Usability and inter-rater reliability of the NeuroMotion app: A tool in General Movements Assessments

Katarina A. Svensson a, b, Maria Örtqvist c, Arend F. Bos d, Ann-Christin Eliasson e, Heléne EK. Sundelin a, c, *

a Linköping University, Linköping, Sweden
b Crown Princess Victoria's Children's and Youth Hospital, University Hospital, Linköping, Sweden
c Neonatal Unit, Department of Women's and Children's Health, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden
d Division of Neonatology, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
e Neuropediatric Unit, Department of Women’s and Children’s Health, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden

A R T I C L E   I N F O

Article history:
Received 27 November 2020
Received in revised form
30 April 2021
Accepted 6 May 2021

Keywords:
Cerebral palsy
General movement assessment
Infant
Inter-rater reliability
Smartphone application

A B S T R A C T

Background: Early intervention after perinatal brain insults requires early detection of infants with cerebral palsy (CP). General Movements Assessments (GMA) in the fidgety movement period has a high predictive value for CP.

Aim: To investigate the NeuroMotion™ app’s usability regarding film quality and user experience and to assess the inter-rater reliability of GMA in a neonatal risk group.

Methods: GMA, inter-rater reliability and film quality was assessed in a cohort consisting of 37 infants enrolled in a multicentre study of GMA as part of the Swedish neonatal follow-up program for high-risk infants. Some of these infants were filmed twice. For evaluation of user experience 95 parents of 52 infants were addressed with a web-based questionnaire. A GMA expert assessed film quality and performed GMA and three on-site assessors, individually performed GMA. Inter-rater reliability was computed using Krippendorff’s alpha (k-alpha).

Results: In all, 45 films showed good or excellent quality. The response rate of the questionnaire survey was 40% and revealed predominantly positive perceptions of the NeuroMotion™ app. GMA in 36 infants resulted in substantial agreement (k-alpha = 0.72, 95%CI = 0.3–1.0) between the three on-site assessors’ consensus and the GMA expert. Inter-rater reliability for GMA between the on-site assessors was moderate (k-alpha = 0.48, 0.18–0.74).

Conclusion: The NeuroMotion™ app produces good technical quality films and the app user experience was overall positive. High agreement was observed between the on-site assessors and the GMA expert. The study design is feasible for more extensive GMA studies in cohorts of infants at risk of CP.

© 2021 The Authors. Published by Elsevier Ltd on behalf of European Paediatric Neurology Society. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

There is a need for early intervention in children with high risk of cerebral palsy (CP) and hence a necessity to identify these children earlier than what is currently the case. Although there has been improvements due to early screening most children are still diagnosed with CP after 1 year of age [1,2]. Prechtl’s General Movement Assessment (GMA) is recommended as one of the most important tools for early detection of CP [3,4]. GMA is a qualitative analysis of an infant’s spontaneous movements from birth until 20 weeks of corrected age [5]. Fidgety movements are continuous small amplitude, moderate speed movements of neck, trunk, shoulders, wrists, hips, and ankles in all directions and of variable accelerations, seen at 3–5 months post term in typically developing infants when they are awake and lying supine with no interference [4]. Several studies have shown that absence of fidgety movements on GMA during the fidgety movement period (9–20 weeks) can identify infants who later develop CP with a sensitivity of 98% and a specificity of 91–94% [6–8].

Studies of GMA inter-rater reliability during the fidgety...
movement period have been done with different methods and with various results [9–12]. To our knowledge, there are no studies on inter-rater reliability of GMA films of high-risk infants evaluated by blinded assessors and recorded in a home setting by parents using a smartphone app.

Film recordings for GMA are usually performed in an outpatient clinic, often not adapted to the infant’s needs and the non-familiar surroundings may influence its mood. Recording at home with a smartphone can adjust the situation to the child’s alertness and wellbeing and overcome inequality due to geographical distances. In Australia, the Baby Moves app has been successfully used for this specific aim with a positive user experience [13]. Nevertheless, it was not possible to use this app in Sweden to upload films to our secure server. Based on our information, there is no secure and efficient way for patients to upload films to Swedish health care providers.

