

Intra-Individual Consistency in the Quality of Neonatal General Movements

Akmer Mutlu^{a, b} Christa Einspieler^b Peter B. Marschik^b Ayse Livanelioglu^a

^aSchool of Physical Therapy and Rehabilitation, Hacettepe University, Ankara, Turkey;

^bInstitute of Physiology, Center for Physiological Medicine, Medical University of Graz, Austria

Key Words

Preterm infants · Inter-scorer agreement · Spontaneous movements · Video analysis

Abstract

Background: The qualitative assessment of general movements (GMs) proved to be a highly sensitive and specific diagnostic tool for the assessment of the integrity of the young nervous system. It is essential that the quality of GMs remains consistent in an individual during a given recording at a certain date. **Objectives:** The aim of the study was to investigate the intra-individual consistency of the quality of GMs during one recording. **Methods:** 39 preterm infants were recorded at least twice; some were recorded three times. In all, 88 recordings were available but three recordings were excluded due to frequent crying, seizures or hypokinesia. Three scorers assessed 2–3 sequences of these 85 GM recordings. **Results:** The inter-scorer agreement was high (κ 0.85–0.94). Intra-individual consistency revealed a κ of 0.90 with a 95% CI (0.51, 1.00) for preterm GMs, 0.96 with a 95% CI (0.57, 1.00) for writhing GMs, and 0.92 with a 95% CI (0.53, 1.00) for fidgety GMs. **Conclusions:** The individual quality of GMs remains consistent for a neonate or young infant at a certain date.

Copyright © 2007 S. Karger AG, Basel

Introduction

Prechtl's method on the qualitative assessment of general movements (GMs) proved to be a highly sensitive and specific diagnostic tool for the assessment of the integrity of the young nervous system [1]. GMs involve the entire body in a variable sequence of arm, leg, neck, and trunk movements [2], and occur in age-specific patterns. The GMs of a preterm infant may have large amplitudes and be fast in terms of speed [3]. During term age and the first 2 months postterm, GMs are characterized by a small to moderate amplitude and slow to moderate speed. Typically they are elliptical in form, which creates the impression of writhing [4]. GMs during 3–5 months postterm are described as fidgety movements with small amplitude, moderate speed, and variable acceleration of the neck, trunk, and limbs in all directions [1, 3]. When the nervous system is impaired, GMs lose their complex and variable character and become monotonous and poor [for a recent review, see 5].

As the qualitative assessment of GMs is based on visual gestalt perception, an average inter-scorer agreement of 90% (average Cohen's κ 0.88) confirmed the objectivity of the method [for a review, see 3, p. 35]. Furthermore, an analysis of more than 8,000 assessments performed by about 800 observers revealed that a few days of training enables clinicians to apply GM assessment accurately [6].

Reliable assessment of GMs requires a standardized recording procedure [7]. A manual on this subject was published in 2004 [3]. Preterm infants are usually recorded for 30–60 min; this time span allows the investigator to record sufficient bouts of activity. The recording is viewed later at fast speed and three GM sequences are copied onto the assessment tape. From term age onwards, Einspieler et al. [3] suggested recording and assessment for 5–10 min during the behavioral state 4 [according to 8].

In order to establish the assessment of GMs as a diagnostic tool it was essential that the quality of GMs remains consistent in an individual during a given recording at a certain date. Intra-individual consistency has been investigated, but never published [G. Cioni and H. F.R. Precht, pers. commun.]. To study this topic systematically, we took advantage of more than 1-hour recordings performed in preterm infants, but also at term age (the period of writhing movements), and at 3–5 months (the period of fidgety movements). Provided that the inter-scorer agreement among three scorers confirmed the previous results [3, 5, 7], we used these long recordings to answer the following question: Does the intra-individual consistency ensure reliable judgment of the normal and abnormal quality of GMs at preterm and the periods of writhing and fidgety movements?

