DR. VIVIANA MARCHI (Orcid ID : 0000-0003-0900-5206)

Automated pose estimation captures key aspects of General Movements at 8-17 weeks from conventional videos

Viviana Marchi*[1,2], Anna Hakala[3], Andrew Knight[3], Federica D'Acunto[2], Maria Luisa Scattoni[4], Andrea Guzzetta[2,5], Sampsa Vanhatalo[6]

1 Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa, Italy;

2 Department of Developmental Neuroscience, IRCCS Stella Maris Foundation Pisa, Italy;

3 Neuro Event Labs Oy (2712284-1), Tampere, Finland;

4 Research Coordination and Support Service, Istituto Superiore di Sanità, Rome, Italy;

5 Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy;

6 Department of Clinical Neurophysiology, Children´s Hospital, University of Helsinki and Helsinki University Hospital (HUH), Helsinki, Finland.

* Corresponding author:

Viviana Marchi, Department of Developmental Neuroscience, Stella Maris Scientific Institute, IRCCS Stella Maris Foundation; Viale del Tirreno 331, 56128 Calambrone, Pisa (Italy)

Tel. 050/886226-239

E-mail address: vmarchi@fsm.unipi.it

ABSTRACT

Aim

General movement assessment requires substantial expertise for accurate visual interpretation. Our aim was to evaluate an automated pose estimation method, using conventional video records, to see if it could capture infant movements using objective biomarkers.

Methods

We selected archived videos from 21 infants aged 8-17 weeks who had taken part in studies at the IRCCS Stella Maris Foundation (Italy), from 2011-2017. Of these, 14 presented with typical low-risk movements, while seven presented with atypical movements and were later diagnosed with cerebral palsy. Skeleton videos were produced using a computational pose estimation model adapted for infants and these were blindly assessed to see whether they contained the information needed for classification by human experts. Movements of skeletal key points were analysed using kinematic metrics to provide a biomarker to distinguish between groups.

Results

The visual assessments of the skeleton videos were very accurate, with Cohen's K of 0.90 when compared with the classification of conventional videos. Quantitative analysis showed that arm movements were more variable in infants with typical movements.

Conclusion

It was possible to extract automated estimation of movement patterns from conventional video records and convert them to skeleton footage. This could allow quantitative analysis of existing footage.

Keywords: automatic pose estimation, cerebral palsy, general movements, infants, spontaneous motor activity

KEY NOTES

- This study evaluated whether using automated pose estimation could effectively assess general movement using conventional video records.
- We used archived videos from 21 infants aged 8-17 weeks who had taken part in Italian studies: 14 had typical low-risk movements, while seven had atypical movements and were later diagnosed with cerebral palsy.
- The visual assessments of the skeleton videos were very accurate when compared with the classification of conventional videos.

BACKGROUND

There is growing awareness of the importance of recognising children who at risk of developing cerebral palsy (CP) at an early-stage and providing therapeutic interventions. Despite all the advances in biochemical and neuroimaging methods, CP remains a condition that is defined by abnormal neurological function and, conversely, spontaneous activity remains the most reliable marker of emerging CP. Assessment of spontaneous general movements (GMs) in an awake infant at the age of the fidgety pattern, from the second to

the fifth month after term, has the highest predictive power and sensitivity for later neurological outcomes (1,2). This assessment has so far been based on visual recognition of infant movement patterns, namely Gestalt perception, while lying supine in standardised settings. Despite its high reliability and reproducibility (3), a wider uptake of this approach has been hampered by the need for substantial training to achieve and maintain sufficient expertise.

Non-invasive recording, combined with computational data analysis, presents a new approach for kinematic analysis and offers an attractive way to study infant mobility (4). Over the last decade, a number of computer-based methods have attempted to emulate visual GM analysis, either by employing miniaturised movement sensors worn on the body (4) or by assessing movements directly from a video image (5). These have been technically challenging because of issues in the sensor systems or computational tasks or because the mere complexity of the instrument precludes its implementation in a real-life clinical setting. Recent progress in computational methods, in particular the introduction of machine learning approaches, has led to dramatically improved automated image analysis. It is now possible to reliably track objects and even body parts with high accuracy from normal two-dimensional or three-dimensional video streams (6). This provides the possibility of converting ordinary video recordings into kinematic time-series, which can be used with the same analytic approaches employed with other biosignals.

