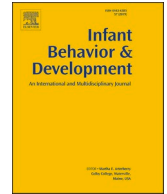




ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

The reliability and predictive ability of the Test of Infant Motor Performance (TIMP) in a community-based study in Bhaktapur, Nepal

Ingrid Kvestad^{a,b,*}, Jaya S. Silpakar^c, Mari Hysing^d, Suman Ranjitkar^c,
Tor A. Strand^{b,e}, Catherine Schwinger^e, Merina Shrestha^c, Ram K. Chandyo^f,
Manjeswori Ulak^{c,e}

^a Regional Centre for Child and Youth Mental Health and Child Welfare, NORCE Norwegian Research Centre, Bergen, Norway

^b Innlandet Hospital Trust, Department of Research, Lillehammer, Norway

^c Department of Pediatrics, Child Health Research Project, Institute of Medicine, Tribhuvan University, Kathmandu, Nepal

^d Department of Psychosocial Science, Faculty of Psychology, University of Bergen, Norway

^e Center for Intervention Science in Maternal and Child Health, Centre for International Health, University of Bergen, Bergen, Norway

^f Department of Community Medicine, Kathmandu Medical College, Sinamangal, Kathmandu, Nepal

ARTICLE INFO

Keywords:

Infant motor development
TIMP
Psychometric properties
Nepal
Low-resource setting

ABSTRACT

Aim: In a Nepalese setting, to measure the reliability of the Test of Infant Motor Performance (TIMP) and its ability to predict development scores at 6 months.

Methods: Nepalese infants ($n = 705$) were assessed by the TIMP when they were 8–12 weeks old and the Bayley Scales of Infant and Toddler Development, 3rd edition (Bayley-III) at 6 months. Inter-rater agreement was expressed by intraclass correlation coefficients (ICCs), the internal consistency by Cronbach's alphas and Pearson correlation coefficients. Predictive ability was estimated in linear regression models.

Results: Inter-rater agreement was excellent ($ICCs > 0.93$). Alphas for the TIMP total scores were 0.76 for infants born to term and 0.72 in those born preterm. Correlation coefficients between TIMP total and Bayley-III subscale-scores ranged from 0.05 to 0.28 for term infants and from 0.15 to 0.43 for preterm infants. Using American norms, 56.3 % had TIMP scores within average and 43.7 % below average range. Bayley-III subscale scores were lower in children with TIMP scores below the average range, with the strongest estimates for Gross motor and Socio-emotional development.

Interpretation: The reliability of the TIMP was acceptable, and the TIMP could be a feasible tool to monitor infant motor development in low-resource settings. Properties of the TIMP differed according to gestational age.

1. Introduction

Motor development in infancy is an important indicator for early child development. Suitable neonatal neurobehavioral and

* Correspondence to: Regional Centre for Child and Youth Mental Health and Child Welfare, NORCE Norwegian Research Centre, Nygårdsgaten 112, 5008 Bergen, Norway.

E-mail address: inkv@norceresearch.no (I. Kvestad).

<https://doi.org/10.1016/j.infbeh.2023.101809>

Received 22 May 2022; Received in revised form 3 January 2023; Accepted 5 January 2023

Available online 9 January 2023

0163-6383/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

neuromotor assessment tools are beneficial both for clinical purpose and to address specific research questions in early infancy (Noble & Boyd, 2012). These tools assess spontaneous behavioral repertoires in newborns and observable behavioral responses to environmental stimuli (Majnemer & Snider, 2005). Early assessment of motor performance in infants is complex and requires reliable and valid tests with adequate psychometric properties.

The Test of Infant Motor Performance (TIMP) is proposed as a clinically useful and psychometrically sound tool for early motor assessment (Heineman & Hadders-Algra, 2008; Majnemer & Snider, 2005; Noble & Boyd, 2012). The tool evaluates postures and movements for functional activities in infants from 34 weeks postmenstrual age to 17 weeks corrected age (Campbell, 2012). The first version of the TIMP was designed in 1993 and the test has since gone through 5 versions. Version 5, finalized in 2005, includes 42 items, where 13 are observed items and 29 elicited items. This version also includes new age standard norms based on a cross-sectional sample of 990 US infants. The test can be used to identify delayed motor development in infants below 4–5 months, track typically developing infants with precision, and to develop and evaluate intervention goals for infants with delayed development. Clinically, the test can also be used to educate parents on infant motor development (Campbell, 2012). The test is typically used by physical and occupational therapists, and no formal training is needed (Haffner & Sankovic, 2022). The TIMP test has been widely used among at-risk infants and is one of few tools with published norms for infants born prematurely (Noble & Boyd, 2012). Previous studies on the ability of the TIMP to differentiate between infants at high and low risk of delays show promising results (Heineman & Hadders-Algra, 2008; Majnemer & Snider, 2005). The TIMP is also one of few neonatal assessment tools that has been used to evaluate treatment programs (Girolami & Campbell, 1994; Lekskulchai & Cole, 2001).