Subjects and Methods

The study population consisted of 39 infants (14 girls and 25 boys) born at the Hacettepe University Hospital, Ankara (Turkey) during 2006. Selection criteria were the following: preterm birth (mean gestational age 30 weeks; SD 2.5 weeks; range 26–34 weeks), assignment to an early intervention program [9, 10] (reported elsewhere), and parental consent to participate in the follow-up examination at 2 years of age. The mean birth weight was 1,417 g (SD 402 g; range 780–2,400 g). Twenty infants were multiples. All infants were videoed according to the standards of Precht's method of GM assessment [3, 7]. All subjects were recorded at least twice; some were recorded three times. In all, 88 recordings were available. According to the age-specific patterns of GMs, we had 27 recordings of preterm GMs, 32 recordings of writhing movements, and 29 recordings of fidgety movements. The duration of video recordings was between 35 and 95 min. Two to three different sequences (A, B, and C – each of 2 min duration) were selected from these tapes meeting the following criteria: state 4, i.e. active wakefulness during the period of writhing and fidgety GMs and sequences were bouts of activity occur during preterm GMs [3]; at least a 20-min interval between two different sequences. Following the standards of GM assessment [3] we excluded prolonged periods of fussing, crying, or hiccupping and sucking on a dummy. Two preterm infants had to be excluded from the GM assessment because 1 infant had seizures and the second was

hypokinetic. Another infant could not be scored during the writhing movement period because of frequent crying. Hence, 85 recordings could be scored twice (sequences A and B); 62 recordings could be assessed three times (sequences A, B, and C). All sequences were re-coded and copied in random order onto the following assessment tapes: preterm GMs A (n = 25), preterm GMs B (n = 25), preterm GMs C (n = 19), writhing movements A (n = 31), writhing movements B (n = 31), writhing movements C (n = 20), fidgety movements A (n = 29), fidgety movements B (n = 29), fidgety movements C (n = 23).

Procedure

Scorer 1 (A.M.) attended a basic (2004) and an advanced (2006) training course on Precht's GM assessment. Scorer 2 (C.E.) and scorer 3 (P.B.M.) are well versed in GM assessment; scorer 2 is an instructor of the method. The three scorers independently assessed the 243 video sequences. Preterm and writhing GMs were scored as normal or abnormal (categories: poor repertoire, i.e. the sequence of the successive movement components is monotonous; or cramped-synchronized, i.e. limb and trunk muscles contract and relax almost simultaneously [3]). Fidgety movements were scored as normal, abnormal, or absent. The time organization of fidgety movements was scored as continual or sporadic [3]. After having completed one of the age-related tapes (preterm or writhing or fidgety GMs; A, B or C), each assessment was sealed in an envelope (per rater and per sequence). Two to four days later the scorers continued the next series.

Statistics

Fleiss' κ , a variant of Cohen's κ for measuring inter-scorer reliability [11], is applicable when more than two scorers assign categorical scores to a fixed number of items [12].

Ethics

The study was approved by the Hacettepe University Medical Faculty Ethics Committee. The infants' parents gave their written informed consent to their children's participation in the study.

Results

Preterm GMs

Four infants were scored as normal, 15 infants had poor-repertoire GMs, and 6 infants had cramped-synchronized GMs. The intra-individual consistency is given in table 1. One infant was scored as cramped-synchronized GMs during the first two sequences and as poor-repertoire GMs during the last sequence. A further infant was scored as normal during the first two sequences and as poor-repertoire GMs during the third sequence.