A particular challenge in computational analyses is how to preserve the details of spontaneous movement patterns with sufficiently high fidelity so that the record can be used for diagnosis by trained clinicians. One technique that could be used to increase the availability of detail is pose estimation, an algorithmic procedure that extracts a skeletal model from an image of a human body. Ideally, the video image that is used should also include in-depth information to allow full three-dimensional recovery of the limb movement trajectories (7). While such camera technology is already available for prospective clinical studies (7), most of the existing GMs video material and current clinical practice is based on

recording conventional two-dimensional colour videos. A human expert trained in GMs analysis is able to make GMs classification from such videos, suggesting that video information on its own does contain the elements needed for GM classification (8).

This study was an exploration to test whether novel pose estimation methods (6) were able to recover an infant skeleton in such detail that just its movements could be classified by a blinded human observer. If that was the case, we wanted to further test whether the quantified metrics of such skeleton movements could be used to distinguish between normal and abnormal spontaneous movements.

METHODS

The experimental flow chart is summarised in Figure 1.

Subjects

We selected 21 GMs video recordings from infants that had taken part in studies at the Infant Neurology Section of the IRCCS Stella Maris Foundation in Calambrone, Pisa, Italy (Table S1). The recordings were selected from two groups of infants that had been recorded at the fidgety age. One group consisted of recordings of children who later developed cerebral palsy, while the other group consisted of recordings of low-risk and term-born infants, recruited as controls for other ongoing studies. The first group of seven infants was selected from a cohort born between 2013 and 2016 with abnormal movement patterns due to absent fidgety movements. They were later diagnosed with cerebral palsy and showed a documented brain lesion on a magnetic resonance imaging (MRI) scan. The second group comprised 14 infants with low neurodevelopmental risk, who presented fidgety movements and a typical motor repertoire. They were recruited in 2011-2017 as controls in an ongoing study on the early signs of autism, which was carried out by the Italian network for the early

recognition of autism spectrum disorders and co-ordinated by Dr Scattoni at the Italian Institute of Health. The parents provided written consent for using the video recordings for clinical and research purposes.

The video recordings were performed using the standard protocol devised by Prechtl and the General Movements Trust (8). General Movements were recorded during a session of spontaneous motility, when the infants were fully awake but were not crying or fussing. The ideal recording setting was with the infant lying supine and possibly naked, or in a bodysuit not covering the limbs, and the camera positioned about one metre above the infant. The surrounding adults were instructed not to touch the infant or enter the video scene. All videos lasted for at least two continuous minutes, while the infants were in a quiet awake state, and the mean length of the recordings was five minutes and 23 seconds (Table S1). However, the data comprised a collection of videos from a number of different years and studies, which means that the setting was less than ideal. There were instances when the infants had alternative clothing, there were occasional interruptions by adults in the scene and the camera, or its position, changed between recordings. These challenged the later analysis, prompted the development of quality control mechanisms and resulted in the exclusion of 14 of the 21 infants from the quantitative assessments. However, variations of these kinds are expected in most historical datasets and they allowed us develop the methodological process more robustly for such wider use. The mean age of the infants in the videos that were selected was 12 weeks post-term (range 8-17 weeks).

Pose estimation and video skeletonisation

Pose estimation was carried out with OpenPose, an open-source computer vision software developed by Cao et al in 2017 (6). It is trained to detect human joint locations, namely key points, on single images. In the version used for our analysis, the 18 key points we used were: the nose, each eye, each ear, the neck, each shoulder, each elbow, each wrist, each

hip, each knee and each ankle. A detailed description of the model can be found in Cao et al (6).