In a systematic review, the TIMP is proposed to be among the tools with the best predictive ability for future development in infants 4 months and below (Spittle et al., 2008). In infants at risk, studies have demonstrated a relationship between TIMP scores in early infancy and developmental scores measured at 6 months (Campbell et al., 2002; Kim et al., 2011). Recent studies in preterm infants have also demonstrated associations between TIMP scores in infancy and scores on the Bayley Scales of Infant and Toddler Development (Bayley) at 2 years (Madayi et al., 2021; Peyton et al., 2018).

In a large population-based double-blinded clinical trial, conducted in Bhaktapur Nepal, we included early motor function as a secondary outcome in infancy when children were approximately 10 weeks old. The TIMP was chosen after a careful review of feasible motor function tests and based on studies showing acceptable psychometric properties and that the test had previously been used to evaluate intervention effects. Being set in a low-to-middle-income (LMIC) setting, the cost-effectiveness of the TIMP and that the examiners could be trained on-site without a more expensive process of certification outside of the country, was also important factors in our decision process. The main aim of the present study was to examine reliability measures such as inter-rater agreement and internal consistency of the TIMP when used by health workers in an urban community-based sample of infants in Nepal. Moreover, we aimed to measure the ability of TIMP scores in infancy to predict Bayley-III subscale scores when the children were 6 months old. Since evidence from previous research on psychometric properties of the TIMP are predominantly from studies in preterm infants, a secondary aim of the study was to examine to what extent the properties of the scale differed according to gestational age.

2. Materials and methods

2.1. Study design and setting

Data for the current study was collected in infants of mothers enrolled in a community-based, randomized double-blind placebo-controlled trial to assess the effect of daily maternal vitamin B12 supplementation from pregnancy to 6 months post-partum on growth and neurodevelopment (Chandyo et al., 2017). The primary outcomes of the trial were linear growth and neurodevelopment measured when the infants were 6 and 12 months of age. The TIMP test was included as a secondary outcome, to measure the effect of the intervention on infant motor performance in infancy. The study was undertaken in Bhaktapur municipality, and the surrounding areas located in the east of the Kathmandu valley of Nepal. Women in early pregnancy, aged between 20 and 40 years, residing and planning to reside in the area for at least the next 2 years, and who consented to participate in the study were included in the study. TIMP was administered when the infants were 10 weeks \pm 7 days of age according to date of birth. Of the 800 women included in the original trial, 760 gave birth to a child that was alive, and of these, 740 were tested with the TIMP. Due to the COVID19 pandemic, all children could not be reached within the window period and children more than 12 weeks according to date of birth ($N = 35$) were excluded from the analyses. The number of infants for the current sub-study was therefore 705, out of these we have Bayley-III measurements at 6 months in 694 infants.

The study was approved by the Nepal Health Research Council (NHRC; number 253/2016) and the Norwegian Regional Committee for Medical and Health Research Ethics (REC; number 2016/1620) and is registered at clinicaltrials.gov (registration number [NCT03071666](https://clinicaltrials.gov/ct2/show/study/NCT03071666) and universal trial number U111-1183-4093). All women participants gave written informed consent for participation in the study.

2.2. Data collection

Demographic information was obtained from mothers during enrollment in their early pregnancy, and birth records were collected after delivery. The demographic information included family socio-economic situation such as parental education, occupation (not working/agriculture, daily wage earner, services, government, and private employment and foreign employment) and whether the families lived in joint families, owned their house and owned land (all yes/no). Information on the child, included gestational age (weeks), birthweight (grams), whether or not there were any congenital anomalies (i.e. congenital heart disease, cleft lip, and

polydactyly) and if the child was hospitalized after birth (yes/no). Preterm birth was defined as gestational age < 37 weeks and low birthweight as birthweight < 2500 g.

