoint; no later timepoint is available early enough to guide the referrals, therapies, and family counseling that matter most during the developmental window when early intervention has the clearest evidence of benefit, and every alternative to a two-year assessment carries costs

of its own, chiefly the years of missed intervention that waiting would impose on children who do, in fact, have a persistent impairment. The more constructive response, in our reading of this literature, is for clinicians, families, and cohort investigators alike to treat a two-year Bayley classification as a probabilistic, revisable estimate rather than a final verdict, to weight it accordingly higher for children with more severe brain injury on neuroimaging and lower for children whose injury markers were mild or absent, and to build the longitudinal infrastructure needed to revisit that estimate at school age as a matter of course rather than only for the subset of children whose two-year result already raised concern.

Making full use of this evidence base depends on treating the two-year assessment as what it actually is: a genuinely informative, internationally comparable, but imperfect and non-final snapshot, sensitive to the specific instrument and edition used to generate it, prognostic on average but not deterministic for any individual child, and best interpreted alongside detailed neonatal-course and brain-injury information rather than as a single number in isolation [11,6]. The consistent convergence of independent national cohorts — EPICure, EPIPAGE-2, EXPRESS, EPI-DAF, and the NICHD Neonatal Research Network among them — on a broadly similar shape for this outcome distribution, despite real differences in resuscitation practice and health-system organization, is in our view the strongest available evidence that the field has identified a genuine, replicable phenomenon rather than an artifact of any single cohort's methodology [4,3,5]. That convergence is a solid foundation. Closing the gap between improving survival and stagnant severe-disability rates, refining measurement so that comparisons across cohorts and calendar time mean what they appear to mean, and extending the intervention evidence base beyond the small number of trials with genuinely randomized, two-year neurodevelopmental endpoints remain, in our reading, the field's most pressing unfinished business.

9. References

1. Cheong, J.L.Y., Spittle, A.J., Burnett, A.C., Anderson, P.J., Doyle, L.W. (2020) Have Outcomes Following Extremely Preterm Birth Improved Over Time? *Seminars in Fetal and Neonatal Medicine*, 25(3): 101114. DOI: [10.1016/j.siny.2020.101114](https://doi.org/10.1016/j.siny.2020.101114)
2. Wood, N.S., Marlow, N., Costeloe, K., Gibson, A.T., Wilkinson, A.R. (2000) Neurologic and Developmental Disability after Extremely Preterm Birth. *New England Journal of*

