

A Protocol for Structured Observation of Motor Performance in Preterm and Term Infants

Interobserver Agreement and Intraobserver Consistency

K. Hammarlund, K. Persson G. Sedin and B. Strömberg

Department of Paediatrics, Uppsala University Children's Hospital, Uppsala, Sweden

ABSTRACT

A new protocol for structured observation of motor performance, for use both in term and preterm infants, has been tested regarding interobserver agreement and intraobserver consistency. Ten different motor items are assessed concerning the developmental level as described in the protocol. Any deviations from the description of the level are noted. The two main observers twice reassessed video recordings of 17 infants who had primarily been assessed by them both in an ongoing follow-up study. The two observers arrived at the same level of development in 75% of the observed items at the first reassessment and in 81% at the second, and agreed concerning deviation in 73 and 79% of the observations at the two reassessments, respectively. Each observer arrived at the same level of development at all three assessments in 62% of the observed items. Thus when used by these two observers the protocol has an acceptable degree of interobserver agreement and intraobserver consistency.

INTRODUCTION

With an increasing number of newborn infants undergoing and surviving neonatal intensive care (7, 10), many studies have been made to establish how the infants develop during later life (6, 8, 11). Different neurological tests (1, 3, 12) and developmental scales (2, 4, 5, 9) have been used for this purpose. To study the motor performance of infants in detail, taking into consideration not only quantitative aspects but also the quality of the motor function, a new protocol for structured observation of motor performance during the first 10 months after term has been developed. It is designed so that each part of the body is assessed separately concerning the level of motor development, any deviation from the description of the level and the type of deviation, and any asymmetry. The protocol has been described in detail elsewhere (13).

The protocol is currently being applied in an ongoing study to evaluate the motor development of all infants who have undergone neonatal intensive care in the county of Uppsala in Sweden. The aim of this paper is to present results of application of the protocol as to interobserver

agreement and intraobserver consistency; these concern judgements made by the two main observers in the follow-up study.

MATERIAL AND METHODS

In the study of motor performance in our follow-up programme a physician and a physiotherapist, both with special interest in developmental neurology, are the main observers. One of them is always present at the assessment of each infant, usually together with a neonatologist. Each assessment is recorded on videotape. In the present study the video recordings of 17 infants, who were primarily assessed in the follow-up study by both main observers together, were sampled from all recordings by an assistant. Reassessments were made separately by these two observers from the video recordings on two occasions more than one year apart. The selected recordings were made when the infants had a corrected age (age from estimated due date of delivery) of 0 (n=3), 2 (n=4), 4 (n=4), or 6 (n=6) months. The first reassessment was made more than a year after the primary assessment. At reassessment, the observers were not informed about the identity, age or history of the infant.

Ten different motor items were assessed in each infant, namely, the motor performance of the a) head, b) arms and hands, c) trunk and d) legs and feet in the supine and the prone position, and the motor performance of the whole body during sitting and locomotion. For each item the developmental level was estimated according to the protocol. The observers also noted whether the motor performance corresponded exactly to the description of the level in question or whether there was any deviation. This result was graded as no deviation, suspected deviation or clear deviation. The type of deviation was also noted, but this question has not been evaluated in the present study. The observers also noted whether there was any asymmetry in the motor performance.

Interobserver agreement was assessed by comparing the results obtained by the two observers at each of the two reassessments from the video recordings. Seventeen recordings were reassessed on the first occasion, but on the second occasion one 6-month infant was not assessed, leaving 16 for comparison.

Intraobserver consistency was assessed by comparing the results of each of the two observers at the two reassessments. Comparison was also made between the primary assessment (made by the two main observers in unison) and the reassessments. The results for 13 infants were included. Of the original 17 recordings four had to be excluded because of errors made by the assistant, once concerning the number of the infant's recording (which meant exclusion of two recordings) and once concerning the age of the infant in the recording shown to both observers (exclusion of two recordings).