2.3. Test of Infant motor performance (TIMP)

The TIMP consists of 42 items with 13 observed items and 29 elicited items (Campbell, 2012). The observed subscale (score of 0 = skill not observed and 1 = skill observed) is used to examine spontaneous movements such as selective control of fingers and ankles, midline head orientation, and development of ballistic, oscillating, and fidgety movements. The elicited subscale (scores vary from 0 to 6) examines movement responses to visual and auditory stimuli in different positions such as prone, supine, and supported sitting and standing, as well as control of the head in a variety of spatial orientations. The total possible TIMP score is 142 (13 for the observed subscale and 129 for the elicited subscale), where higher scores indicate better motor performance. The TIMP has age specific norms developed in the US, categorizing the motor performance into average range (± 1 SD of the normative mean), low average (-0.5 SD to -1 SD of the normative mean), below average (-1 to -2 SD below the normative mean), and far below average range (< -2 SD below the normative mean) (Campbell, 2012).

Before the study, the two examiners (a pediatrician and a psychologist) were trained by a clinical child psychologist. The training was followed by standardization procedures in which 20 infants were tested, and the two examiners scored the test individually. During the study, as part of study procedures (Chandyo et al., 2017) approximately 5 % of the infants were randomly selected to be double scored by the second examiner to measure inter-observer agreement.

During the study, the trained staff tested the infants in the presence of the caregiver in designated rooms at the study clinic. Prior to testing, the caregiver received information on the test, and it was made sure that the child was well-fed, healthy, and not sleepy. The test took between 20 and 50 min to complete, and scoring was done during testing. Upon completion, caregivers were given brief feedback and answered any parental concern.

2.4. Bayley Scales of Infant and Toddler Development, 3rd edition (Bayley-III)

Child development at 6 months was assessed by the Bayley-III, which is a widely used assessment tool for children 1–42 months (Bayley, 2006a). The Bayley-III consists of a series of developmental tasks and takes between 45 and 60 min to administer. The tool has previously been used and evaluated at the current study site (Ranjitkar et al., 2018). Three psychologists performed the Bayley-III assessments. They were standardized in the assessments reaching an intra-class correlation (ICC) > 0.85 when compared with a gold standard ahead of the study and > 0.94 for the double-scoring during the study. The tool consists of a cognitive, language and motor scales administered with the child, and a socio-emotional scale based on parent-report (Bayley, 2006b). Raw scores were converted to subscale scores (expected mean (SD) of 10 (3) based on American norms (Bayley, 2006a). In the current analysis, we used the 6 subscale scores: Cognitive, Receptive and Expressive Language, Gross and Fine Motor and Socio-emotional.

Table 1
Demographic and perinatal characteristics of 705 infants from Bhaktapur, Nepal.

Characteristics		N (%) / mean (SD)
Infant demographics		
Male child		367 (52.1 %)
Age of infants in weeks, mean (SD)	Un-corrected age	10.27 (0.4)
	Corrected age in weeks	9.52 (1.6)
Born preterm (< 37 weeks)		61 (8.7 %)
Birth weight in gram, mean (SD)		3197.9 (1444.9)
Low birth weight (< 2500 g)		91 (12.9 %)
Congenital anomaly		17 (2.4 %)
Hospitalization to NICU as newborn		43 (6.1 %)
Family demographics		
Age of mother, years, mean (SD)		27.6 (4.0)
Maternal education	Up to grade 10	291 (41.3 %)
Occupation of mother	Housewife/Agriculture	239 (33.9 %)
	Daily wage earner	100 (14.2 %)
	Business	149 (21.1 %)
	Government and private employment	217 (30.8 %)
Age of father, years, mean (SD)		30.13 (4.3)
Paternal education	Up to grade 10	338 (47.9 %)
Occupation of father	Not working/Agriculture	23 (3.3 %)
	Daily wage earner	146 (30.7 %)
	Business	306 (43.4 %)
	Government and private employment	200 (28.4 %)
	Foreign employment	30 (4.3 %)
Living in joint family		520 (65.0 %)
Owned their house		597 (74.6 %)
Ownership of land		489 (69.4 %)

Values are numbers (%) unless otherwise specified.

2.5. Statistical analysis

Analyses were performed using Stata®, version 17 (Stata Corporation, College Station, Texas, USA) and JASP, version 0.16.2 (JASP team). Demographic characteristics are presented as means and standard deviations (SD) for continuous data, and frequency counts and percentages (%) for categorical data. Inter-rater agreement was expressed by ICCs obtained from one-way random effects models for single measurements (Barnhart et al., 2007).