Writhing Movements

Fifteen infants had normal GMs, 10 had poor-repertoire GMs, and 6 had cramped-synchronized GMs. Only 1 infant was inconsistently scored during the three video sequences. The first two assessments revealed normal

Table 1. Intra-individual consistency given in Fleiss' κ (95% CI)

	Preterm GMs	Writhing GMs	Fidgety movements
Number of infants	25	31	29
Normal vs. subcategories of abnormal GMs	0.90 (0.51, 1.00)	0.96 (0.57, 1.00)	0.92 (0.53, 1.00)

Table 2. Inter-scorer agreement (Fleiss' κ , 95% CI) among the three observers

	Preterm GMs	Writhing GMs	Fidgety movements
Number of infants assessed	25	31	29
Normal vs. subcategories of abnormal GMs	0.85 (0.46, 1.00)	0.94 (0.55, 1.00)	0.92 (0.53, 1.00)

GMs, the assessment of the third video sequence showed poor-repertoire GMs (table 1).

Fidgety Movements

Twenty-one infants were scored as normal, 6 infants had sporadic fidgety movements (i.e. interspersed with longer pauses [3]), and 2 infants were scored as 'no fidgety movements'. Only 1 infant was inconsistently scored by one observer (table 1). This disagreement was due to sporadic versus no fidgety movements.

Inter-Scorer Agreement

The inter-scorer agreement (Fleiss' κ) was between 0.85 and 0.94 (table 2). The disagreement was due to poor-repertoire GMs versus cramped-synchronized GMs in 5 infants, 3 recorded during preterm age and 3 infants recorded during term age. The κ for differentiation between normal and abnormal GMs was 1.00 with a 95% CI (0.61, 1.00). Scoring the fidgety movements revealed one disagreement (sporadic versus no fidgety movements).

Discussion

Reliability is an essential aspect when dealing with clinicians' assessments of discrete categories. The present investigation yielded an inter-scorer agreement of $\kappa = 0.85$ (95% CI 0.46, 1.00) to 0.94 (95% CI 0.55, 1.00), which confirms previously published κ values for GM assessment [for review, see 3].

The quality of GMs may improve or worsen within an individual developmental trajectory [1, 3, 5]. Hence, calculation of intra-individual reliability over time is of lim-

ited value. On the other hand, high intra-individual reliability (i.e. the consistency of GM quality during one recording at a certain date) is a prerequisite for clinical application of GM assessment. With a κ of 0.90 (95% CI 0.51, 1.00) for preterm GMs, 0.96 (95% CI 0.57, 1.00) for writhing GMs, and 0.92 (95% CI 0.53, 1.00) for fidgety GMs, this prerequisite is fulfilled. κ values from 0.41 to 0.60 indicate moderate agreement, from 0.61 to 0.80 substantial agreement, and >0.80 an almost perfect agreement [13]. Recently, Sim and Wright [14] mentioned that the prevalence, bias, and non-independence of rating might influence the magnitude of κ . For a situation in which scorers choose between classifying cases as either positive or negative in respect of an attribute (normal or abnormal GMs), a prevalence effect exists when the proportion of agreement on the positive classification differs from that of the negative classification [14]. Our prevalence indices were between 0.03 and 0.40. The effect on the κ for the assessment of writhing GMs was negligible. However, the larger prevalence indices for preterm GMs and fidgety movements resulted in a lower κ [15]. Accordingly, bias indices, which express the extent to which scorers disagree on the proportion of positive or negative cases [14], are low. The third effect, namely non-independent rating, may also be disregarded. Each scorer generated an assessment without knowledge of the other scorer's assessment. In addition, a time interval of at least 2 days between two assessments of the same individual (sequence A or B or C) is deemed adequate [16]; the time interval in our study was 2–4 days.

Only 1 preterm infant was scored as having normal GMs and then scored as having poor-repertoire GMs about 1 h later. A further 2 infants were inconsistently

within the abnormal categories of poor-repertoire and cramped-synchronized preterm or writhing GMs. In this context, it should be noted that the prediction of the subsequent neurological outcome is always based on developmental trajectories rather than a single recording [2, 3].

