Developmental trajectories of infants born at less than 30 weeks gestation on the Bayley-III scales

Chloe Durrant[1], Hilary S. Wong[2], Tim J. Cole[1], Betty Hutchon[3], Lynn Collier[3], Anna Wright[3], Cara George[3], Michelle de Haan[1] & Ángela Huertas[3]

UCL Great Ormond Street Institute of Child Health[1], University of Cambridge[2], University College London Hospital[3]

Ángela Huertas - Corresponding Author
    a.huertas-ceballos@nhs.net
    Neonatal Service, University College London Hospital
    2nd floor North Wing
    250 Euston Road
    London
    NW1 2PG
    https://orcid.org/0000-0002-6608-3772-

Word Count – Abstract: 209
            Main text: 2,334

Objective: To describe the cognitive, language and motor developmental trajectories of children born very preterm, and to identify perinatal factors that predict the trajectories.

Design: Data from a cohort of 1,142 infants born at <30 weeks' gestation who were prospectively assessed on the Bayley Scales of Infant and Toddler Development, third edition (Bayley-III) at 3, 6, 12 and 24 months corrected age, were analysed using the Super Imposition by Translation and Rotation (SITAR) growth curve analysis model.

Main outcome measures: Developmental trajectory SITAR models for Bayley-III cognitive, language (receptive and expressive communication sub-scales) and motor (fine and gross motor sub-scales) scores.

Results: The successfully fitted SITAR models explained 62% of variance in cognitive development, 68% in receptive communication, 53% in fine motor and 68% in the gross motor development. There was too much variation in the expressive communication, sub-scale to fit a SITAR model. The rate of development (gradient of the curve) best explains the variation in trajectories of development in all domains. Lower gestational age, lower birth weight and male sex significantly predicted a slower rate of development.

Conclusions: The rate of development, rather than single time point developmental assessment, best predict the very preterm infant's developmental trajectory and should be the focus for monitoring and early intervention.

What is known about the topic?

- Research on the neurodevelopmental outcomes of infants born preterm typically assess and report outcomes at single time points.

- Neurodevelopment is a dynamic process and assessment of developmental trajectories is more informative in identifying different types of developmental delay.

- The variation in developmental trajectories for the preterm population is unknown.

What does the study add?

- The rate of development over time, rather than developmental score in early childhood, account for the variation in neurodevelopmental outcomes in the very preterm population.
- We describe the normative data approximation of developmental trajectories in the cognitive, receptive communication, fine motor and gross motor sub-scales for very preterm infants.
- The normative data can be used as a reference to aid early identification of deviation from the expected developmental trajectory in this population.

In England and Wales, 8% of infants are born preterm (before 37 weeks gestation).[1] These infants are at higher risk for negative neurodevelopmental outcomes than those born at term.[2,3] The majority of neurodevelopmental outcomes research on infants born preterm has reported outcomes at single time points. As the impact of developmental disorders may be different at different stages of development, there is increasing emphasis on developmental studies to examine development as dynamic processes, and to move away from a focus on static single timepoints.[4–6] The assumption often made is that if two children score in the normal range they are using the same processes to obtain those scores. However, trajectories measured over time may show different developmental routes which may mean different processes lead to scores in the normal range[5] Developmental trajectories can also be highly informative in identifying different types of developmental delay, such as between children with a stable developmental delay and those who are declining.[7] It is therefore important not just to compare an atypical population to a typical population but also to understand the variation within an atypical population. The typical developmental trajectory during infancy, and the underlying variations, for children born very preterm (<30 weeks gestation) are unknown.

SuperImposition by Translation and Rotation (SITAR) growth curve analysis is used to model trajectories. SITAR analysis has previously been used to describe the postnatal weight gain

and construct longitudinal gestational age-specific growth curves for infants born preterm.[8,9] We aim to model the developmental trajectories of children born very preterm based on their scores on the Bayley Scales of Infants and Toddler Development, third edition (Bayley-III), using SITAR growth curve analysis, and to identify perinatal factors that predict the trajectories.

Method

We used data from a historical cohort of infants born at less than 29 weeks gestation between 1st January 2005 and 31st December 2013 and at less than 30 weeks between 1st January 2014 and 11th April 2017 in five neonatal units in the North Central London Perinatal Network, United Kingdom. These infants were invited to attend developmental assessments using the Bayley-III [10] at 3, 6, 12 and 24 months corrected (post-term) age as part of routine clinical follow-up. The assessments were completed by a small team of trained assessors who had consistently achieved inter-rater reliability of greater than 90%. The data of participants who attended and had scores recorded for at least two Bayley-III assessments were included. The study was categorised as an audit and service evaluation by the University College London Hospital Research and Development department. Consent to use data for these purposes were received from parents of infants attending the follow-up assessments.

