Early Human Development 104 (2017) 17–21

Contents lists available at ScienceDirect

Early Human Development

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

# Inter-observer agreement of the General Movements Assessment with infants following surgery

Cathryn Crowle [a,b,*], Claire Galea [c], Catherine Morgan [b,c], Iona Novak [b,c], Karen Walker [a,b,c], Nadia Badawi [a,b,c]

[a] Grace Centre for Newborn Care, The Children's Hospital Westmead, Sydney, Australia
[b] University of Sydney, Sydney, Australia
[c] Cerebral Palsy Alliance Research Institute, Sydney, Australia

ARTICLE INFO

*Article history:*
Received 26 April 2016
Received in revised form 2 November 2016
Accepted 13 November 2016
Available online xxxx

*Keywords:*
Reliability
Inter-observer agreement
General Movements Assessment
Prechtl
Infant

ABSTRACT

*Background:* The General Movements Assessment (GMA) is a validated and reliable method of identifying infants at risk of adverse neurodevelopmental outcomes, however there is minimal data available on the use of the GMA with infants following surgery.
*Aims:* The aim of this study was to investigate the inter-observer agreement for the GMA with this infant population.
*Study design:* Reliability and agreement study.
*Subjects:* This was a prospective cohort study of 190 infants (male n = 112) born at term (mean 38 weeks, SD 2 weeks).
*Outcome measures:* A GMA was conducted in the Neonatal Intensive Care Unit (NICU) following either cardiac surgery (n = 92), non-cardiac surgery (n = 93) or both types of surgery (n = 5), and then again at three months of age. All videos were independently assessed by three advanced trained clinicians. Agreement and reliability statistics were calculated between each pair.
*Results:* We found moderate to substantial levels of agreement in the writhing period (66–77%, $AC_1 = 0.53–0.69$). For fidgety general movements, agreement was classified as almost perfect, ranging from 86 to 89% ($AC_1 = 0.84–0.88$).
*Conclusions:* The GMA has high levels of inter-observer agreement when used with infants who have undergone surgery in the neonatal period, making it a valid, complementary assessment tool. Research is now underway to determine the ability of the GMA to predict neurodevelopmental outcomes in this population.


## 1. Introduction

Prechtl's Method on the Qualitative Assessment of General Movements (GMA) is a valid, reliable, non-invasive assessment tool for identifying infants at risk of poor neurodevelopmental outcomes, specifically cerebral palsy, with the aim of intervening early and improving outcomes. Whilst extensive evidence exists for use of the GMA, particularly with the preterm infant population [1–5] there is a paucity of published data on the use of the GMA with the infant surgical population, despite the known risk of later neurodevelopmental problems following surgery in the neonatal period [6].

We know from population-based data of infants in New South Wales that at one year of age, infants who have undergone surgery in the neonatal period demonstrated a delay (ranging from mild to severe) on all subscales of the Bayley Scales of Infant & Toddler Development [6]. For infants who had undergone cardiac surgery, the greatest difference was on the gross motor subscale, with 50% of the infants demonstrating a delay, compared with 20% of controls. Infants who underwent non-cardiac surgery performed worse in four of the five subtests (cognition, receptive language, fine motor and gross motor) with statistically significant lower mean scores than the control infants. Gross motor delay was evident in 38% of these infants, compared with 20% of the control group [6].

The GMA is an assessment of quality of movement and can be used with infants from birth until approximately 20 weeks post-term age. It involves video recording an infant's spontaneously generated movements as they lay supine in a quietly awake state, in order to evaluate the integrity of the infant nervous system. It is a clinically feasible tool to use with fragile post-surgical infants as it does not require any handling [7]. Movement quality is categorised by trained observers as outlined by Prechtl and colleagues [8]. From preterm age, up until approximately nine weeks post-term age, 'writhing' general movements are categorised as either normal, poor repertoire, cramped-

* Corresponding author at: Grace Centre for Newborn Care, The Children's Hospital at Westmead, Locked Bag 4001, Westmead 2145, Australia.
*E-mail addresses:* Cathryn.crowle@health.nsw.gov.au (C. Crowle), cgalea@cerebralpalsy.org.au (C. Galea), cmorgan@cerebralpalsy.org.au (C. Morgan), inovak@cerebralpalsy.org.au (I. Novak), Karen.walker@health.nsw.gov.au (K. Walker), nadia.badawi@health.nsw.gov.au (N. Badawi).

http://dx.doi.org/10.1016/j.earlhumdev.2016.11.001
0378-3782/

synchronised or chaotic. From nine to twenty weeks post-term age, 'fidgety' general movements are categorised as either normal, abnormal or absent. Normal general movements are associated with a normal outcome, whilst infants demonstrating persistently cramped-synchronised movements and/or absent fidgety movements, are at high risk of cerebral palsy [8].

The GMA is increasingly being used in clinical settings to identify infants who would benefit from targeted early intervention. We know the GMA has high reliability in the preterm infant population and those with Hypoxic Ischaemic Encephalopathy (HIE) [4,9]. Previous studies have reported inter-observer agreement rates ranging from 87 to 93% [10–12], and studies reporting on inter-observer reliability using a kappa statistic report fair to almost perfect agreement, with kappa between 0.36 and 0.94 [10, 12–17]. To date this remains unreported in the infant surgical population. As the profile of abnormal neurodevelopmental outcomes has different prevalence's in this surgical group, it is important to study this group separately, rather than assume rates of agreement will be the same. Just as high prevalence affects the predictive accuracy of cerebral palsy using the GMA [18], studies of inter-observer agreement are also affected by prevalence. Establishing strong inter-observer agreement between trained clinicians using the GMA with the infant surgical population is one important step in determining validity.