We present data separately for preterm infants and term-born infants; results for the total group can be found as Supplementary material. We graphically present the distribution of the total TIMP scores in smoothed density plots using univariate kernel density estimation. To examine the internal consistency of the TIMP total subscale scores, we calculated Cronbach's alphas and intra-correlations with Pearson correlation coefficients. Alpha values < 0.6, 0.6–0.8, and > 0.8 were considered questionable, acceptable, and good. Correlation coefficients < 0.3, 0.3–0.5, 0.5–0.7 and > 0.7 were considered weak, moderate, strong, and very strong. The relationship between the TIMP scores and the Bayley subscale scores at 6 months, was estimated using Pearson correlation coefficients. The extent to which gestational age at birth (preterm vs. term) modified the association between the TIMP total and Bayley subscale scores was assessed in linear regression models by including an interaction term.

Based on the total TIMP scores, we calculated the number (%) of study children who were within average range, low average, below average, and far below average, corrected for gestational age using US norms (Campbell, 2012). One child received a total score above the average range for its age and was coded as with scores within the average range. In linear regression models, we assessed the difference in Bayley-III subscale scores at 6 months in children with TIMP scores below average range (i.e., scores < – 1 SD below the normative mean) compared to those within the average range (i.e., scores \pm 1 SD of the normative mean).

3. Results

The baseline characteristics of the 705 infants are shown in Table 1. Participants had a mean age of 10 weeks and 3 days, ranging from 8.2 to 12 weeks at assessment. Correcting for gestational age, mean (SD) age was 9 weeks and 5 days (1.63), ranging from – 2.05 to 14.10. Among the participants, 52 % were male. The average birth weight was 3197 (1444) g, with 13 % born with low birthweight, and 9 % born preterm. The mean age (SD) of the mothers was 28 (Campbell, 2012) years, 65 % lived in joint families, 75 % owned a house, and 41 % had an educational level less than or equal to grade ten.

The ICCs on the TIMP total and subscale scores between the examiners' scoring during standardization procedures and throughout the study period demonstrate an excellent inter-rater agreement (ICCs > 0.93) (Table 2).

The mean (SD) TIMP total and subscale scores, alphas and intra-correlations coefficients are shown in Table 3 and in Supplemental Table 1. The distribution of the TIMP total scores by gestational age are also depicted in Fig. 1. The mean (SD) TIMP total scores in infants born to term were 77.73 (9.66), while the scores for preterm infants was 70.70 (9.29) with a mean difference (95 % CI) of 7.02 (95 % CI 4.49, 9.55) (Table 3).

The alpha (95 % CI) for the TIMP total score for infants born to term was 0.76 (95 % CI 0.74, 0.79) and for preterm infants 0.72 (95 % CI 0.61, 0.81) indicating acceptable internal consistency in both groups. The alphas for the elicited subscale were similar to those of the total score, while the alphas for the observed subscale indicated questionable internal consistency (< 0.31). The correlation coefficients between the total and elicited subscale scores were very strong (0.99), while the correlation coefficients between the TIMP total and the observed subscale scores were weak (0.32 (95 %CI 0.25, 0.39) and 0.28 (95 %CI 0.03, 0.50) for infants born to term and preterm respectively).

The correlation coefficients between the TIMP total and subscale scores and the Bayley-III subscale scores stratified by infants born term and preterm are shown in Table 4 and for the total sample in Supplemental Table 2. For infants born to term, all coefficients between the total TIMP score and the Bayley-III subscale scores were weak ranging from 0.05 (95 %CI – 0.03, 0.13) for the Socio-emotional subscale to 0.28 (95 %CI 0.20, 0.35) for the Gross motor subscale. For infants born preterm, the correlation coefficients between the total TIMP score and the Bayley-III subscale scores ranged from 0.15 (95 %CI – 0.11, 0.39) for the Socio-emotional subscale to 0.43 (95 %CI 0.14, 0.58) for the Cognitive subscale with varying precision of the estimate. In this group, four of the estimates was of moderate strength (for Cognitive, Expressive language, Fine and Gross motor subscale). Gestational age (born preterm vs. to term) modified the association between the TIMP total score and the Bayley-III Cognitive subscale score ($p = 0.005$) and the Fine motor subscale ($p = 0.04$), but not the other subscales.