Clinical data including the sex, gestational age, birth weight, multiplicity (singleton, twin or triplet) and neonatal morbidities of the infants were retrieved from medical records. The data were cleaned prior to analysis and summary statistics were calculated.

The raw Bayley-III scores from the cognitive scale, receptive and expressive communication sub-scales and fine and gross motor sub-scales were used for the analysis, so that each infant's set of scores increased with age in a trajectory like a growth curve. SITAR growth curve analysis was used to analyse these trajectories. SITAR does two things: it estimates

the mean trajectory as a cubic spline curve and it summarises in a simple way how individual trajectories differ from the mean trajectory.[8] Cubic splines are flexible smooth curves that are made up of sections of cubic curves that are joined together to be smooth at the join points. The number of sections joined together (the degrees of freedom) determines how complex the curve shape is. SITAR assumes that an individual trajectory differs from the mean in three respects, random effects which are termed size (the intercept, i.e. the individual curve is higher/lower than the mean), timing (the curve is shifted left/right relative to the mean), and intensity (the curve is steeper/shallower than the mean). If the model fits well, the individual curves adjusted for the random effects will be superimposed on the mean curve. Separate SITAR models were calculated for each scale. The analysis was carried out using the statistical computer program `R` and the `sitar` package.[11,12]

Multiple regression analysis was used to test whether sex, gestational age, birth weight and multiplicity predicted the individual random effects for each SITAR model. Due to insufficient power, subgroup analyses of SITAR models stratified by these perinatal factors were not performed.

Results

Figure 1 shows the study participation flow chart and the characteristics of the participants are summarised in table 1. Data from 1,142 infants (586 male) were analysed, with mean gestational age 27.1 (SD 1.8) weeks and mean birth weight 944g (256g). 295 participants attended two assessments, 389 attended three and 458 attended all four. The mean Bayley-III composite scores achieved by the participant at each time point are presented in table 2.

SITAR models were successfully fitted on the cognitive, receptive communication, fine motor and gross motor scales, but the expressive communication model failed to converge. The

optimal models all included only the intensity random effect, not the size effect, and the mean spline curve was fitted with just two degrees of freedom, the simplest possible nonlinear curve. The size effect corresponds to a random intercept, so that omitting it means that individuals all shared the same intercept, with the fitted curves all fanning out from the same point at age 0.

Focussing first on the cognitive scale, five outliers were excluded with standardised residuals exceeding ± 4. Figure 2 illustrates how the model works. In Figure 2a the raw trajectories for each individual are shown colour coded to distinguish between them, and they can be seen to fan out over time, with some slopes steeper than others. The model estimates these individual slopes and compares them to the mean slope (mean fractional difference 0, SD 0.16). Simultaneously the model estimates the mean curve, as shown in Figure 2b. Figure 2c shows the predicted curves for each individual, which can be compared with the raw curves in Figure 2a. Figure 2d shows how the variability in the raw curves of 2a (in grey) is reduced when the individual differences in curve slope are adjusted for (in colour). The adjustment process reduces the variance between the raw curves by 62% (Table 3). The key assumption of the SITAR model is that all the individual curves are essentially the same shape, so curves that are unusual in shape will generally fit less well.

The predicted individual (coloured) and mean (black) curves for the receptive communication, fine motor and gross motor scales are shown in Figure 3 The final models accounted for 68% of variance in reception communication, 53% in fine motor and 68% in gross motor (Table 3). The variance values are a numerical summary of the reduction in scatter seen in Figure 2, comparing the unadjusted curves in grey with the adjusted curves in colour. The lower value for fine motor suggests that fine motor is noisier and harder to measure.

Gestational age, birth weight and sex independently predicted the individual intensity random effects in the cognition, receptive communication, fine motor and gross motor scales (Table 3). Being a twin also influenced cognitive, fine motor and gross motor developmental trajectories.