The aim of this study was to determine the inter-observer agreement and reliability of the GMA with infants who have undergone surgery in the neonatal period.

## 2. Methods

Procedures and reporting for this study followed the guidelines outlined by Kottner and colleagues in 2011; GRRAS (Guidelines for Reporting Reliability and Agreement Studies) [19]. This paper proposed the items which should be addressed when reliability and agreement studies are reported, such as the subject and rater population, the process for rating, the statistical analysis and the reporting of estimates of reliability and agreement.

### 2.1. Study design

This was a prospective cohort study of 190 infants whose GMA videos were independently assessed by three advanced trained observers, two of whom were blinded to the infant's medical history. Levels of inter-observer agreement and reliability were measured.

### 2.2. Participants

We conducted a prospective study on the use of the GMA with infants following surgery in the neonatal period, which recruited 304 infants from a level 6 NICU in NSW, Australia. Infants were enrolled if they required surgery within the first 30 days of life, and were eligible for follow-up in the development clinic. The development clinic sees infants >30 weeks gestation with congenital cardiac conditions, major surgical anomalies, or significant neurological problems. Written consent for GMs video assessment was gained from parents/carers and ethics approval for the study was obtained through The Sydney Children's Hospital Network, Human Research Ethics Committee. No family declined to participate.

From the study sample of 304 infants, 190 infants were eligible and included in analysis for this sub-study. We excluded the 82 infants who had missing data for the GMA in the writhing period, and nine infants where the GMA was completed outside the optimum age range to more reliably assess writhing or fidgety movements (refer to procedure below). We then omitted infants whose videos were scored as 'un-assessable' due to behavioural state (crying, fussing) (n = 18); infants who did not proceed to surgery (n = 3), and infants whose videos were not available to be viewed by all three observers (n = 2). This left a sample of 190 infants with writhing and fidgety videos that were able to be scored (illustrated in Fig. 1).

### 2.3. Observers

The three observers had completed both the basic and advanced training courses offered by the GMs Trust. There were two occupational therapists and one physiotherapist, all experienced in the clinical use of the GMA. Two external observers, three years post-Advanced training certification, were blinded to the infant's medical history and any risk factors, and only provided with the age of the infant at the time of assessment, as recommended during GMA training. The third observer was the infant's treating clinician and could not be blinded to clinical details.

### 2.4. Procedure

In order to ensure consistency with the procedure, a protocol was developed. This included guidelines for inclusion/exclusion of videos. Writhing videos needed to be longer than 2 min duration and taken prior to 46 weeks gestational age. From this age onwards, writhing movements involving complex, variable rotations along the axis of the limbs and through the trunk, slowly begin to disappear, as they are replaced by smaller, circular movements that are characteristic of the fidgety period [8]. By about nine weeks of age, fidgety movements become more apparent, and are at their peak by 12 weeks of age. Similarly, these movements fade out as voluntary, goal directed movements take over. Based on this, for the purposes of evaluating inter-observer agreement, fidgety videos needed to be taken between 10 and 15 weeks of age.

Videos were taken following the guidelines outlined by Prechtl and colleagues [8], which included infants lying on their back with minimal clothing, no dummy (pacifier) or toys, and in an adequate behavioural state to lie quietly whilst unwrapped. Videos were scored as 'un-assessable' (and excluded) when there was persistent crying or unsettled behaviour, despite several attempts.

Observers assessed the videos in independent locations, but were aware that their scores would be compared. There was a Gestalt setting of typical cases [8] prior to each rating period, which lasted a maximum time of 1 h. Rating categories were those described by Prechtl and colleagues [8]. In the writhing stage the categories were normal, poor repertoire (PR), cramped synchronised (CS) or chaotic; in the fidgety stage, categories were normal, abnormal or absent. For the analysis of writhing data, there were only five infants that were rated as having 'chaotic' general movements. A decision was made to omit these from the final analysis due to the very small number, and the very low prevalence across all populations [8]. This left 185 infants for the analysis of writhing GMA.

### 2.5. Statistical analysis

Studies reporting on inter-observer agreement and reliability traditionally report the Cohen's kappa statistic, however when the prevalence of a condition is low, there is ample information available regarding the problems with this statistic [20,21]. For this study, an alternative reliability coefficient, the AC1 statistic was used, as it adjusts for chance agreement more appropriately than Cohen's kappa in this population [22,23]. However, Cohen's kappa (kappa) and percentage agreement were also calculated for inter-observer reliability. Linear weighted analysis for kappa was performed. Descriptive statistics were used to profile demographic characteristics of the sample.

Gwet's AC1 statistic and weighted kappa inter-observer agreement coefficients were interpreted using benchmark scales of Landis & Koch, and Altman [24] (refer to Table 1). 95% confidence intervals (p = 0.05) were calculated for weighted kappa and Gwet's AC1 statistic. Analysis was performed using Agree-Stat 2015 (Advanced Analytics,