According to US norms, 397 (56.31 %) had scores within the average range and a total of 308 (43.68 %) had scores below the average range, in which 18 (2.55 %) had scores far below average range. Overall, although negligible, mean Bayley subscale scores were lower in children who had TIMP scores below average range in infancy than in children with TIMP scores within the average range (Table 5). The mean differences were largest for the Gross motor (– 0.65, (95 %CI – 1.02, – 0.29)) and Socio-emotional

Table 2

Inter-agreement between the two examiners during standardization procedures and throughout the study period.

TIMP score	ICC ^a (95 %CI) during standardization (N = 20)	ICC during study (N = 40)
Observed	0.94 (0.89, 0.98)	0.93 (0.88, 0.97)
Elicited	0.98 (0.97, 0.99)	0.96 (0.94, 0.99)
Total	0.98 (0.97, 0.99)	0.97 (0.95, 0.99)

^a Intraclass correlation.

Table 3

Mean (SD) and the internal consistency of the total TIMP, observed and elicited scores in 705 Nepalese infants 8–12 weeks uncorrected age born to term (N = 644) and preterm (N = 61).

TIMP Score	Mean (SD)	Range	Mean diff. Preterm vs. born to term ¹	Alpha ²	Intra-correlations ³	
					Observed	Elicited
Infants born to term (n = 644)						
Observed	8.10 (1.28)	5–12	0.48 (0.14, 0.81)	0.31 ⁴ (0.23, 0.38)		
Elicited	69.62 (9.33)	39–104	6.54 (4.10, 8.99)	0.77 (0.74, 0.79)	0.20 (0.12, 0.27)	
Total	77.73 (9.66)	45–114	7.02 (4.49, 9.55)	0.76 ⁴ (0.74, 0.79)	0.32 (0.25, 0.39)	0.99 (0.99, 0.99)
Infants born preterm (n = 61)						
Observed	7.62 (1.12)	6–11		0.26 ⁴ (– 0.07, 0.51)	-	-
Elicited	63.08 (9.04)	41–83		0.72 (0.61, 0.81)	0.17 (– 0.09, 0.40)	
Total	70.70 (9.29)	49–90		0.72 ⁴ (0.61, 0.81)	0.28 (0.03, 0.50)	0.99 (0.99, 0.99)

¹ Linear regression models, ² Cronbach alpha, ³ Pearson correlation coefficients, ⁴ item 1, 2, 3, 6, 7, 8 dropped from the analysis due to lack of variance for preterm infants, item 1 dropped from the analysis due to lack of variance for preterm infants.

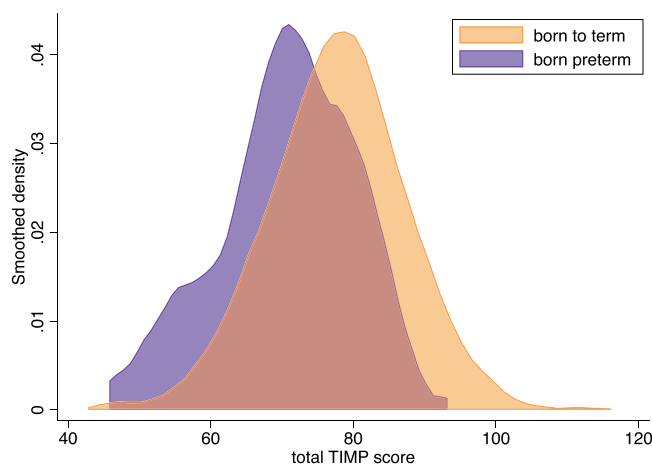


Fig. 1. Distribution of the TIMP total scores in Nepalese infants born preterm (N = 61) and infants born to term (N = 643).

Table 4

Correlations between Test of Infant Motor Performance total and subscale scores at 8–12 weeks uncorrected age and Bayley-III scaled scores at 6 months in 694 Nepalese infants.