Telemedicine and mHealth solutions are expanding and important technological tools in the development of pediatric health care [14]. To enable nationwide studies using GMA in screening infants at an increased risk of CP a mHealth app (NeuroMotion™) was developed to film and transfer film files to a secure server. However, mHealth tools need to be evaluated for usability such as accuracy, efficiency, effectiveness and user satisfaction [15].

Our study aims to investigate the usability of the NeuroMotion™ app and determine the inter-rater reliability of Prechtl’s GMA in a neonatal high-risk group.

2. Materials and methods

2.1. Study design and study population

This is a feasibility study using sub-cohorts from an ongoing prospective nationwide multicentre study on the use of the Prechtl GMA at a part of the Swedish national neonatal follow-up programme for high-risk infants, see Appendix A.1.

GMA, inter-rater reliability and film quality was assessed in a cohort consisting of the first 37 infants enrolled in the multicentre study. Some of these infants were filmed twice, and all received films were used for film quality analysis.

The first 6 app users were contacted by the researchers to ensure there were no technical difficulties. The cohort used for evaluation of user experience consist of parents to 56 infants enrolled in the multicentre study. Both parents to the remaining 52 infants were invited separately to participate. For details of included participants, see flowchart in Fig. 1.

2.2. The mHealth app - NeuroMotion™

The development of NeuroMotion™ was done by the company Unitalent, owned by Linköping University in cooperation with the first and last authors of this study. NeuroMotion™ is compatible with iOS and Android operating systems available at the App Store and Google Play. Personal login credentials are used simultaneously for every child in the different systems that allow uploading of several films. The app gives reminders by push notices when to record the infants and instructions according to GMA [5], which are short and intended to be intuitive. When the app is in filming mode, a screen filter facilitates capturing the infant’s whole body on the screen. Personal information is not handled or saved in the app. The films last for a minimum of 2 and a maximum of 3 min. This timeframe was chosen, based on studies recommending at least 2 min [16,17], together with difficulties holding the smartphone steady and keep an infant satisfied for much longer than 3 min. When recording is finished, the app user can upload the film to a secure web-based platform supporting research data management at Linköping University using REDCap (Research Electronic Data Capture) [18]. REDCap is a browser-based, metadata-driven electronic data capture software to design clinical and translational research databases.

After consent, parents to the participating infants received a letter with login credentials, recording instructions and information about downloading the app. A push notice appeared in the app when the infant was 12 weeks corrected age reminding the parents to film. Because the app was newly developed, the parents also received an email reminder in case the push notification messages were not functioning appropriately.

2.2.1. Film quality assessment

Quality assessment of the films was performed by a GMA expert (a licensed GMA tutor, advanced course) using a structured form with questions on sharpness, camera position, clothing, surrounding influences and the infants’ position and mood, according to GMA guidelines for recording [5].

2.2.2. User experience of NeuroMotion™

A web link to the questionnaire about user experience was sent by email 1–7 months after recording to one or both parents depending on available email addresses. Email reminders were sent twice. The web-based questionnaires were anonymous and not coupled to a specific film or infant. The questionnaire was supported by the tool “Survey&Report” [19] licensed for Linköping University, enabling answers customised for smartphones and computers. The survey investigated parents’ experience of the app, perceptions of security, recording instructions and technique. Most questions were designed to fit into a five-point Likert scale with possibilities to make individual comments etc (Table S2). Questionnaire items were adapted from the Baby Moves app survey described by Kwong et al [13]

2.3. GMA

2.3.1. GMA assessors

There were three on-site assessors (two paediatric physiotherapists and one paediatric neurologist) with several years of experience with children suffering from neurological deficits. All assessors had certificates from GM trust, advanced course but with
limited clinical experience using the GMA. Moreover, a GMA expert performed the same assessments as the three on-site assessors.

