Early Human Development 93 (2016) 43–46

Contents lists available at ScienceDirect

Early Human Development

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

Test–retest reliability of the Test of Infant Motor Performance Screening Items in infants at risk for impaired functional motor performance

Tordis Ustad [a,b], Jorunn L. Helbostad [a,c], Suzann K. Campbell [e], Gay L. Girolami [f], Lone Jørgensen [g,h], Gunn Kristin Øberg [g,h], Kari Anne I. Evensen [b,d,i]

[a] Department of Clinical Services, St. Olavs Hospital, Trondheim University Hospital, Pb 3250 Sluppen, N-7006 Trondheim, Norway
[b] Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, N-7491 Trondheim, Norway
[c] Department of Neuroscience, Faculty of Medicine, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[d] Department of Public Health and General Practice, Faculty of Medicine, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[e] University of Illinois at Chicago, 1301 W. Madison Street, Unit 526, Chicago IL 60607, USA
[f] Department of Physical Therapy (MC 898), College of Applied Health Sciences, University of Illinois at Chicago,1919 West Taylor Street, Chicago, Illinois 60612, USA
[g] Department of Health and Care Sciences, University of Tromsø, The Arctic University of Norway, Pb 6050 Langnes, N-9037 Tromsø, Norway
[h] Department of Clinical Therapeutic Services, University Hospital of North Norway, N-9038 Tromsø, Norway
[i] Department of Physical Therapy, Trondheim Municipality, Pb 2300 Sluppen, N-7004 Trondheim, Norway

ARTICLE INFO

Article history:
Received 10 June 2015
Received in revised form 11 December 2015
Accepted 15 December 2015

Keywords:
Test–retest
Preterm/high-risk infants
Motor assessment

ABSTRACT

Objective: To examine test–retest reliability of the TIMPSI in infants at risk for impaired functional motor performance.
Methods: The TIMPSI was administered twice to 51 infants from two different hospitals in Norway.
Results: The intra-class correlation coefficient was 0.99.
Conclusion: Test–retest reliability of the TIMPSI was excellent.

1. Introduction

Motor assessments in infants at risk for developmental delay are primarily performed to discriminate between typically developing infants and infants with suspected neurological dysfunction. This is important when planning intervention, predicting motor difficulties and evaluating change over time [1,2]. In order to direct resources towards infants likely to gain most from intervention, while avoiding intervention on infants with typical development, it is essential that assessment tools are reliable and valid. The prevalence of developmental difficulties in infants born preterm increases with decreasing gestational age at birth (GA) [3], and the incidence of motor disabilities such as cerebral palsy and developmental coordination disorder is particularly high [3–5]. Systematic reviews of neonatal assessments tools conclude that the Test of Infant Motor Performance (TIMP) is one of the best motor assessment tools to discriminate between infants with age-appropriate motor development and infants with delayed motor performance. Further it is also useful for planning interventions and evaluating change over time [1,2,6,7].

The TIMP was developed to assess functional motor performance in new-borns and infants from 34 weeks postmenstrual age (PMA) to 17 weeks corrected age (CA). Conducted at 3 months CA the TIMP was predictive of children's motor performance at 4-5 years, as measured by The Peabody Developmental Motor Scales [8]. A test–retest reliability study of the TIMP in 106 infants PMA 32 weeks to CA 16 weeks with varying risk and ethnicity demonstrated a high correlation between scores on two different days ($r = 0.89$) [7].

Average time to conduct the TIMP is 25–35 minutes, which for the youngest and most fragile infants may be too demanding. Therefore, a shorter version was developed, the Test of Infant Motor Performance Screening Items (TIMPSI) [6,9], for identifying infants for whom the full verison should be conducted. In a group of low birth weight infants between PMA 34 weeks and CA 17 weeks total TIMPSI scores correlated well with total TIMP scores ($r = 0.88$) [9]. A test–retest reliability study of the TIMPSI in infants at risk for long-term motor difficulties has not yet been carried out, but should be performed before routinely implementing this test for the assessment of fragile infants. We wanted to explore the clinical utility of the TIMPSI by investigating the stability in scores and the measurement error when the same tester conducted two consecutive tests. The aim of this study was to examine test–retest

E-mail addresses: tordis.ustad@ntnu.no (T. Ustad), jorunn.helbostad@ntnu.no (J.L. Helbostad), skc@uic.edu (S.K. Campbell), gayg@uic.edu (G.L. Girolami), lone.jorgensen@uit.no (L. Jørgensen), gunn.kristin.oeberg@uit.no (G.K. Øberg), karianne.i.evensen@ntnu.no (K.A.I. Evensen).

http://dx.doi.org/10.1016/j.earlhumdev.2015.12.007
0378-3782/

reliability of the TIMPSI in a group of infants at high to moderate risk for long-term motor developmental difficulties.

2. Methods and participants

This study used an observational design to investigate test–retest reliability of the TIMPSI within a period of three days. This specific time frame was chosen because developmental changes are expected to be minimal over such a short interval [7]. In order to generalise the findings to clinical work, infants between PMA 36–37 weeks and CA 12–13 weeks, with varying risk for neurologic diagnosis or motor delay, were recruited.