TIMP scores		Bayley subtests					
		Cognitive	Receptive	Expressive	Fine motor	Gross motor	Socio-emotional
Infants born to term (N = 635)							
Total	Pearson's r	0.16	0.09	0.14	0.20	0.28	0.05
	95 % CI	0.08, 0.23	0.01, 0.17	0.07, 0.22	0.12, 0.27	0.20, 0.35	-0.03, 0.13
Observed	Pearson's r	0.11	0.05	0.10	0.11	0.11	0.04
	95 % CI	0.03, 0.18	-0.02, 0.13	0.02, 0.17	0.04, 0.19	0.03, 0.19	-0.04, 0.12
Elicited	Pearson's r	0.15	0.08	0.14	0.19	0.27	0.05
	95 % CI	0.07, 0.22	0.01, 0.16	0.06, 0.21	0.11, 0.26	0.20, 0.34	-0.03, 0.13
Infants born preterm (N = 59)							
Total	Pearson's r	0.43	0.24	0.41	0.37	0.39	0.15
	95 % CI	0.20, 0.62	-0.02, 0.47	0.17, 0.60	0.13, 0.57	0.14, 0.58	-0.11, 0.39
Observed	Pearson's r	0.19	-0.00	0.01	-0.12	0.03	0.16
	95 % CI	-0.07, 0.42	-0.26, 0.26	-0.25, 0.26	-0.37, 0.14	-0.22, 0.29	-0.11, 0.40
Elicited	Pearson's r	0.42	0.25	0.42	0.39	0.39	0.14
	95 % CI	0.19, 0.61	-0.01, 0.47	0.18, 0.61	0.15, 0.59	0.15, 0.59	-0.13, 0.38

Gestational age (born preterm vs. to term) significantly modified the association between the total TIMP score and the cognitive (p = 0.005) and fine motor (p = 0.04) subscales, but not receptive (p = 0.13), expressive (p = 0.16), gross motor (p = 0.44) and socioemotional (p = 0.53) subscales.

(-0.73 , (95 %CI -1.24 , -0.21)) subscales with the weakest estimate for the Receptive language subscale (-0.10 (95 %CI -0.38 , 0.19)).

4. Discussion

In the current study, we used the TIMP to evaluate motor performance in Nepalese infants at 8–12 weeks in a peri-urban community-based sample. This is the first paper to evaluate the feasibility of the TIMP among Nepalese infants. The inter-rater agreement between the two examiners were excellent and our findings demonstrate that with training, practice, and standardization exercises, it is possible to measure motor development reliably in this setting using the TIMP. Our findings are in agreement with previous reviews (Heineman & Hadders-Algra, 2008; Majnemer & Snider, 2005). Notably, while the estimates suggest acceptable internal consistency for the total scale and elicited subscale, the internal consistency of the observed subscale was questionable. Alpha values are sensitive to the number of items in the scale (Cortina, 1993), which could partly explain the low alphas for the observed subscale (13 items compared to 29 and 42 items in the elicited subscale and the total scale). The difference in internal consistency could also be explained by the difference of scoring in the two subscales, where the dichotomized scoring (yes/no) in the observational items leads to less variance than the elicited items (scoring options of 0–6). Following this, items of the observational scale were dropped from the alpha calculations due to lack of variance, leaving even fewer items for the analyzes. Finally, the low alpha value could be due to the fact that this subscale involves observation of spontaneous movements. Such observations could be more sensitive to length and context of the assessments, and perhaps best reported by parents. Our finding of low consistency of the observed subscale is in contrast to a recent study in Brazilian infants, demonstrating excellent internal consistency for all TIMP subscales (Chiquetti et al., 2020). Lack of internal consistency of a scale could question the validity of findings in terms of what the scale is measuring. Further examination of the TIMP in similar settings as the current is needed to better understand the low consistency of the observed items. Presently, the total and elicited scales seem to be more reliable measures of early motor development in the current setting.

As expected, our findings show higher mean total and subscale scores among infants born to term than in preterm infants, mirroring previous findings of delayed motor development in preterm infants (Spittle et al., 2008; Van Haastert et al., 2006). We found moderate correlation between the TIMP scores and the 6 months Bayley scores in preterm infants, and weak correlations in infants born to term. This difference between gestational groups was significant for later cognitive and fine motor skills, but not for other developmental domains. A range of previous studies suggest a relationship between early motor development and later development in preterm infants (Spittle et al., 2008), and in the current study we find moderate correlations in particular for cognitive and fine motor development. Hence, although the long-term implications of altered patterns of development in preterm infants are not fully understood, our findings suggest that this pattern could persist until children are 6 months related to cognitive and motor development. Our findings in conjunction with previous literature is interesting and provide support to that the feasibility of the TIMP tool among Nepalese preterm infants for early identification should be explored further.