2.3.2. GMA procedure

All on-site assessors were blinded to the infants’ medical history, except for gestational age at birth and corrected age at assessment. However the GMA expert was totally blinded to medical history. The GMA scoring of fidgety movements was dichotomised into two categories: present (normal or abnormal (exaggerated)) or absent (sporadic or absent) [20,21]. The on-site assessors scored the films individually and in the same order as uploaded. After separate, individual scorings, a consensus decision was reached based on discussions between the on-site assessors. This approach was used to resemble a clinical setting where assessors may have varying degrees of experience and in line with the recommendation by Crowle et al. [11]. The GMA expert scored the films individually on one occasion. A second film was requested if fidgety movements were lacking, the film quality was poor or instructions were not followed. This approach was chosen to ensure the results instead of asking for longer film sequences with better possibilities for parents to film the infant in a good mood and enable filming adjustments.

2.3.3. Inter-rater reliability

Inter-rater reliability was calculated both between the on-site assessors and between the on-site assessors and the GMA expert. In addition, the assessors’ on-site consensus was also compared to the assessment of the GMA expert. This analysis resulted in six alternative calculations for inter-rater reliability.

The on-site assessors discussed the film and their ratings, resulting in a lack of blindness of the raters’ further assessments regarding the same infant. In cases in which a second film (n = 8) was requested, only the first was used to analyse inter-rater reliability.

2.4. Statistical analysis

Krippendorff’s alpha (K-alpha) was used to calculate the inter-rater reliability of the assessors [22]. Values for alpha range from –1 to 1, and the guidelines for interpreting degree of agreement described by Landis and Koch were adopted: K-alpha <0.00 = total disagreement, 0.00–0.20 = slight agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement, 0.81–1.00 = almost perfect agreement [23]. Calculations were performed using SPSS version 24 (IBM Corp, Armonk NY, USA).

2.5. Ethics

The study was approved by the Regional Ethics Review Board in Linköping, Sweden (Dnr 2018/236-31). Written consent after information about the study was mandatory. By answering the questionnaire and returning it, the parents implicitly consented to participate in the questionnaire part of the study.

3. Results

3.1. Usability of NeuroMotion™

3.1.1. Film quality

In total there were 45 films available for film quality assessment and most of these films (93.3–95.6%) showed excellent picture quality (Table 2). One film was defined as “not approved” because the “camera was not held in a stationary position”. Infant setting was excellent in almost all films (89%–100%) (Table 2). Before filming, all infants were correctly placed in a supine position.

3.1.2. User experience

Altogether 95 parents of 52 infants were addressed by email with the web-based questionnaire whereof 38 parents responded. As 31 were mothers, we assume at least one parent of 31 infants responded, hence the response rate of the individual infants is somewhere between 60 and 73%. Some 92% totally or strongly agreed that the app was easy to use and that it provided instructions easy to understand and follow (97% totally or strongly agreed) (Table 3). All parents experienced that the screen filter made it easy to understand how to position the infant. Using the app, made three parents worried about their baby's development. A minority (11% totally or strongly agreed) of the parents preferred visits from health care personnel rather than sending films through the app. One parent found that the app did not facilitate sending films to health care personnel. Most parents totally or strongly agreed (37 of 38) that it felt good to be a part of their baby’s assessment and 78% reported that the assessment of their baby’s movements using the NeuroMotion™ felt safe and secure (Table 3).

3.2. Inter-rater reliability of GMA

For the GMA phase of the study, 37 infants were initially included. One infant was excluded because the film was not possible to access and the second film asked for was not uploaded by the parents. Of the 36 remaining infants, 15 were born extremely preterm (<28 weeks), 4 very preterm (28–31 weeks), 4 moderately preterm (32–36 weeks) and 13 at term (≥37 weeks) (Table 1). Films lasted between 2 and 3 min (mean 2.54 min). The first films of the infants were recorded at a median age of 12 weeks post-term. A second film was required in eight infants and was sent at a median age of 14.5 weeks (Table 1). Absent fidgety movements were found in 3 of the 36 infants as reported together with the individual GMA ratings and final consensus in the appendix (Supplementary Table S-3).