- Medicine, 343(6): 378-384. DOI: [10.1056/NEJM200008103430601](https://doi.org/10.1056/NEJM200008103430601)
3. Pierrat, V., Marchand-Martin, L., Marret, S., Arnaud, C., Benhammou, V., Cambonie, G., et al. (2017) Neurodevelopmental Outcome at 2 Years for Preterm Children Born at 22 to 34 Weeks' Gestation in France in 2011: EPIPAGE-2 Cohort Study. *BMJ*, 358: j3448. DOI: [10.1136/bmj.j3448](https://doi.org/10.1136/bmj.j3448)
 4. Serenius, F., Källén, K., Blennow, M., Ewald, U., Fellman, V., Holmström, G., et al. (2013) Neurodevelopmental Outcome in Extremely Preterm Infants at 2.5 Years After Active Perinatal Care in Sweden. *JAMA*, 309(17): 1810-1820. DOI: [10.1001/jama.2013.3786](https://doi.org/10.1001/jama.2013.3786)
 5. van Beek, P.E., Rijken, M., Broeders, L., ter Horst, H.J., Koopman-Esseboom, C., de Kort, E., et al. (2022) Two-Year Neurodevelopmental Outcome in Children Born Extremely Preterm: The EPI-DAF Study. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 107(5): F467-F474. DOI: [10.1136/archdischild-2021-323124](https://doi.org/10.1136/archdischild-2021-323124)
 6. Rysavy, M.A., Colaizy, T.T., Bann, C.M., DeMauro, S.B., Duncan, A.F., Brumbaugh, J.E., et al. (2021) The Relationship of Neurodevelopmental Impairment to Concurrent Early Childhood Outcomes of Extremely Preterm Infants. *Journal of Perinatology*, 41(9): 2270-2278. DOI: [10.1038/s41372-021-00999-7](https://doi.org/10.1038/s41372-021-00999-7)
 7. Marlow, N., Wolke, D., Bracewell, M.A., Samara, M. (2005) Neurologic and Developmental Disability at Six Years of Age after Extremely Preterm Birth. *New England Journal of Medicine*, 352(1): 9-19. DOI: [10.1056/NEJMoa041367](https://doi.org/10.1056/NEJMoa041367)
 8. Moore, T., Hennessy, E.M., Myles, J., Johnson, S.J., Draper, E.S., Costeloe, K.L., Marlow, N. (2012) Neurological and Developmental Outcome in Extremely Preterm Children Born in England in 1995 and 2006: The EPICure Studies. *BMJ*, 345: e7961. DOI: [10.1136/bmj.e7961](https://doi.org/10.1136/bmj.e7961)
 9. Månsson, J., Fellman, V., Stjernqvist, K. (2014) Children Born Extremely Preterm Show Significant Lower Cognitive, Language and Motor Function Levels Compared with Children Born at Term, as Measured by the Bayley-III at 2.5 Years. *Acta Paediatrica*, 103(5): 504-511. DOI: [10.1111/apa.12585](https://doi.org/10.1111/apa.12585)
 10. Younge, N., Goldstein, R.F., Bann, C.M., Hintz, S.R., Patel, R.M., Smith, P.B., et al. (2017) Survival and Neurodevelopmental Outcomes among Periviable Infants. *New England Journal of Medicine*, 376(7): 617-628. DOI: [10.1056/NEJMoa1605566](https://doi.org/10.1056/NEJMoa1605566)
 11. Moore, T., Johnson, S., Haider, S., Hennessy, E., Marlow, N. (2012) Relationship between Test Scores Using the Second and Third Editions of the Bayley Scales in Extremely Preterm Children. *Journal of Pediatrics*, 160(4): 553-558. DOI: [10.1016/j.jpeds.2011.09.047](https://doi.org/10.1016/j.jpeds.2011.09.047)
 12. Volpe, J.J. (2009) Brain Injury in Premature Infants: A Complex Amalgam of Destructive and Developmental Disturbances. *The Lancet Neurology*, 8(1): 110-124. DOI: [10.1016/S1474-4422\(08\)70294-1](https://doi.org/10.1016/S1474-4422(08)70294-1)
 13. Zhou, M., Wang, S., Zhang, T., Duan, S., Wang, H. (2024) Neurodevelopmental Outcomes in Preterm or Low Birth Weight Infants with Germinal Matrix-Intraventricular Hemorrhage: A Meta-Analysis. *Pediatric Research*, 95(3): 625-633. DOI: [10.1038/s41390-023-02877-8](https://doi.org/10.1038/s41390-023-02877-8)
 14. Jang, Y.H., Ham, J., Kasani, P.H., Kim, H., Lee, J.Y., Lee, G.Y., Han, T.H., Kim, B.-N., Lee, H.J. (2024) Predicting 2-Year Neurodevelopmental Outcomes in Preterm Infants Using Multimodal Structural Brain Magnetic Resonance Imaging with Local Connectivity. *Scientific Reports*, 14(1): 9331. DOI: [10.1038/s41598-024-58682-8](https://doi.org/10.1038/s41598-024-58682-8)
 15. Crowther, C.A., Middleton, P.F., Voysey, M., Askie, L., Duley, L., Pryde, P.G., Marret, S., Doyle, L.W. (2017) Assessing the Neuroprotective Benefits for Babies of Antenatal Magnesium Sulphate: An Individual Participant Data Meta-Analysis. *PLOS Medicine*, 14(10): e1002398. DOI: [10.1371/journal.pmed.1002398](https://doi.org/10.1371/journal.pmed.1002398)
 16. Schmidt, B., Roberts, R.S., Davis, P., Doyle, L.W., Barrington, K.J., Ohlsson, A., Solimano, A., Tin, W. (2006) Caffeine Therapy for Apnea of Prematurity. *New England Journal of Medicine*, 354(20): 2112-2121. DOI: [10.1056/NEJMoa054065](https://doi.org/10.1056/NEJMoa054065)
 17. Schmidt, B., Roberts, R.S., Davis, P., Doyle, L.W., Barrington, K.J., Ohlsson, A., Solimano, A., Tin, W. (2007) Long-Term Effects of Caffeine Therapy for Apnea of Prematurity. *New England Journal of Medicine*, 357(19): 1893-1902. DOI: [10.1056/NEJMoa073679](https://doi.org/10.1056/NEJMoa073679)
 18. Schmidt, B., Anderson, P.J., Doyle, L.W., Dewey, D., Grunau, R.E., Asztalos, E.V., et al. (2017) Academic Performance, Motor Function, and Behavior 11 Years After Neonatal Caffeine Citrate Therapy for Apnea of Prematurity: An 11-Year Follow-up of the CAP Randomized Clinical Trial. *JAMA Pediatrics*, 171(6): 564-572. DOI: [10.1001/jamapediatrics.2017.0020](https://doi.org/10.1001/jamapediatrics.2017.0020)
 19. Gao, S., Zhang, X., Du, W., Zhou, X., Xi, Y., Ju, R. (2021) Systematic Review and Meta-Analysis: The Effect of Bronchopulmonary Dysplasia on Neurodevelopment in Very Low Birth Weight Premature Infants. *Translational Pediatrics*, 10(11): 3023-3033. DOI: [10.21037/tp-21-449](https://doi.org/10.21037/tp-21-449)
 20. Chawla, S., Laptook, A.R., Smith, E.A., Tan, S., Natarajan, G., Wyckoff, M.H., et al. (2021) In-Hospital Mortality and Morbidity among Extremely Preterm Infants in Relation to Maternal Body Mass Index. *Journal of Perinatology*, 41(5): 1014-1024. DOI: [10.1038/s41372-020-00847-0](https://doi.org/10.1038/s41372-020-00847-0)
 21. Yang, J.Y., Youn, Y.M., Kang, J.I., Han, Y.J., Lee, D.K., Shim,

S.-Y. (2025) Gestational Age-Specific Sex Difference in Neurological Outcomes of Very Low Birth Weight Preterm Infants -- A Nationwide Study. Scientific Reports, 15(1): 39939. DOI: [10.1038/s41598-025-23735-z](https://doi.org/10.1038/s41598-025-23735-z)

10. Declarations

Funding

This literature synthesis received no specific grant or funding from agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no competing interests relevant to the content of this article.

Data Availability Statement

This is a literature synthesis and no new data were generated or analyzed by the authors; readers are referred to the individual cited sources for their respective data and code availability statements.

Ethical Approval

Not applicable. This article did not involve the collection of primary data, and no human participants or animal subjects were recruited by the authors.

Citation: Anderson I, Ingram C, Anderson R. "Neurodevelopmental Outcomes of Extremely Preterm Infants at Two Years Corrected Age." *Pediatr Neonatal Res J* (2026): 001-011 doi: [10.5281/zenodo.22849496](https://doi.org/10.5281/zenodo.22849496)