

P R A C E O R Y G I N A L N E
neonatologia

Usefulness of the Apgar score: a national survey of Polish neonatal centers

Ocena przydatności skali Apgar: ogólnopolska ankieta dotycząca oddziałów noworodkowych

Kornacka Maria K.¹, Musialik-Świetlińska Ewa², Świetliński Janusz³, Książyk Janusz⁴, Migdał Marek³, Brożek Grzegorz⁵, Zejda Jan⁵

¹ Neonatal and Intensive Care Department, Warsaw Medical University, Poland

² Neonatal Intensive Care Unit, Medical University of Silesia, Katowice, Poland

³ Anesthesiology and Intensive Care Department, Children's Memorial Health Institute, Warsaw, Poland

⁴ Department of Pediatrics, Children's Memorial Health Institute, Warsaw, Poland

⁵ Department of Epidemiology, Medical University of Silesia, Katowice, Poland

Abstract

Introduction: The paper discusses the reliability of the Apgar score for evaluating newborns, particularly its usefulness in assessing the state of preterm or full term newborns born with hypoxia.

Aim: The paper provides a sum-up of the opinions on usefulness and reliability of the Apgar score given by doctors from 255 NICUs.

Material and methods: Data for the analysis were gathered by means of a questionnaire opinion poll sent to 158 primary referral centers, 71 secondary referral centers and 26 tertiary referral centers. Respondents answered questions about value of the Apgar score assessment (highly valuable, limited value, always reliable), as well as overrating and underrating children born in good overall condition and children born with clinical and biochemical indicators of hypoxia. In the group of prematurely born babies, the data concerning newborns with very low (VLBW) and extremely low birth weight (ELBW) were analyzed separately.

Results: 88,5% neonatologists claimed the use of the Apgar score in assessing newborn condition to be of little value and only 11,5% found this indicator useful and reliable.

Conclusions: According to the majority of Polish neonatologists, Apgar score is not reliable in the assessment of term and preterm hypoxic newborns.

Key words: **Apgar score newborns / neonatal hypoxia / questionnaire / neonatal referral center /**

Corresponding author:

Maria K. Kornacka
Neonatal and Intensive Care Department, Warsaw Medical University
Duchess Anna Mazowiecka Clinical Hospital,
Karowa 2; 00-315 Warszawa, Poland
tel +48-22-59-66-155, fax +48-22-59-66-484, mobile +48-693539492

Otrzymano: 01.10.2010
Zaakceptowano do druku: 20.12.2010

Streszczenie

Ocena noworodków według skali Apgar wydaje się być coraz mniej przydatna w ocenie noworodka niedotlenionego.

Cel pracy: Celem pracy była sumaryczna opinia neonatologów zatrudnionych w 255 oddziałach noworodkowych dotycząca przydatności i wiarygodności skali Apgar w ocenie stanu ogólnego noworodków.

Materiał i metody: Badania oparto na ankietach przesłanych do 158 oddziałów I stopnia referencyjności, 71 oddziałów II stopnia referencyjności i 26 oddziałów III stopnia referencyjności. Neonatolodzy przedstawili swoje opinie dotyczące wiarygodności skali Apgar w ocenie donoszonych i przedwczesnie urodzonych noworodków z objawami niedotlenienia wewnątrzmacicznego. Respondenci udzielali odpowiedzi oceniając wartość skali Apgar jako bardzo wiarygodną, o ograniczonej wiarygodności lub zawsze miarodajną, a także jako niedoszacowaną lub przeszacowaną w stosunku do biochemicznych i klinicznych wykładników niedotlenienia. W grupie noworodków przedwczesnie urodzonych opinie dotyczyły zarówno oceny noworodków z bardzo małą jak i ekstremalnie małą masą ciała.

Wyniki: 88,5% neonatologów podkreśla ograniczoną wartość skali Apgar w ocenie stanu ogólnego noworodków, a tylko 11,5% uważa ją stale za przydatną i wiarygodną.

Wnioski: Według większości polskich neonatologów ocena według skali Apgar noworodków niedotlenionych tak donoszonych jak i przedwczesnie urodzonych nie jest wiarygodna.

Słowa kluczowe: **skala Apgar / noworodek / niedotlenienie / ankieta / oddziały neonatologiczne /**

Introduction

For many years newborns in the first minutes of life have been assessed according to the score which was created in 1952 by Virginia Apgar. The score, based on clinical evaluation of respiratory, circulatory and nervous systems in the first minute of a newborn life, reflects the fetus condition and the influence of factors present during labor. It is generally believed that it correlates well with in-utero indicators of fetal well-being [1, 2].

The assessment of newborns according to the Apgar scale in the first minute of their life is an indicator of the in-utero condition of the fetus and different factors present during labor. Low 5-minute Apgar score is said to be directly related to further development of the child [3, 4].

However, more and more often physicians claim little usefulness of the Apgar score in evaluating the condition of preterm babies, as well as considerable difficulties in its interpretation. Questioning the credibility of the Apgar score, medical professionals point out lack of correlation between the score and the neonatal period of life, the clinical state and morbidity of neonates, as well as the further development of in-utero hypoxic babies. Skeptical opinions concern also high subjectivity in awarding points to neonates and differences in scoring, stemming from discrepancies in the experience of doctors present during labor. Although the Apgar score ceased to be used in some departments of neonatology altogether, e.g. in France, it is still widely used and none of the numerous other scales created to assess the clinical state of a neonate, such as CRIB and SNAP, managed to replace it [5].

The aim of the present work was to determine the opinions of neonatal centers in Poland on the Apgar score as part of a national survey on clinical practices. The authors were particularly interested in which populations, with regard to maturity and hypoxemia, the Apgar score was perceived as reliable and whether the perceptions differed among the tertiary, intermediate and primary care centers.