3.2.1. Inter-rater reliability

Inter-rater reliabilities were moderate between the on-site assessors (k-alpha = 0.48; 95% confidence interval [CI] = 0.18–0.74) and between all on-site assessors and the GMA expert (k-alpha = 0.54; 95% CI = 0.34–0.71) (Table 4). Consensus reliability between the on-site assessors and the GMA expert was substantial (k-alpha = 0.72, 95% CI = 0.30–1.00) (Table 4).

4. Discussion

Overall, the app user experience was very positive, and the film quality was excellent in most films the parents recorded. According to the Landis and Koch guidelines [23], this study showed a substantial agreement between the on-site assessors’ consensus and the GMA expert for the GMA results in high-risk infants using the NeuroMotion™ app.

4.1. Usability of NeuroMotion™

4.1.1. Film quality

The NeuroMotion™ app is well suited for GMA because of the excellent film quality recorded with the app, making it possible to detect even subtle movements. According to GMA guidelines, the films were in accord with the infant and the surrounding’s necessary standardised settings [5]. Our findings may be compared to those in a study that used the Baby Moves app [13], although the two apps differs in some aspects. NeuroMotion™ has a minimalistic design with a simple colour scheme, brief step-by-step instructions
with the intention to lead intuitively to the right actions and thereby optimizing the films (Fig. 2). In that study 21/180 films could not be analysed due to poor film quality and were non-scorable. In contrast, all films in our study could be scored based on film quality. A success factor could be the simplified instructions or better intuitive visual cues. On the other hand, our study is much smaller, which needs to be considered when comparing the results of the two studies. Another explanation may be that many people today are skilled and experienced smartphone users. Moreover, smartphone software is becoming more technically sophisticated.

Table 1
Characteristics of study participants in the inter-rater reliability analysisa and film quality.

<table>
<thead>
<tr>
<th>Gestational age group at birth, n (%)</th>
<th>Extremely preterm &lt;28</th>
<th>Very preterm 28–31</th>
<th>Moderately preterm 32–36</th>
<th>Term ≥37</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 (40.5)</td>
<td>4 (10.8)</td>
<td>5' (13.5)</td>
<td>13 (35.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex, n (%)</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22 (59.5)</td>
<td>15 (40.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corrected age at recording, weeks</th>
<th>First film, n – 37</th>
<th>Second film, n – 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median – 12</td>
<td>Median – 14.5</td>
</tr>
<tr>
<td></td>
<td>Mean – 12.6</td>
<td>Mean – 14.5</td>
</tr>
<tr>
<td></td>
<td>Range – 11–15</td>
<td>Range – 13–16</td>
</tr>
</tbody>
</table>

* One participant was excluded in the inter-reliability analysis due to inaccessibility and lack of a second film.

Table 2
Quality assessment performed by a GMA expert (N = 45 films)a.

<table>
<thead>
<tr>
<th>Quality</th>
<th>Approved</th>
<th>Excellent</th>
<th>Good</th>
<th>Not approved</th>
<th>Indistinct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharpness</td>
<td>43 (95.6)</td>
<td>2 (4.4)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camera is held in a correct position</td>
<td>42 (93.3)</td>
<td>3 (6.7)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camera is held in a stationary position</td>
<td>42 (93.3)</td>
<td>2 (4.4)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clothing</td>
<td>44 (97.8)</td>
<td>1 (2.2)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background/blanket</td>
<td>40 (88.9)</td>
<td>5 (11.1)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supine position</td>
<td>45 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surrounding influences</td>
<td>43 (95.6)</td>
<td>2 (4.4)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby’s mood</td>
<td>44 (97.8)</td>
<td>1 (2.2)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* All 37 infants were included, 8 infants were filmed twice due to lacking fidgety movements, poor film quality or because instructions were not followed.

Table 3
Results from the questionnaire on the usability of the NeuroMotion™ app.