Discussion

The study used SITAR analysis to calculate mean developmental trajectories for the cognition, receptive communication, fine motor and gross motor scales on the Bayley-III in children born at less than 30 weeks gestation. For all four scales, the variation in trajectory between individuals was best explained by a model considering only their intensity random effects, i.e. their slopes and not the intercept., This means that the individual predicted curves all start at the same point at age 0, and splay out with increasing age. The implication of this is that the variation in outcome between individuals is best explained by the rate of development, not the initial value. This is clinically significant as it suggests the development score at the first assessment is not the most useful predictor of developmental trajectories. Instead, how infants develop over time is the best predictor of their developmental outcome. Therefore, sequential assessments are necessary in clinical practice to monitor the developmental progress of very preterm infants. Nevertheless, whilst the initial assessment may not be the most useful predictor it is still of value in identifying atypical development to allow for early intervention and gives a first score which can then be used to examine rate of development over time. The results suggest future research should focus on resilience factors and how interventions can be designed to increase the rate of development, i.e. the trajectory intensity.

The approach of this study differed from the small number of previous studies that had researched neurodevelopmental trajectories in children born preterm. Previous approaches have usually involved comparing trajectories of children born preterm to those born at

term.[13–15] For example, Yaari et al. [13] assessed groups of infants on the Mullen Scale of Early Learning (MSEL) and found that those born at less than 34 weeks gestation scored lower than those born at term with the gap increasing over infancy. Linsell et al. [15] compared cognitive scores of children born at less than 26 weeks and children born at term measured at 2.5, 6, 11 and 19 years of age. Children born preterm scored significantly lower than those born at term across all age assessments and showed little evidence of catching up. Sansavini et al. [14] calculated trajectories on the Bayley-III, from measurements at 12, 24 and 30 months corrected age for children born extremely ($\leq$28 weeks gestation) preterm and full term. They found children born extremely preterm had a lower trajectory than those born at term, the gap remaining stable over time on the cognitive and language scales but increasing on the motor scales. However, they assessed only 17 children born extremely preterm and 11 born at term. The current study used a large population to model the developmental trajectories of children born very preterm rather than compare them to children born at term. However, Figure 3 shows the developmental age equivalent scores from the Bayley-III to generally be higher than the average trajectory modelled for children born very preterm, supporting the findings of this previous research. Figure 3 also highlights the variation in the individual trajectories of children born very preterm both above and below the developmental age equivalent scores. Wang et al [16] described this variation in a large sample of 887 very low birth weight infants (<1500g) who were assessed sequentially between 6 and 24 months. Five patterns of cognitive trajectories were identified: average-stable (20% of children), average-decline to borderline delay (34%), borderline delay-catch-up to average (20%), borderline delay-decline to significant delay (17%) and significant delay-stable (8%).

To our knowledge, this is the first study to fit a mean developmental curve and individual predictive developmental curves for infants born very preterm. The strengths of the study

include the use of a large cohort of infants who were assessed longitudinally with multiple assessments undertaken by a small team of experienced, trained clinicians, minimising inter-rater variability.

One important consideration is that the data were collected at routine follow-up assessments, where the purpose was to identify infants requiring early intervention and support. The impact of any instituted intervention is not known. Only 40% of infants attended all four assessments, which may have contributed to selection bias. A term group was not available for comparison; however, the focus of the study was to describe what is typical for this atypical population. General issues with the Bayley-III have been widely discussed in the literature.[17–19] However, these issues are less applicable to the present study as only the growth element of the scores was considered. Therefore, the relationship to previous versions of the Bayley or developmental impairment boundaries is of less importance.

We did not have sufficient power in our current study to undertake subgroup analyses as well as to examine the dynamic influence of postnatal risk factors such as neonatal morbidities, socioeconomic status and maternal mental health on developmental trajectories. Research has shown that these postnatal events alter prognosis over time and may account for the large variability about the SITAR mean curve. Therefore, whilst it is possible to obtain an anticipated developmental trajectory using two sequential assessment on the current models, we do not recommend using these curves for clinical prediction purposes. Future prediction modelling that combines postnatal influences may aid to develop individualised predictive outcome trajectories. [13,16,20–22]

Our study population was from a geographical area with high levels of multilingualism. A 2011 census of the five local authorities the study population covers found only 75-86%, and in one area less than 75%, of people speaking English as their main language, compared to the national average of 92%.[23] This may have contributed to the large variation observed in the

Bayley-III expressive communication sub-scale in this cohort, resulting in a SITAR model not being successfully fitted. Infants born preterm with bilingual parents score lower in cognitive and language developmental assessments.[24] As the administration of the Bayley-III language scales requires specific phrasing and translation it is likely that the inclusion of a high proportion of infants exposed to multilingualism alongside monolingual children contributed to the variation. We did not have data on language exposure in the infants to assess this effect. We also acknowledge that the generalisability of our findings to areas with lower rates of multilingualism may be limited.