In this study, the OpenPose framework was further enhanced with additional software elements customised for infants by Neuro Event Labs Oy, Tampere, Finland. In particular, this customisation included the ability to estimate different infant body-to-limb proportions, as well as to accommodate movement ranges that differed from adults and to present frequent mutual occlusions among limbs. In addition, customised adaptation was needed to compensate for technical issues, such as instability in handheld video recordings, as well as varying video resolutions and suboptimal colour contrasts between the infant's clothing and the background. Accordingly, the process was adjusted to stabilise and normalise the view in order to produce comparable skeletons. The infant body model was stabilised by fixing the origin at the neck key point, providing image stabilisation to counteract camera movements. To reduce variance due to different video resolutions and camera distances, the infant's size was normalised by quantifying the distance between the shoulders as the base unit of distance measurement. Notably, most of these technical issues could be readily overcome in prospective video recordings by controlling the camera position and video quality.

The locations of the skeletal key points were processed for quantified movement analysis using standard kinematic metrics, such as velocity, acceleration and jerk. We elaborate on this later. As the anatomical key points were often located in areas with low visual contrast, a high-frequency jitter could occur between consecutive frames. This jitter was removed with a moving average filter with a window size of 30 frames. The moving average filter also decreased, but did not resolve, errors induced by the occlusion of key points.

The consecutive frames of the pose estimated skeleton were used to generate a skeleton video (Figure 2b), which was used for the inter-rater agreement study. An example of the skeleton video is provided (Video S1), while the time-series signal extracted from the key points was recorded in order to compute quantitative descriptors of the movements.

Visual video assessments

All the original videos that were selected were re-assessed to generate the clinical benchmark, namely the gold standard. The GMs' assessment was performed by two experienced experts (VM, AG), certified by the General Movements Trust, according to the standard methodological principles of Prechtl's method. All the videos were renamed for the blind assessments and the assessors did not know the child's identity or any of their clinical information. The scorers evaluated spontaneous motility as the presence or absence of fidgety movements. The third option, abnormal fidgety, was not observed in our cohort. In cases of disagreement the final score was reached by consensus and the final categories were globally classified as typical, when fidgety movements were present, and atypical, if fidgety movements were absent.

After the videos had been processed, two experts certified by the General Movements Trust (VM, FD), carried out blind, independent assessments of the skeleton videos that had been re-coded for full anonymity. One of the scorers (FD) had not even participated in the visual analysis of the conventional videos in this dataset. The scoring of the skeleton videos was only carried out to determine a dichotomic group assignment by assessing the quality of the overall movement repertoire. While the skeleton video assessment could not be fully equivalent to assessing movements in a real human subject, our experts were intuitively guided by the Assessment of Motor Repertoire developed by the Prechtl group (9). Verbal descriptions of the perceived normality included motor repertoires that were fluent, smooth and variable, while abnormal movements were monotonous, jerky or stiff (10–15). Notably, the experts did not attempt to recognise particular fidgety movements as they were technically more challenging due to their temporal and spatial ambiguity. They also demonstrated a subtle physical appearance relative to the spatial accuracy available in the pose estimation from our present two-dimensional video recordings, as discussed later.

No consensus was sought, as we wanted to measure both the classification accuracy and inter-rater agreement between our experts. The benchmark, or gold standard, for correct classification was taken from the assessment of original GMs videos.

Quantitative characterisation of the GM pattern

We plotted all the key points generated by the pose estimation for every frame of the video, recording them as a point cloud in order to visually estimate the spatial spread of the movement (Figure 3b). After a visual inspection of these point clouds, we used kinematic metrics to select a set of measures to be computed. These metrics included the time-series values of: the distance between the wrists, the spatial centroid of the key points, the normalised locations of the elbows and wrists, the distance between shoulders and wrists, the spectrum of the location of wrists and the standard deviations of velocity and acceleration of the shoulder, elbow and wrist movements. After inspecting the quality of the derived metrics, we chose to focus on the upper limb data, as the pose estimation of the lower limbs was not accurate enough with the regular two-dimensional videos. Despite the apparently good-quality cloud plots, there seemed to be disturbing levels of technical noise from jitter in the pose estimation output. We therefore needed to develop a quality control method at this stage. The first step was to identify the extent of the jitter in the cloud plots generated from apparently good-quality video recordings. This allowed us to define limits for physically impossible movements in the videos (Figure 3a). For instance, we empirically placed a threshold based on velocity estimation of the wrists at 3.3 m/s between consecutive frames. Videos with more than 5% of data defined as physically impossible were excluded from the analysis to limit the effect of outliers. This yielded seven recordings, four from the typical group and three from the atypical group.