Using US norms, around 45 % of the infants had total TIMP scores below the average range. Notably, using these norms, infants' age is corrected for gestational age. In general, although for some subscales only negligible differences are suggested, our findings suggest lower scores on all Bayley-III subscales in children with TIMP scores below average compared to within average range in infancy. These differences were most pronounced for gross motor and socio-emotional development, corresponding to a difference of approximately 3.5 composite scores (mean (SD) scores of 100 (Ranjitkar et al., 2018)) which could be considered a clinical meaningful difference on a population level. The predictive ability of early motor development on later gross motor development was expected and supported by the findings in fine motor development although the latter had a weaker estimate. Interestingly, our findings also show a relationship with more general development such as socio-emotional, and to a certain extent cognitive, and expressive language development. Few studies have investigated the association between TIMP score and later development as measured with the Bayley scales. One study from South Korea suggest that the TIMP is highly sensitive and specific with 6 months scores on the Bayley-II (Kim et al., 2011). Using the Bayley-II, this study was not able to differentiate developmental domains as detailed as in the current study. Recent studies have also demonstrated a high specificity of low TIMP scores in infancy for low Bayley-III subscale scores at 2 years, in one study this included all subscale scores (Peyton et al., 2018), and in a second the motor and cognitive subscales, but not the language subscales (Madayi et al., 2021). Further studies on the predictive ability of TIMP scores in early infancy on development beyond 6 months incorporating various developmental domains, is needed to track early child development and mechanisms involved.

Strengths of the study include a large and population-based sample in a low to middle income setting increasing the generalizability of the results, as well as the comprehensive assessment with both the TIMP and the Bayley-III in infancy within a narrow age window. Limitations of the study are the relatively small number of preterm children which limits statistical power in some of the analyzes and the lack of norms for a South Asian population.

5. Conclusion

To conclude, our results show that, with proper training and standardization exercises, the reliability of the TIMP was acceptable, and that the TIMP could be a feasible tool to monitor infant motor development in low resource settings. Our findings suggest stronger correlations between early motor performance and later cognitive and fine motor development among preterm infants compared to term infants, and the usefulness of the TIMP in preterm infants in LMIC settings should be explored further. The poor internal consistency of the observed subscale should also be subject to further investigation.

Table 5
Mean Bayley-III subscale scores at 6 months by TIMP range in early infancy in 694 Nepalese infants.

	Bayley-III subscale scores	
	Mean (SD)	Mean difference (95 %CI) ^a
TIMP range		
Average	Cognitive 10.04 (0.11)	Ref.
Below average	9.74 (0.12)	-0.30 (– 0.62, 0.02)
Average	Receptive language 9.63 (0.10)	ref.
Below average	9.53 (0.11)	-0.10 (– 0.38, 0.19)
Average	Expressive language 6.52 (0.08)	ref.
Below average	6.31 (0.09)	-0.21 (– 0.46, 0.04)
Average	Fine Motor 9.74 (0.15)	ref.
Below average	9.37 (0.15)	-0.37 (– 0.77, 0.04)
Average	Gross motor 10.52 (0.12)	ref.
Below average	9.87 (0.14)	-0.65 (– 1.02, – 0.29)
Average	Socio-emotional development 10.52 (0.17)	ref.
Below average	9.80 (0.20)	-0.73 (– 1.24, – 0.21)

^a Linear regression models.

Funding and conflict of interest

This work was supported by the Research Council of Norway through its Centres of Excellence scheme (Project no. 223269) and the University of Bergen (UiB), Norway, to the Centre for Intervention Science in Maternal and Child Health (CISMACH).

Data sharing statement

Data available on request. To meet ethical requirements for the use of confidential patient data, requests must be approved by the Nepal Health Research Council (NHRC) and the Regional Committee for Medical and Health Research Ethics in Norway. Requests for data should be sent to the authors, by contacting NHRC (<http://nhrc.gov.np>), or by contacting the Department of Global Health and Primary Care at the University of Bergen (post@igs.uib.no).

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jaya S. Silpakar, Suman Ranjitkar, Catherine Schwinger, Merina Shrestha, Ram K. Chandyo and Manjeswori Ulak reports financial support was provided by Research Council of Norway through its Centres of Excellence scheme (project number 223269). Manjeswori Ulak reports financial support was provided by University of Bergen (UiB), Norway, to the Centre for Intervention Science in Maternal and Child Health (CISMACH).

Data Availability

The authors do not have permission to share data.