In conclusion, the study is believed to be the first to use a large cohort to calculate predicted mean developmental trajectories of children born at less than 30 weeks gestation for the cognitive, receptive communication, fine motor and gross motor scales on the Bayley-III. The rate of development over time best explains the variation in outcome. This highlights the importance of repeated assessments across infancy to calculate developmental trajectories in children born very preterm. Future research efforts should focus on identifying factors that influence developmental trajectories in order to develop clinically effective interventions that will improve the long-term neurodevelopmental outcomes of infants born preterm.

Acknowledgments

References

1   Office for National Statistics. Birth characteristics.

        https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/l

        ivebirths/datasets/birthcharacteristicsinenglandandwales (accessed 29 Jun 2018).

2   Woodward LJ, Moor S, Hood KM, *et al.* Very preterm children show impairments

        across multiple neurodevelopmental domains by age 4 years. *Arch Dis Child Fetal*

        Neonatal Ed 2009;**94**.

3   Woythaler MA, McCormick MC, Smith VC. Late preterm infants have worse 24-

        month neurodevelopmental outcomes than term infants. *Pediatrics* 2011;**127**.

4   Karmiloff-Smith A. Development itself is the key to understanding developmental

        disorders. Trends Cogn. Sci. 1998;**2**:389–98.

5   Karmiloff-Smith A. Nativism versus neuroconstructivism: rethinking the study of

        developmental disorders. *Dev Psychol* 2009;**45**:56–63.

6   Karmiloff-Smith A, D'Souza D, Dekker TM, *et al.* Genetic and environmental

        vulnerabilities in children with neurodevelopmental disorders. Proc. Natl. Acad. Sci.

        U. S. A. 2012;**109**:17261–5.

7   Thomas MSC, Annaz D, Ansari D, *et al.* Using developmental trajectories to

        understand developmental disorders. *J Speech, Lang Hear Res* 2009;**52**:336–58.

8   Cole TJ, Donaldson MDC, Ben-shlomo Y. SITAR-a useful instrument for growth

        curve analysis. *Int J Epidemiol* 2010;**39**:1558–66.

9   Cole TJ, Statnikov Y, Santhakumaran S, *et al.* Birth weight and longitudinal growth in

        infants born below 32 weeks' gestation: a UK population study. *Arch Dis Child Fetal*

        Neonatal Ed 2014;**99**:F34-40.

10   Bayley N. *Bayley Scales of Infant and Toddler Development (third edition) UK and*

        Ireland supplement manual. 3rd ed. San Antonio, TX: : Psychological Corporation

        2010.

11   R Core Team. *R: A language and environment for statistical computing*. Vienna,

        Austria: : R Foundation for Statistical Computing 2017.

12   Cole T. Super Imposition by Translation and Rotation Growth Curve Analysis R

        package sitar. 2018.

13   Yaari M, Mankuta D, Harel-Gadassi A, *et al.* Early developmental trajectories of

        preterm infants. *Res Dev Disabil* 2018;**81**:12–23.

14   Sansavini A, Pentimonti J, Justice L, *et al.* Language, motor and cognitive

        development of extremely preterm children: Modeling individual growth trajectories

        over the first three years of life. *J Commun Disord* 2014;**49**:55–68.

15   Linsell L, Johnson S, Wolke D, *et al.* Cognitive trajectories from infancy to early

        adulthood following birth before 26 weeks of gestation: A prospective, population-

        based cohort study. *Arch Dis Child* 2018;**103**:363–70.

16   Wang LW, Wang ST, Huang CC. Preterm infants of educated mothers have better

        outcome. *Acta Paediatr Int J Paediatr* 2008;**97**:568–73.

17   Moore T, Johnson S, Haider S, *et al.* Relationship between test scores using the

        second and third editions of the Bayley Scales in extremely preterm children. *J

        Pediatr* 2012;**160**:553–8.

18   Anderson PJ, De Luca CR, Hutchinson E, *et al.* Underestimation of developmental

        delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med* 2010;**164**:352–6.

19   Vohr BR, Stephens BE, Higgins RD, *et al.* Are outcomes of extremely preterm infants

        improving? Impact of Bayley assessment on outcomes. *J Pediatr* 2012;**161**:222–8.

20   Xiong T, Gonzalez F, Mu D-Z. An overview of risk factors for poor

        neurodevelopmental outcome associated with prematurity. *World J Pediatr*

        2012;**8**:293–300.