<table>
<thead>
<tr>
<th>Statements</th>
<th>Totally agree</th>
<th>Strongly agree</th>
<th>Partly agree</th>
<th>Low degree of agreement</th>
<th>Disagree</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In general, the NeuroMotion™ app was easy to use.</td>
<td>23 (60.5)</td>
<td>12 (31.6)</td>
<td>3 (7.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The push notices were helpful as a reminder when to film.</td>
<td>15 (39.5)</td>
<td>6 (15.8)</td>
<td>2 (5.3)</td>
<td>4 (10.5)</td>
<td>1 (2.6)</td>
<td>11 (28.9)</td>
</tr>
<tr>
<td>The number of push notices to remind of filming was sufficient.</td>
<td>15 (39.5)</td>
<td>6 (15.8)</td>
<td>1 (2.6)</td>
<td>1 (2.6)</td>
<td>4 (10.5)</td>
<td>4 (11)</td>
</tr>
<tr>
<td>It was good to have an email with information about the filming date to prepare me for filming.</td>
<td>27 (71.1)</td>
<td>5 (13.2)</td>
<td>2 (5.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was easy to understand the instructions in the app.</td>
<td>28 (73.7)</td>
<td>9 (23.7)</td>
<td>1 (2.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The instructions were easy to follow.</td>
<td>28 (73.7)</td>
<td>8 (21.1)</td>
<td>2 (5.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The screen filter made it easy to understand how to position my baby when filming.</td>
<td>33 (86.8)</td>
<td>5 (13.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had access to further information about the NeuroMotion™ app.</td>
<td>8 (21.1)</td>
<td>3 (7.9)</td>
<td>18 (47.4)</td>
<td>2 (5.3)</td>
<td>7 (18.4)</td>
<td></td>
</tr>
<tr>
<td>Reminders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was easy to film our baby in this specific way.</td>
<td>20 (52.6)</td>
<td>16 (42.1)</td>
<td>2 (5.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There was no problem uploading the film.</td>
<td>26 (68.4)</td>
<td>7 (18.4)</td>
<td>3 (7.9)</td>
<td>1 (2.6)</td>
<td>1 (2.6)</td>
<td></td>
</tr>
<tr>
<td>The NeuroMotion™ app facilitates the transferring of film/films to health care personnel.</td>
<td>25 (65.8)</td>
<td>9 (23.7)</td>
<td>3 (7.9)</td>
<td>1 (2.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptions of participation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I prefer visits from health care personnel to film and see my baby than to send films with the NM app.</td>
<td>3 (7.9)</td>
<td>1 (2.6)</td>
<td>10 (26.3)</td>
<td>7 (18.4)</td>
<td>17 (44.7)</td>
<td></td>
</tr>
<tr>
<td>To use the NeuroMotion™ app worried me about my baby's development.</td>
<td>1 (2.6)</td>
<td>2 (5.3)</td>
<td>9 (23.7)</td>
<td>11 (28.9)</td>
<td>15 (39.5)</td>
<td></td>
</tr>
<tr>
<td>The NeuroMotion™ app increased my awareness of my baby’s development.</td>
<td>6 (15.8)</td>
<td>8 (21.1)</td>
<td>10 (26.3)</td>
<td>9 (23.7)</td>
<td>5 (13.2)</td>
<td></td>
</tr>
<tr>
<td>It feels good to be a part of the assessment of our baby.</td>
<td>29 (76.3)</td>
<td>8 (21.1)</td>
<td>1 (2.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To know that the assessment of my baby’s movements was made using the app felt safe and secure.</td>
<td>20 (52.6)</td>
<td>10 (26.3)</td>
<td>6 (15.8)</td>
<td>1 (2.6)</td>
<td>1 (2.6)</td>
<td></td>
</tr>
<tr>
<td>The questions in this survey were easy to understand.</td>
<td>28 (73.7)</td>
<td>8 (21.1)</td>
<td>2 (5.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Respondents did not receive push notices.
* Respondents did not receive an email.
4.1.2 User experience