Statistical analysis

Inter-rater reliability was assessed using skeletonised videos of the infants. We computed Cohen's K, inter-rater agreement and binomial probability. Between-group differences in the upper limb kinematic parameters were tested by using a non-parametric Mann-Whitney U test for small samples. The software we used was the Stats package in R version 3.4 (R Foundation for Statistical Computing, Vienna, Austria; 16).

RESULTS

The two scorers were able to distinguish normal from abnormal movements with very high accuracy, identifying the correct categories in 18 and 19 of the 21 videos respectively. The global inter-rater agreement was 95% and the skeleton videos agreed with the original GM assessments in 90% of cases. Taken all together, we obtained Cohen's K of 0.90, with a combined error in only three cases. This was highly significant, with a binomial probability of p<0.001.

The errors in the skeleton video scores occurred both ways: one scorer assigned normal movement to an infant judged abnormal in the original video. The other scorer stated that two skeleton videos showed abnormal movement patterns when the original videos had been classified as normal.

In some cases, the quantitative assessment of the movements showed estimated movements patterns that could not reflect physiological movements. They were obviously caused by technical issues in the video recording that could not be fully compensated for in the pose estimation. After strict and objective exclusions of cases with such technical uncertainties, we could only include seven videos in our preliminary assessment of quantitative comparisons: four with normal and three with abnormal movements. Assessment of the spatial distributions of key points showed that the overall scatter might be

different between groups and an example of a cloud plot is shown in Figure 3b. However, the distributions of key point coordinates appeared to overlap too much between the infant groups (Figure S1) to allow us to determine direct classifications from the key point coordinates. Therefore, we further assessed the movements and found that the wrist-related measures were significantly different between the groups and the standard deviations of velocity and acceleration were significantly higher in infants with typical movements (p=0.03) (Figure 3c). The distinction between the groups seemed to increase from proximal to distal anatomical key points (Figure 3c).

DISCUSSION

Our explorative study shows that it is possible to perform accurate automatic pose estimation from a selective sample of video records. That is, the generated skeleton videos retain salient information allowing for an accurate visual classification by human experts. The inter-rater agreement rate was high, even in the absence of details on the fidgety movements, which are typically considered crucial for assessing the hallmarks of GMs. Moreover, the quantitative measures generated by the automatic pose estimation provided a means for classifying the extremes of the spectrum of GMs quality.

We observed a high level of accuracy in the visual assessment of the skeleton videos. This was in keeping with the knowledge that the holistic perception of GMs, namely the visual Gestalt, adapts to ongoing changes in recording angles, as well as to jittering during recordings. For example, the visual assessments seemed to automatically ignore obvious and rare misallocations of the limbs generated in the skeleton videos. The high inter-rater agreement seemed to outlast the technical inaccuracies, which were probably related to the overall sub-optimal quality of the records. This suggested an underestimation of the real potential of the presently used pose estimation method. Furthermore, the fact that they were able to selectively disregard unnatural movements and locations could inform the

development of quantitative analysis methods that target the jitter and inter-frame disparity by building models that work on series of images (videos) rather than on single frames.

It is interesting that our visual assessment did not focus on the specific features of Fidgety movements, namely the multi-planar, small amplitude movements of all the joints visible in typically developing babies between nine and 20 weeks of post-term age. Indeed, Fidgety movements are considered to be the hallmark of normal movement at about three months of age in traditional GMs analyses (18). However, they were hardly visible in the skeleton videos produced by our pose estimation method, if at all. Yet, the experts were able to identify normal and abnormal infants, suggesting that there are latent features in the movement repertoire that are less consciously identified, but are nevertheless part of the Gestalt perception in the visual GMs analysis. Interestingly, several studies have attempted to improve qualitative movement analysis by including features of movement beyond the pure evaluation of the Fidgety pattern (11–15,19).