21   Ambalavanan N, Carlo WA, Tyson JE, *et al.* Outcome trajectories in extremely

        preterm infants. *Pediatrics* 2012;**130**:e115-25.

22   Cheong JLY, Lee KJ, Boland RA, *et al.* Changes in long-term prognosis with increasing postnatal survival and the occurrence of postnatal morbidities in extremely preterm infants offered intensive care: a prospective observational study. *Lancet Child Adolesc Heal* 2018;**2**:872–9.

23   Office for National Statistics. Language in England and Wales: 2011. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/articles/languageinenglandandwales/2013-03-04#main-language-by-local-authority (accessed 17 Aug 2018).

24   Walch E, Chaudhary T, Herold B, *et al.* Parental bilingualism is associated with slower cognitive development in very low birth weight infants. *Early Hum Dev* 2009;**85**:449–54.

Tables

Table 1 Characteristics of the study participants

Characteristic	n=1,142
Gestation, median (IQR), weeks^{+days}	27^{+3} (25^{+6} – 28^{+4})
Birthweight, median (IQR), g	910 (740.5 – 1120.75)
Male sex, n (%)	586 (51.3)
Multiplicity	
Singleton, n (%)	902 (79)
Twin, n (%)	214 (18.7)
Triplet, n (%)	26 (2.3)
Neonatal morbidities*	
Oxygen dependency at 36 weeks post-menstrual age, n (%)	364 (31.9)
Intra-ventricular haemorrhage grade 3 – 4, n (%)	90 (7.9)
Cystic periventricular leukomalacia, n (%)	32 (2.8)
Post-haemorrhagic hydrocephalus, n (%)	24 (2.1)
Retinopathy of prematurity stage 3 – 5, n (%)	60 (5.3)
Necrotising enterocolitis, n (%)	39 (3.4)
Number of Bayley-III assessments attended, n (%)	
Two	295 (25.8)
Three	389 (34.1)
Four	458 (40.1)

*Data on these neonatal morbidities were missing for 28 participants

Table 2 Mean Bayley-III composite scores of participants at each assessment

	Bayley-III Scales		
	Cognitive	**Language**	**Motor**
3-month assessment			
Mean (SD)	105.5 (13.7)	91.3 (9.8)	106.3 (12.3)
N	853	853	854
6-month assessment			
Mean (SD)	98.4 (13.0)	96.1 (12.8)	95.2 (16.7)
N	915	909	912
12-month assessment			
Mean (SD)	96.1 (14.0)	91.3 (14.0)	88.7 (13.5)
N	923	907	920
24-month assessment			
Mean (SD)	92.8 (16.2)	90.6 (18.3)	93.2 (15.8)
N	824	779	819

Table 3 Multiple regression analysis results for the intensity random effect as measured on each scale, including the coefficients, t-values and p-values

	Cognition	Receptive Communication	Fine Motor	Gross Motor
Variance Explained by SITAR Model	62%	68%	53%	68%
Standard Deviation of Intensity Random Effect	0.16	0.24	0.15	0.23
Gestational Age (weeks)				
Coefficient	0.014	0.016	0.010	0.012
t-value	4.3	3.2	3.5	2.4
p-value	<0.001	<0.01	<0.001	0.02
Birth Weight (kg)				
Coefficient	0.085	0.099	0.056	0.096
t-value	3.8	2.9	2.9	2.9
p-value	<0.001	<0.01	<0.01	<0.01
Sex - Male				
Coefficient	-0.023	-0.063	-0.019	-0.027
t-value	-2.8	-5.1	-2.7	-2.2
p-value	<0.01	<0.001	<0.01	0.03
Multiple - Twin				
Coefficient	-0.023	-.0003	-0.026	-0.050
t-value	-2.2	-0.02	-2.9	-3.2
p-value	0.03	>0.9	<0.01	<0.01
Multiple - Triplet				
Coefficient	0.003	.004	0.003	0.029
t-value	0.1	0.1	0.1	0.7
p-value	0.9	0.9	0.9	0.5

Figure legends

Figure 1 Flowchart of the follow-up of participants and inclusion into study analysis

Figure 2 (a) Unadjusted cognition developmental trajectories for each individual, (b) mean spline curve of the model, (c) predicted curves for each individual based on the SITAR model, (d) adjusted trajectories shown in colour on top of the unadjusted trajectories in grey.

Figure 3 Predicted (colour) and mean (black) curves based on the SITAR model for (a) cognition, (b) receptive communication, (c) fine motor and (d) gross motor domains. The

black dots represent developmental age equivalent scores from the Bayley-III for each age of assessment.