RESULTS

Interobserver agreement

At the first reassessment the two observers arrived at the same level in 75% of their observations (128/170) and at the second reassessment in 81% (131/160). There was no clear difference in the agreement at different ages, nor was there any item that differed markedly from the others. Observer A arrived at a higher level than observer B in 14% of the observations at the first reassessment and in 9% at the second, and at a lower level than observer B in 9% at the first reassessment and in 7% at the second.

The observers agreed in their assessment regarding the degree of deviation from the description of the level in 79% of the observations at the first reassessment and in 73% at the second.

Observer A found a higher degree of deviation than observer B in 6% of the observations at the first reassessment and in 11% at the second, and a lower degree in 14% of the observations at the first reassessment and in 15% at the second.

Regarding asymmetry, the two observers arrived at the same result in 87% of the observations at the first reassessment and in 82% at the second.

Intraobserver consistency

On comparison of the two reassessments, it was found that observer A came to the same level of motor function in 78% of the observations, and observer B in 77%. Observer A found a higher level at the second reassessment in 11% of the observations and a lower level in 12%, while the corresponding figures for observer B were 11 and 13% respectively.

Observer A found the same degree of deviation from the description in 75% of the observations, increased the degree (no deviation → suspected deviation → clear deviation) at the second reassessment in 18% and decreased it in 7%. The corresponding figures for observer B were 74, 16 and 10% respectively.

At the two reassessments of asymmetry, observer A came to the same result in 88% of the observations and observer B in 85%.

The primary assessment, made by the two main observers in unison, was also compared with each reassessment for each observer. The same level of motor development was noted at all three assessments in 62% of the observations by both observers. Observer A noted the same level in 75% when the primary assessment was compared with the first reassessment, and in 72% when it was compared with the second one. The corresponding figures for observer B were 77 and 72%.

Observer A noted the same degree of deviation from the described level of motor development in 78% when the primary assessment was compared with the first reassessment, and in 72% when it was compared with the second one. The corresponding figures for observer B were 71 and 68%. Observer A increased the degree in 12% of the observations at the first reassessment, compared with the primary assessment, and in 21% at the second, while observer B increased the degree in 22% at the first reassessment and in 25% at the second.

DISCUSSION

Many different factors can affect the results of a test for motor development. Some factors are related to the variations in performance of the infant under study, some depend on the construction of the instrument and some on the ability and alertness of the observer. The ideal instrument should be so well composed that it discriminates accurately between levels of development regardless of who conducts the test. The present protocol was devised with these aims in sight.

In 75 to 80% of the observations the two observers participating in this study arrived at the same level of development when assessing the infants from the video recordings. This must be considered an acceptable degree of conformity, and it is interesting that the conformity was slightly higher at the later reassessment, as experience with the protocol increased. Where they did not agree there did not seem to be any consistent difference in their assessment of the level of motor development.

In infants whose development falls between two described levels, it can sometimes be difficult to decide to which level the development should be allocated. For example the examiner could consider one level of development without deviation, or alternatively a higher level with some degree and type of deviation. As the description of the type of deviation was not included in the study, it is not known how often this occurs. If a further clinical evaluation of the protocol should indicate that a difference of one step in the assessments made by the main observers is acceptable, an interobserver agreement of 96% for level of development and 98% for deviation of motor performance will have been achieved in the first reassessment. The corresponding figures in the second reassessment will be 97 and 95% respectively.

The infants sampled for this test were taken at random from the population of tested preterm and term infants, the only requirement being that the primary assessment was made by the two main observers together. This meant that the material could have included infants who in the longitudinal study could prove to be either normal or abnormal in their motor performance.

In 62% of the observations the two observers arrived at the same level of motor development at all three assessments. In the remaining observations there was no clear trend regarding the changes made from one assessment to another. There was a tendency however, for more deviations from the description of the level to be noted as experience with the protocol increased. One of the observers made the greatest change between the primary assessment and the first reassessment and added only a few more changes at the second reassessment, while the other observer made a smaller change between the primary assessment and the first reassessment, and then a further change of the same size at the second reassessment. It is possible that familiarity with the protocol and the effect of training make the observers more observant regarding deviations from the description of the motor performance.