The survey results reveal overall parental satisfaction with NeuroMotion™. To our knowledge, the Baby Moves app is the only app used for remote GMA with previously published results about app usability [13]. Parents found the instructions, supported by the screen filter (for positioning of the infant), within NeuroMotion™ easy to understand and follow, which was also the case with the Baby Moves app [13]. The format of the screen filters in the two apps is somewhat different and the filter in NeuroMotion™ accommodates infants’ different positions. The NeuroMotion™ app serves the infants’ needs as parents can choose when and where to film and when the child is awake and alert. Parents’ filming experiences at home and being part of the assessment are important. Our findings of high parental satisfaction are consistent with a previous study [24] investigating parents’ perspectives on taking part in early neurodevelopmental screening. Opinions varied on whether the filming with the NeuroMotion™ app increased the parents’ awareness about their babies’ development. This discrepancy in results was also found when evaluating the Baby Moves app [13]. Having a child being at risk of permanent disability due to a brain insult and staying at a neonatal intensive unit, may influence parental anxiety and vulnerability. Parents’ participation in early screening may increase their awareness of their infant’s risk for neurodevelopmental disability, prepare them emotionally and motivate them to early interventions [24]. Our results indicate that a minority of parents would have preferred visits from health care personnel instead of filming their infant with an app. Using the NeuroMotion™ and Baby Moves app resulted in a minority of the parents feeling worried about their baby’s development [13]. Nevertheless, some parents’ comments demonstrated the importance of a combination of using the app and having face-to-face meetings.

Telemedicine and mHealth solutions are expanding and are seen as important technological tools in pediatric health care development [14]. Parents of children with genetic diseases [25] view mHealth as a time saver and an essential health care component. Thus, research on usability and implementation of these tools is vital regarding utility issues such as accuracy, efficiency, effectiveness and user satisfaction [15]. The development of family-centred care, regardless of geographical distance, can also be facilitated using technical solutions (such as the NeuroMotion™ app and mHealth). Such solutions can provide specialized care opportunities, offer access to targeted resources and streamline health care resources.

Technical problems such as push notices sometimes malfunctioning (an issue now resolved) can have an impact on usability. Approximately 13% of users had some difficulties uploading the films but eventually succeeded, which might be due to issues with the internet connection. Work is underway for a solution that will allow the film uploading when the internet connection is stable. Stability is important so that the app can also be used where the internet connection is less reliable.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krippendorff’s alpha (K-alpha) was calculated for inter-rater reliability (n = 36).</td>
</tr>
<tr>
<td>Assessors</td>
</tr>
<tr>
<td>Assessor 1 + 2 + 3</td>
</tr>
<tr>
<td>Assessor 3 + GMA expert</td>
</tr>
<tr>
<td>Assessor 2 + GMA expert</td>
</tr>
<tr>
<td>Assessor 1 + GMA expert</td>
</tr>
<tr>
<td>Assessors 1 + 2 + 3 + GMA expert</td>
</tr>
<tr>
<td>Consensus on-site assessors + final decision GMA expert</td>
</tr>
</tbody>
</table>

* According to guidelines by Landis and Koch [23].
* Consensus was reached in eight cases based on a second film.
4.2. Inter-rater reliability in GMA

GMA is a qualitative analysis very much based on the assessor's visual perceptions of movements. Our results show that the final decision should be based on a consensus of at least two assessors. There was substantial agreement between consensus decisions but only moderate agreement between all assessors. Our on-site consensus results agree with two studies estimating GMA's inter-rater agreement in the flidgety period showing substantial agreement [10,12] while a third showed almost perfect agreement [11]. The variability in the results of inter-rater agreement can be influenced by statistical methods for estimation, study population, blindness and experiences of the assessors.

Having all on-site assessors blinded to medical history, except for gestational age and a GMA expert totally blinded, is unique to our inter-rater reliability study compared to other similar studies [10–12]. This blind approach is important when studying the true reliability of an assessment. With background knowledge of an infant's medical history, there is a risk of unintentionally allowing other known risk factors for CP to influence decisions on the absence of flidgety movements.

For statistical analysis of inter-rater reliability, we chose the K-alpha method as it considers sample size and number of assessors, not just percentage of agreement. Many studies report the percentage of agreement between assessors, making it difficult to use when comparing studies of different cohort sizes and predicting inter-rater reliability in future studies. Other statistical methods can be used to calculate the strength of agreement (reliability). We believe that K-alpha is best suited given that it can be used for two or more assessors and handle missing data [22]. The strength of agreement for categorical data can be compared using cut-off values for kappa statistics (as described in Landis and Koch) [23] between the results of Cohen's kappa and K-alpha.