Our human rater assessments suggest that our skeleton videos captured the diagnostically essential information needed, which would support their use in developing novel computerised classification software. However, building such classifiers with machine learning methods requires much larger datasets. These need to provide high technical quality and well documented multi-rater classifications for testing human equivalence in diagnostic accuracy, as explained by Tapani et al (20). Indeed, the relatively low number of videos with high technical quality was a limitation of our present study and did not allow us to develop a diagnostic classifier of GMs *per se.* However, the quantitative measures generated by the automatic pose estimation of the high-quality videos, available in about one-third of the sample, meant that we were able to classify the extremes of quality of the GMs spectrums by using the information from the movement of the upper extremities. Furthermore, the present mix of lower and higher quality video clips supported the development of automated quality control, which could be implemented in future studies.

Our study supports the work by Adde et al (5,17), by showing that it was possible to obtain sensor-free kinematic quantification of infant movements from conventional video records. However, our pose estimation produced a much more detailed model of infant movements than Adde et al, which allowed us to quantify more complex movement patterns and to estimate a higher number of kinematic metrics.

In addition to providing a way to quantify movement patterns, our study also suggests that GMs videos could be converted into skeleton videos that are fully anonymised, yet retain clinically relevant movements. This may be an attractive alternative in larger scale data sharing or transmission of GMs records between sites, for example in larger clinical trials where privacy protection would limit the transmission of conventional video clips.

A sensor-free kinematic analysis based on video records is an optimal approach for developing automatic assessment of movement patterns in high-risk infants, as reviewed by Marcroft et al (4). Video recordings can be easily adapted and widely available in clinical practice. The new three-dimensional cameras that are already available (21) will allow the easy creation of three-dimensional models with better spatial resolution. The remaining challenge is the availability of reliable pose estimation methods, which are necessary if we are to produce kinematic measures for computational analyses. Our study shows that the OpenPose estimator (6) was reliable for mapping and recognising infant movements even on two-dimensional video recordings. These could be significantly improved with three-dimensional camera technology, as an additional degree of freedom would allow for assessing movements orthogonal to the viewing plane.

CONCLUSION

Automatic pose estimation from video records is available and it seems a promising and objective way to assess movements in infants with neurodevelopmental risks. Adapting the OpenPose estimator (6) so that it could be used within infants was sensitive enough to

detect and reproduce Gestalt classifiers that drive the human eye during the diagnostic decision making. This suggests that it may also be used to train automated classifiers so that they can perform the same task. This could be a promising way to develop easily accessible, non-invasive methods for screening infants with suspected compromised neurodevelopment.

ABBREVIATIONS

GMs, General movements, CP; Cerebral palsy.

CONFLICTS OF INTEREST

AH and AK work for Neuro Event Labs, which developed the pose estimation used in the study.

REFERENCES

1.   Novak I, Morgan C, Adde L, Blackman J, Boyd RN, Brunstrom-Hernandez J, et al. Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. *JAMA Pediatr* 2017 Sep 1; 171(9): 897–907.

2.   Kwong AKL, Fitzgerald TL, Doyle LW, Cheong JLY, Spittle AJ. Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review. *Dev Med Child Neurol* 2018; 60(5): 480–9.

3.   Bernhardt I, Marbacher M, Hilfiker R, Radlinger L. Inter- and intra-observer agreement of Prechtl's method on the qualitative assessment of general movements in preterm, term and young infants. *Early Hum Dev* 2011 Sep; 87(9): 633–9.

4.   Marcroft C, Khan A, Embleton ND, Trenell M, Plötz T. Movement recognition technology as a method of assessing spontaneous general movements in high risk infants. *Front Neurol* 2014; 5: 284.

5.   Adde L, Helbostad JL, Jensenius AR, Taraldsen G, Støen R. Using computer-based video analysis in the study of fidgety movements. *Early Hum Dev* 2009 Sep; 85(9): 541–7.

6.   Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *CoRR*. 2016; abs/1611.08050

7.   Olsen MD, Herskind A, Nielsen JB, Paulsen RR. Using Motion Tracking to Detect Spontaneous Movements in Infants. In *Image Analysis: 19th Scandinavian Conference,*

SCIA 2015 Copenhagen, Denmark, June 15–17, 2015 Proceedings. Springer Science+Business Media. 2015, p. 410-417.