The agreement between the primary assessment and the assessments made from the video recordings shows that accurate assessments can be made from the recordings. It was felt before hand that assessment from recordings would probably reveal deviations not observed during the primary assessment. To make the assessment from the recordings as realistic as possible, the recordings were played through twice, and the observers were not allowed to interrupt the recording to study a certain part more closely. Even so it is difficult to know whether the increase in the noted number of deviations during the reassessments is an effect of training or an effect of the recording as such. A new study with assessment of recordings taken more recently could resolve this problem, but it would then be difficult to find infants whom the observers could not identify from the recordings.

It can be concluded that for the purpose of our current project the protocol allows an acceptable degree of interobserver agreement and intraobserver consistency when used by the two main observers. Before a wider use of the protocol can be recommended for clinical purposes or research, the results of the ongoing follow-up study must be further examined regarding both reliability and validity.

ACKNOWLEDGEMENTS

This study was supported by the Swedish Medical Research Council (project 19M-7544), the Sven Jerring Foundation, the Folke Bernadotte Foundation, the Linnea and Josef Carlsson Foundation and the Hildur Wallén's Donation to Perinatal Research.

REFERENCES

1. Amiel-Tison, C.: A method for neurologic evaluation within the first year of life: experience with full-term newborn infants with birth injury. *Curr Prob Pediatr* 7: 1-50, 1976.
2. Bayley, B.: Manual for the Bayley scales of infant development. Psychological Corporation, New York, 1969.
3. Beintema, D.J. & Precht, H. F. R.: A neurological study of newborn infants. Heinemann, London, 1968.
4. Frankenburg, W., Dodds, J.B., Fondal, A. W., Kazuk, E. & Cohrs, M.: Denver Developmental Screening Test: Reference Manual. University of Colorado Medical Center, Denver, 1975.
5. Griffiths, R.: The Abilities of Babies. University of London Press, London, 1954.
6. Hack, M., Fanaroff, A. A. & Merkatz, I. R.: The Low-Birth-Weight infant - Evolution of a Changing Outlook. *N Engl J Med* 301: 1162-1165, 1979.
7. Hagberg, B., Hagberg, G. & Zetterström, R.: Decreasing perinatal mortality - Increase in cerebral palsy morbidity? *Acta Paediatr Scand* 78: 664-670, 1989.
8. Kitchen, W. H., Orgill, A., Rickards, A., Ryan, M. M., Lissenden, J. V., Keith, C. G., Yu, V. Y. H., Ford, G. W., Astbury, J., Russo, W., Bajuk, B. & Nave, J. R. M.: Collaborative study of very-low-birthweight infants: outcome of two-year-old survivors. *Lancet* 1157-1160, 1982.
9. McCarthy, D.: Manual for the McCarthy Scales of Children's Abilities. The psychological Corporation, New York, 1972.
10. Sedin, G. & Hammarlund, K.: Neonatal Factors in Mild Mental Retardation. *Ups J Med Sci Suppl* 44: 183-187, 1987.
11. Stewart, A. L., Reynolds, E. O. R. & Lipscomb, A. P.: Outcome for infants of very low birthweight: survey of world literature. *Lancet* 1038-1040, 1981.
12. Stuber, W., Dehne, P., Miedaner, J. & White, P.: The Milani-Comparetti Motor Developmental Screening Test. (Manual revised edition). Meyer Children's Rehabilitation Institute, University of Nebraska Medical Centre, Omaha, Nebraska, 1987.
13. Persson, K. & Strömberg, B.: A protocol for structured observation of motor performance in preterm and term infants. *Ups J Med Sci* 98: this issue, 1993.

Correspondence to:

Gunnar Sedin, M.D., Med.Dr., Professor

Department of Paediatrics, Uppsala University Children's Hospital,

S-751 85 Uppsala, Sweden.