Between April 2013 to December 2014, fifty-one infants from two hospitals in Norway, the University Hospital of North Norway (n = 14) and St. Olavs Hospital, Trondheim University Hospital (n = 37), were recruited for this study. Infants with high or moderate risk for long term motor development difficulties were eligible for inclusion. High risk was defined as infants born prior to 28 weeks GA with a birth weight <1000 g, infants with Grade III or IV intraventricular haemorrhages or periventricular leukomalacias and term infants with severe asphyxia treated with hypothermia. Moderate risk was defined as GA from 28 to 33 weeks. Parents were required to understand Norwegian or English. Medically unstable infants, infants who had undergone surgery and infants with genetic syndromes were excluded. With the exception of holidays and periods when the testers were on leave, eligible infants were continuously recruited. The sample was a convenience sample depending on availability of infants and parents at two time points as well as testers.

The study protocol was reviewed by the Regional Committees for Medical and Health Research Ethics (REC) January 2012, which concluded that the study did not require approval but should be reported to the Data Protection Officer at the Hospital.

2.1. The assessment tool

The TIMPSI is comprised of 29 of the 42 items from the TIMP. There are observed items scored during the observation of spontaneous movements and elicited items designed to assess the responses to visual and auditory stimuli, handling and changes of position [6]. The test is divided into three subsets: a Screening set, an Easy set, and a Hard set. The Screening set consists of 11 items with rating scales from five to seven points, score range 0–51. All infants are first assessed with the Screening Set. Based on the raw score of the Screening Set, a second set of either 10 easier or 8 harder items is administered to obtain a total score for motor performance [6]. The Easy set has four dichotomously scored items and six items with a five- or six-point rating scale, score range 0-31. The Hard set has eight items: five dichotomously scored and three with a five-point rating scale, score range 0–17 [10]. The scores for the administered items are summed with higher scores indicative of better motor performance, maximum score 99. TIMPSI age standards are available in the TIMP manual [6] based on the motor performance of 990 U.S. infants [9]. Average scores for infants PMA 36-37 weeks is 42 (SD: 16) and 79 (SD: 13) for infants CA 12–13 weeks.

2.2. Procedure

One tester from each hospital participated. Both testers were experienced paediatric physiotherapists who had attended workshops on the TIMP and had been using the test regularly for several years. A physiotherapist unknown to the parents in the neonatal intensive care unit (NICU) or Follow-up clinic invited all parents of eligible infants to participate in the study and a written consent was obtained. Because we aimed to minimize the burden for each infant and parents, test 1 was administered as part of ordinary clinical practice, either at week 36–37 PMA or at week 12-13 CA. Approximately half of the infants were tested at week 36–37 PMA and half tested at week 12–13 CA. The infants should be in "State of arousal level" three (eyes open, no movements) or four (eyes open, large movements) according to Prechtl's States [11]. The ideal time of the day for most of the infants was following a period of sleep and before meals. Test 2 was carried out within three days after test 1. In case of two tests carried out on the same day, pauses of several hours between the tests ensured the infants were rested and in the proper behavioural state for testing. In addition, testers would not remember scoring details of the previous test.

2.3. Statistical analysis

Sample size was estimated a priori according to Walter [12]. With a power of 80% and a significance level of 5%, we needed 45 participants to achieve an intra-class correlation coefficient (ICC) ≥0.8. Normality of the data was examined by the Shapiro-Wilk test. Relative reliability between Test 1 and Test 2 for within-subject differences was assessed by calculating $ICC_{1.1}$ [13]. Relative reliability refers to consistent ranking of scores for an individual in a group by repeated measurements. Absolute reliability, the standard error of measurement, was calculated as the square root of the mean within-subject variance (S_W) [14,15]. S_W is expressed in the original measurement scale with a low value expressing a small degree of measurement error. The difference between a subject's measurement and the true value would be expected to be less than $1.96 \times S_W$ for 95% of the observations [14]. The difference between the two measurements for the same subject is then expected to be less than $\sqrt{2} \times 1.96 \times S_W = 2.77 \times S_W$ for 95% of the pairs of observations [14]. Bland Altman plot was used for verifying the consistency of the measurements [16]. This plot gives a graphical presentation of the differences between two tests plotted against the mean difference of the two tests allowing visual assessment of the scoring distribution and potential measurement bias [16]. The software IBM SPSS statistics version 22 was used to perform the statistical analyses.

3. Results

The mean time interval between Test 1 and Test 2 was 1 day (SD: 0.84). Thirteen (25%) of the infants had both tests administrated the

Table 1
Neonatal characteristics and age of subjects tested using the TIMPSI.

	High risk (n = 27)	Moderate risk (n = 24)	Total (n = 51)
Birth weight (grams): mean (SD)	1499 (1158)	1546 (292)	1524 (814)
Gestational age at birth (weeks): mean (SD)	29.8 (6.2)	30.4 (1.7)	30.1 (4.4)
Bronchopulmonary dysplasia: n (%)	12 (24%)	0 (0%)	12 (24%)
Abnormal caput ultrasound: n (%)	9 (18%)	4 (8%)	13 (25%)
Intracranial bleed Grade III or IV: n (%)	2 (4%)	0 (0%)	2 (4%)
Periventricular leucomalasia: n (%)	3 (6%)	2 (4%)	5 (10%)
Infants tested at postmenstrual age 36-37 weeks: n (%)	6 (12%)	21 (41%)	27 (53%)
Infants tested at post-term age 12-13 weeks: n (%)	11 (22%)	13 (25%)	24 (47%)

SD: Standard deviation.