8.   Prechtl HF, Einspieler C, Cioni G, Bos AF, Ferrari F, Sontheimer D. An early marker for neurological deficits after perinatal brain lesions. *Lancet.* 1997 May 10; 349(9062): 1361–3.

9.   Einspieler C, Prechtl HFR, Bos AF, Ferrari F. Prechtl's method on the qualitative assessment of general movements in preterm, term and young infants. In *Clinics In Developmental Medicine.* London, UK. Mac Keith Press, 2005.

10.   Bruggink JLM, Einspieler C, Butcher PR, Van Braeckel KNJA, Prechtl HFR, Bos AF. The quality of the early motor repertoire in preterm infants predicts minor neurologic dysfunction at school age. *J Pediatr.* 2008 Jul; 153(1): 32–9.

11.   Bruggink JLM, Einspieler C, Butcher PR, Stremmelaar EF, Prechtl HFR, Bos AF. Quantitative aspects of the early motor repertoire in preterm infants: do they predict minor neurological dysfunction at school age? *Early Hum Dev.* 2009 Jan; 85(1): 25–36.

12.   Bruggink JLM, Van Braeckel KN, Bos AF. The early motor repertoire of children born preterm is associated with intelligence at school age. *Pediatrics.* 2010 Jun; 125(6): e1356-1363.

13.   De Vries N, Bos A. The motor repertoire of extremely low-birthweight infants at term in relation to their neurological outcome. *Dev Med Child Neurol.* 2011 Oct; 53(10): 933–7.

14.   Fjørtoft T, Grunewaldt KH, Løhaugen GCC, Mørkved S, Skranes J, Evensen KAI. Assessment of motor behaviour in high-risk-infants at 3 months predicts motor and cognitive outcomes in 10 years old children. *Early Hum Dev.* 2013 Oct; 89(10): 787–93.

15.   Yuge M, Marschik PB, Nakajima Y, Yamori Y, Kanda T, Hirota H, et al. Movements and postures of infants aged 3 to 5 months: to what extent is their optimality related to

perinatal events and to the neurological outcome? *Early Hum Dev.* 2011 Mar; 87(3): 231–7.

16.   R: The R Project for Statistical Computing. Vienna, Austria, 2008, url: https://www.R-project.org, RRID:AB_853317.

17.   Adde L, Yang H, Sæther R, Jensenius AR, Ihlen E, Cao J-Y, et al. Characteristics of general movements in preterm infants assessed by computer-based video analysis. *Physiother Theory Pract.* 2018 Apr; 34(4): 286–92.

18.   Einspieler C, Peharz R, Marschik PB. Fidgety movements - tiny in appearance, but huge in impact. *J Pediatr* (Rio J). 2016 Jun; 92(3 Suppl 1): S64-70.

19.   Einspieler C, Marschik PB, Pansy J, Scheuchenegger A, Krieber M, Yang H, et al. The general movement optimality score: a detailed assessment of general movements during preterm and term age. *Dev Med Child Neurol.* 2016 Apr; 58(4): 361–8.

20.   Tapani KT, Vanhatalo S, Stevenson NJ. Time-Varying EEG Correlations Improve Automated Neonatal Seizure Detection. *Int J Neural Syst.* 2018 Jun 24; 1850030.

21.   Olsen MD, Herskind A, Nielsen JB, Paulsen RR. Model-Based Motion Tracking of Infants. In Agapito L, Bronstein MM, Rother C, editors. *Computer Vision – ECCV 2014 Workshops: Proceedings*, Part III. Springer. 2014. p. 673-685.

Figure Legends:

Figure 1: Schematic drawing of the experimental flow chart. We conventionally selected a dataset of 21 GMs videos recorded at fidgety age and those videos were visually scored according to the GMs assessment protocol. The same videos were processed with pose estimation algorithm to obtain skeleton videos that were then judged visually, and the results were compared to the qualitative scores on the original videos, reaching an agreement of Cohen's K of 0.90. Skeletal movements with sufficiently high quality were also processed by quantitative analysis to extract potential features that could help to distinguish normal from abnormal movements, as explained in the Methods section.
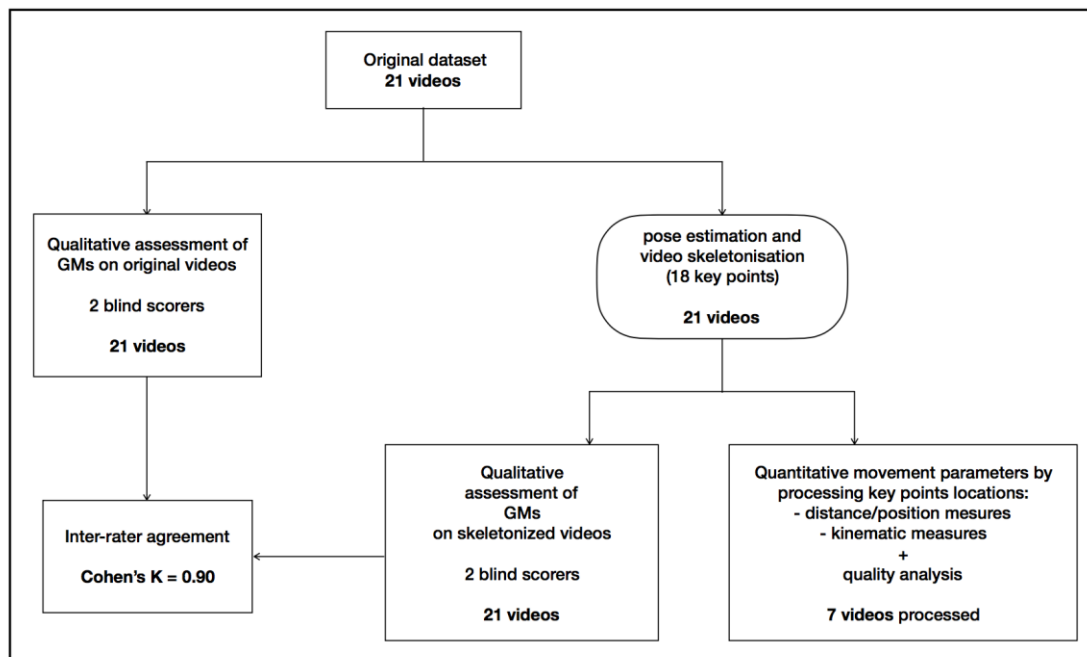

Figure 1

Figure 2: Pose estimation procedure from the video image to skeleton videos. In the original video image (**a.** left side), anatomical key points are identified (**a.** middle and right side with dots in infant's joints) and a skeleton is formed by drawing lines between these key points. This is repeated for all consecutive video frames (**b.**) that are then converted into a skeletonised video (see Video S1) for later assessment. The red dots in the middle figure of **a.** depict the wrist key points as these were the movements that were most robustly different between the infant groups.
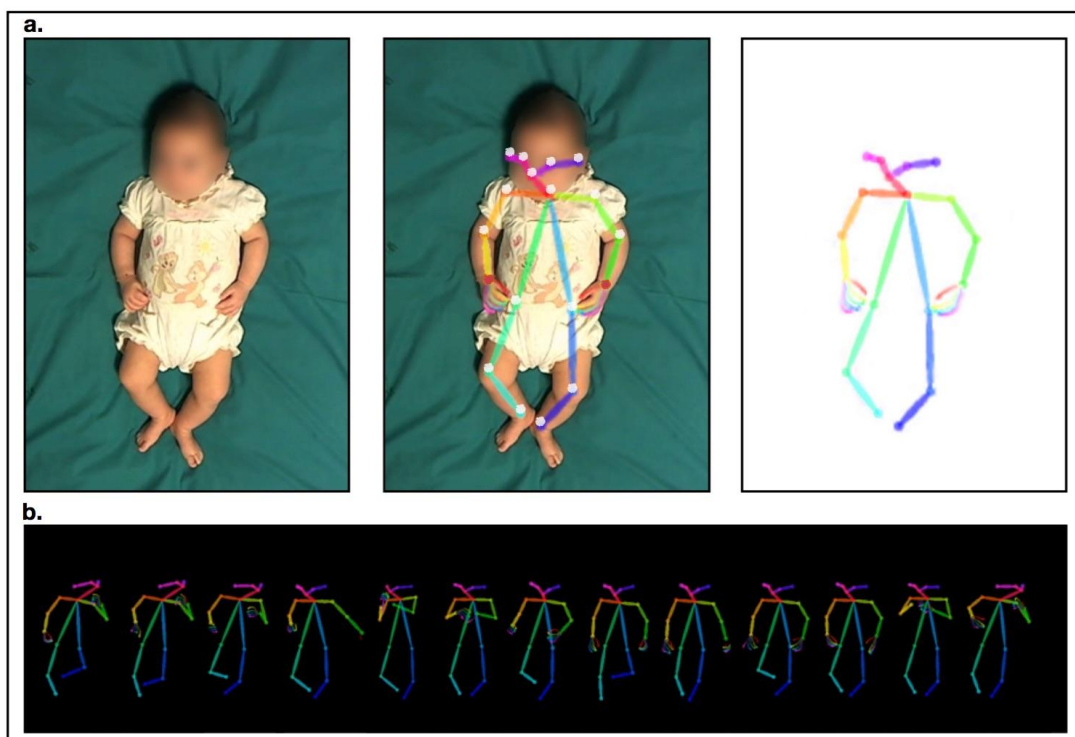