Acknowledgement

We would like to express our gratitude to all participating infants and families in the study. We thank the study physician Kumari Twayana for her participation in the early phases of the project. We are also grateful to fieldworkers, supervisors and general staffs at the Child Health Research Project, Department of Child Health at the Institute of Medicine, Tribhuvan University and Siddhi Memorial Foundation and its founder Shyam S. Dhaubhadel.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.infbeh.2023.101809](https://doi.org/10.1016/j.infbeh.2023.101809).

References

- Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics*, *17*(4), 529–569.
- Bayley, N. (2006a). *Manual for the Bayley scales of infant and toddler development*. San Antonio, TX: NCS Pearson.
- Bayley, N. (2006b). *Bayley scales of infant and toddler development: Bayley-III*. San Antonio, Texas: Harcourt Assessment.
- Campbell, S. K. (2012). *The test of infant motor performance: Test user's manual version 3.0 for the TIMP version 5: Infant motor performance scales*. LLC.
- Campbell, S. K., Kolobe, T. H., Wright, B. D., & Linares, J. M. (2002). Validity of the test of infant motor performance for prediction of 6-, 9- and 12-month scores on the Alberta Infant Motor Scale. *Developmental Medicine and Child Neurology*, *44*(4), 263–272.
- Chandyo, R. K., Ulak, M., Kvestad, I., Shrestha, M., Ranjitkar, S., Basnet, S., et al. (2017). The effects of vitamin B12 supplementation in pregnancy and postpartum on growth and neurodevelopment in early childhood: Study protocol for a randomized Placebo controlled trial. *BMJ Open*, *7*, 8.
- Chiquetti, E., Valentini, N. C., & Sacconi, R. (2020). Validation and reliability of the test of infant motor performance for Brazilian infants. *Physical & Occupational Therapy in Pediatrics*, *40*(4), 470–485.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.
- Girolami, G. L., & Campbell, S. K. (1994). Efficacy of a neuro-developmental treatment program to improve motor control in infants born prematurely. *Pediatric Physical Therapy*, *6*(4), 175–184.
- Haffner, D. N., & Sankovic, A., editors (2022). *A neurologist's guide to neonatal neurodevelopmental assessments and preterm brain injury*. *Seminars in pediatric neurology*. Elsevier.
- Heineman, K. R., & Hadders-Algra, M. (2008). Evaluation of neuromotor function in infancy—A systematic review of available methods. *Journal of Developmental & Behavioral Pediatrics*, *29*(4), 315–323.
- Kim, S. A., Lee, Y. J., & Lee, Y. G. (2011). Predictive value of test of infant motor performance for infants based on correlation between TIMP and Bayley scales of infant development. *Annals of Rehabilitation Medicine*, *35*(6), 860.
- Lekskulchai, R., & Cole, J. (2001). Effect of a developmental program on motor performance in infants born preterm. *Australian Journal of Physiotherapy*, *47*(3), 169–176.
- Madayi, A., Shi, L., Zhu, Y., Daniel, L. M., Noordine, A. A., Sherwood, S. A. M., et al. (2021). The test of infant motor performance (TIMP) in very low birth weight infants and outcome at two years of age. *Journal of Perinatology*, *41*(10), 2432–2441.
- Majnemer, A., & Snider, L. (2005). A comparison of developmental assessments of the newborn and young infant. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(1), 68–73.
- Noble, Y., & Boyd, R. (2012). Neonatal assessments for the preterm infant up to 4 months corrected age: A systematic review. *Developmental Medicine & Child Neurology*, *54*(2), 129–139.
- Peyton, C., Schreiber, M. D., & Msall, M. E. (2018). The test of infant motor performance at 3 months predicts language, cognitive, and motor outcomes in infants born preterm at 2 years of age. *Developmental Medicine & Child Neurology*, *60*(12), 1239–1243.
- Ranjitkar, S., Kvestad, I., Strand, T. A., Ulak, M., Shrestha, M., Chandyo, R. K., et al. (2018). Acceptability and reliability of the Bayley scales of infant and toddler development-III among children in Bhaktapur, Nepal. *Frontiers in Psychology*, *9*.
- Spittle, A. J., Doyle, L. W., & Boyd, R. N. (2008). A systematic review of the clinimetric properties of neuromotor assessments for preterm infants during the first year of life. *Developmental Medicine & Child Neurology*, *50*(4), 254–266.
- Van Haastert, I., De Vries, L., Helders, P., & Jongmans, M. (2006). Early gross motor development of preterm infants according to the Alberta Infant Motor Scale. *Journal of Pediatrics*, *149*(5), 617–622.