Another difficulty comparing studies is differences in the expected prevalence of flidgety movements in the study cohort. The children in the Swedish national neonatal follow-up programme for high-risk infants [26] have a future prevalence of 7–10% for CP [27,28]. Our cohort was recruited from this group of children and would therefore have a similar prevalence of CP, and consequently, a lack of flidgety movements. The prevalence rate of CP in the above-mentioned GMA studies on inter-rater reliability is most likely higher than we expect it to be in our multicentre study. We intend to include more than 1000 infants incorporated in the Swedish national neonatal follow-up programme for high-risk infants [26], allowing us to estimate inter-rater reliability in GMA for different subgroups.

4.3. Strengths and limitations

A reason for the high usability of NeuroMotion™ might be that it was developed in close cooperation between a computer coding specialist, a cyber security engineer and specialists in the field of pediatric neurology and licenced GMA assessors. Inter-rater reliability analysis was based on small sample size. Still, the statistical method used for this purpose was suitable for the small sample size.

Having information on how to make films in your own language is important. In the NeuroMotion™ app you can only choose between Swedish and English. This language restriction is a limitation because only participants mastering these languages can use it.

A limitation of the survey was the low response rate and lack of the respondents' descriptives on social demographic data, leading to a risk of response bias. Since respondents are anonymous, we cannot provide descriptives of the respondents nor clarify reasons for not responding.

It might be that parents who decided to answer the questionnaire were more concerned about their infant's health and development than those who did not respond. Thus, the participating parents might be more motivated, positive and confident using the app. Another reason for low response rate could be that only one of the parents decided to answer. The wide time window between filming and receiving the questionnaire, may have led to a response bias as well as influenced the response rate, although for the majority of infants the time window was less than 4 months. Another limitation might be that we were aware of the gestational age. However, according to the inclusion criteria all the participants have a high risk of neurological sequel as CP also of other causes than prematurity. Awareness of gestational age should therefore not influence the assessment of flidgety movements significantly.

Further studies are needed to evaluate GMA in the neonatal high-risk infant population in light of early diagnosis supported by telemedicine to improve knowledge about the assessment's accuracy in this specific setting, striving for effective and qualitative health care. Assessment of NeuroMotion™ has covered important usability measures, such as accuracy (film quality), efficiency and effectiveness (recording instructions and technical issues) and user satisfaction.

5. Conclusion

The NeuroMotion™ app is an excellent tool for filming infants both at home and in a clinical setting thanks to good technical quality of the films as well as intuitive instructions according to standardised GMA guidelines. Furthermore, parents report high degree of user satisfaction. The approach with GMA assessors reaching a consensus is recommended given the achieved substantial agreement with a GMA expert. Using the NeuroMotion™ app in a home setting makes it possible to study GMA in larger cohorts of neonatal infants.

Authors contribution

KAS: Conception, study planning, literature search, study design, data collection, data analysis, data interpretation, writing. (area of expertise: pediatric neurology), AFB: Data collection, critical review of the manuscript, A-CE Interpreting results, critical review of the manuscript (area of expertise: pediatric neurology), MO: Study design, data collection, data interpretation, critical review of the manuscript (area of expertise: pediatric physiotherapy), HEK: Conception, study planning, study design, data collection, data analysis, data interpretation, critical review of the manuscript (area of expertise: pediatric neurology).

Writing Assistance

None.

Grant support (funding)

KAS: A private donation through the Knut and Alice Wallenberg Foundation for research on stroke and stroke-causing factors in infancy and a grant from Medical Research Council of Southeast Sweden (FORSS–859871), AFB: None, ACE: None, MO: None, HEK: A private donation through the Knut and Alice Wallenberg Foundation for research on stroke and stroke-causing factors in infancy, a grant from the Region Östergötland Research Council and a grant from Medical Research Council of Southeast Sweden.
Appendix A. Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejpn.2021.05.006.

References