Figure 2

Figure 3: a. Video quality assessment based rejecting movements that were physically impossible, shown in good quality (left) and poor quality (right) videos. I. global distribution of movements of the key point for the left elbow. II. distribution of the movements of key points after defining a threshold for physically impossible movements, based on maximal velocity estimation between consecutive frames to define possible infant movements: threshold for wrists at 3.3 m/s and for elbows at 1.7 m/s. The excluded dots are then considered as physically impossible key point movements. Videos containing more that 5% of physically impossible movements were excluded from the analysis to limit the effect of outliers. **b.** density plot of the locations of key points over time for typical (left) and atypical (right) GMs pattern. **c.** distribution of the standard deviation of velocity (left) and acceleration (right) for the shoulder, elbows and wrists key-points in typical (orange) and atypical (grey) GMs. Subjects with typical GMs presented larger standard deviations for wrists movements compared to atypical subjects.
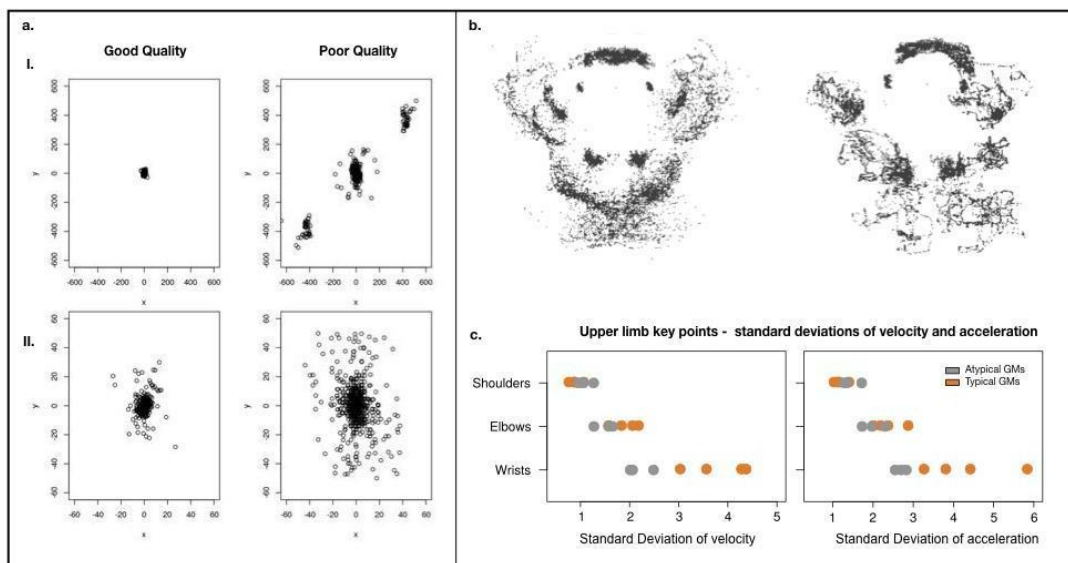

Figure 3