A different situation exists for fidgety movements. Infants are usually re-assessed at the age of about 12 weeks to evaluate the presence, quality or absence of fidgety movements [1]. All infants with normal fidgety movements in the present study were 100% reliably scored (inter- and intra-individual reliability). This is important because normal fidgety movements are highly predictive of a normal neurological outcome irrespective of the infant's history or the GM quality assessed during preterm and the period of writhing movements [1, 3, 5]. The only disagreement in both inter-rater and intra-individual reliability was registered in 1 infant who was recorded at 12 weeks' postterm age and scored in respect of sporadic fidgety movements and the absence of fidgety move-

ments. In the clinical setting, sporadic fidgety movements at 12 weeks can by no means be considered fully normal. Rather, they would have been a cause of concern, as the absence of fidgety movements would.

In conclusion, the present study showed that the assessment of GMs is a reliable diagnostic tool and the quality of GMs remains consistent within one individual during a single recording.

Acknowledgements

During her stay at the Institute of Physiology, Medical University of Graz, Akmer Mutlu was supported by the Leonardo da Vinci Mobility Project (Forming an Educational Model for Paediatric Physical Therapists in Early Rehabilitation; TR/06/A/F/EX1-0650). This study was supported by the Austrian Science Fund (FWF P16984-B02 and P19581-B02) and the Lanyar Foundation Graz, Austria (P325). The first author thanks Professor Gul Sener, Project Coordinator and Director of School of Physical Therapy and Rehabilitation, Hacettepe University, Ankara. Our gratitude goes to Mrs. Sujata Wagner for editing the English.

References

- 1 Prechtl HFR, Einspieler C, Cioni G, Bos AF, Ferrari F, Sontheimer D: An early marker for neurological deficits after perinatal brain lesions. *Lancet* 1997;349:1361–1363.
- 2 Prechtl HFR: Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction. *Early Hum Dev* 1990;23:151–158.
- 3 Einspieler C, Prechtl HFR, Bos AF, Ferrari F, Cioni G: Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants (incl. CD-ROM). London, MacKeith Press/Cambridge University Press, 2004.
- 4 Hopkins B, Prechtl HFR: A qualitative approach to the development of movements during early infancy; in Prechtl HFR (ed): *Continuity of Neural Functions from Prenatal to Postnatal Life*. Clin Dev Med 94. Oxford, Blackwell, 1984, pp 179–197.
- 5 Einspieler C, Prechtl HFR: Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system. *Ment Retard Dev Disabil Res Rev* 2005;11:61–67.
- 6 Valentin T, Uhl K, Einspieler C: The effectiveness of training in Prechtl's method on the qualitative assessment of general movements. *Early Hum Dev* 2005;81:623–627.
- 7 Einspieler C, Prechtl HFR, Ferrari F, Cioni G, Bos AF: The qualitative assessment of general movements in preterm, term and young infants – review of the methodology. *Early Hum Dev* 1997;50:47–60.
- 8 Prechtl HFR: The behavioural states of the newborn infant (a review). *Brain Res* 1974;76:185–212.
- 9 Bobath B: The early treatment of cerebral palsy. *Dev Med Child Neurol* 1967;9:373–390.
- 10 Yigit S, Kerem M, Livanelioglu A, Oran O, Erdem G, Mutlu A, Turanli G, Tekinalp G, Yurdakok M: Early physiotherapy intervention in premature infants. *Turk J Pediatr* 2002;44:224–229.
- 11 Fleiss JL, Cohen J: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Ed Psychol Meas* 1973;33:613–619.
- 12 Fleiss JL: Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378–382.
- 13 Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
- 14 Sim J, Wright CC: The kappa statistic in reliability studies: use, interpretation, and sample size requirement. *Phys Ther* 2005;85:257–268.
- 15 Feinstein AR, Cicchetti DV: High agreement but low kappa. I. The problem of two paradoxes. *J Clin Epidemiol* 1990;43:543–549.
- 16 Streiner DL, Norman GR: *Health Measurement Scales: A Practical Guide to Their Development and Use*, ed 3. Oxford, Oxford University Press, 2003.