Material and Methods

The study was conducted under the auspices of the Polish Neonatal Society, as a part of a development of Neonatal Clinical Procedure Recommendations Program.

A detailed questionnaire was sent to all 420 neonatal units in Poland. While the surveys were anonymous, coded envelopes permitted the main data center to make follow-up calls to non-responders at 4 and 6 weeks. The questionnaires were directed to the chief of each unit, and were to be completed by all members of neonatal staff. Questions concerning the Apgar score constituted a part of a much larger survey and were included in the section on perinatal asphyxia. There were 6 questions about the Apgar score (4 single choice and 2 multiple choice questions) and they addressed opinions about the general usefulness of the Apgar score in different neonatal populations (full term, preterm, in good overall condition and hypoxemic, preterm and lower birth weight) and the tendency to over- or underestimate it in cases of perinatal asphyxia. The responses were also evaluated with regard to the three levels of neonatal centers. The data were analyzed statistically using the Chi² test. The level of significance was set at $p < 0.05$.

Results

274 questionnaires were returned but 19 questionnaires were excluded due for not having been completed properly. Therefore, responses from 255 NICUs were evaluated, representing 60.7% of all neonatal care centers in Poland. The response was representative of all levels of care: 158 from primary (Level I), 71 from intermediate (Level II), and 26 from tertiary care units (Level III). Due to the fact that some centers did not answer every question, not all of the totals equal 255.

Only a few (29, 11.5%) centers found the Apgar assessment of every newborn reliable. The opinion was slightly, but not statistically significantly different among neonatologists in various level centers [Level I (13.5%), Level II (11.1%), and Level III (0.0%), $p = 0.1$]. (Table I).

In contrast, nearly half (42%) of the centers found the Apgar results in full term infants reliable. (Table II).

Usefulness of the Apgar score: a national survey of Polish neonatal centers.

Table I. Reliability of the Apgar score newborns assessment from the point of view of NU doctors in various referral centers.

In your opinion, is Apgar score assessment always reliable?	NEONATAL UNITS				P value
	ALL N=253 (100%)	I N=155 (100%)	II N=72 (100%)	III N=26 (100%)	
yes	29 (11.5)	21 (13.5)	8 (11.1)	0 (0.0)	0,1
no	224 (88.5)	134 (86.4)	64 (88.9)	26 (100)	

Table II. Opinion about Apgar score value in assessing the degree of post-natal hypoxia in full term babies expressed by NU doctors in various referral centers.

	NEONATAL UNITS				P value
	ALL N=255 (100%)	I N=158 (100%)	II N=71 (100%)	III N=26 (100%)	
Very valuable	107 (42.0)	76 (48.1)	25 (35.2)	6 (23.1)	0,02
Limited value	148 (58.0)	82 (51.9)	46 (64.8)	20 (76.9)	

Table III. Opinion about the Apgar score value in assessing the degree of post-natal hypoxia in premature babies expressed by NU doctors in various referral centers.

	NEONATAL UNITS				P value
	ALL N=254 (100%)	I N=156 (100%)	II N=72 (100%)	III N=26 (100%)	
Valuable in assessing the state of every prematurely delivered newborn	16 (6.3)	11 (7.0)	5 (6.9)	0 (0.0)	0,06
Valuable for prematurely delivered neonates except for VLBW	3 (1.2)	2 (1.3)	0 (0.0)	1 (3.8)	
Valuable for premature neonates except for ELBW	3 (1.2)	0 (0,0)	3 (4,2)	0 (0.0)	
Limited value for assessing the state of prematurely born children	232 (91.3)	143 (91.7)	64 (88.9)	25 (96.1)	

Table IV. Various referentiality of NU doctors opinions concerning the Apgar score for neonates .

	NEONATAL UNITS				P value
	ALL N=224 (100%)	I N=134 (100%)	II N=64 (100%)	III N=25 (100%)	
It is reliable for babies delivered in good clinical condition	212 (93.8)	125 (92.6)	62 (95.4)	25 (96.1)	0.6
It is more often overstated in children with perinatal hypoxia	175 (77.4)	106 (78.5)	48 (73.8)	21 (80.8)	0.6
It is more often lowered in children with perinatal hypoxia	49 (21.7)	29 (21.5)	15 (23.1)	5 (19.2)	0.9

Table V. Opinions on the Apgar score provided by NU doctors claiming that the score is not always reliable (combined answers).

	NU N=224 (100%)	I N=134 (100%)	II N=64 (100%)	III N=25 (100%)	P value
It is reliable in case of babies delivered in good clinical condition	150 (66.4)	88 (65.6)	43 (67.2)	19 (76)	0,8
It is more often overstated in children with perinatal hypoxia					
It is reliable for babies delivered in good clinical condition	31 (13.4)	18 (13.4)	10 (15.6)	3 (12)	
It is more often lowered in children with perinatal hypoxia					
Not useful for healthy newborns	43 (20.2)	28 (21.0)	11 (17.2)	3 (12)	

There was a statistically significant difference among the three center levels ($p=0.02$), with less confidence in higher level centers [Level I – 48.1%, Level II – 35.2%, Level III – 23.1%].

Only some (8.7% – [6,3%+1,2%+1,2%]) of the centers, regardless of the level, found the Apgar score reliable for assessing preterm infants. A few more (1,2%) found Apgar scores useful in evaluating preterm infants who were larger than 1500 grams. (Table III).

In contrast, 93.8% of the centers found the Apgar scale reliable for assessing infants born without hypoxemia. Most (77.4%) believed the Apgar score was inappropriately high, but still a significant number answered it was inappropriately low (21.7%) in case of neonates with hypoxemia. There was no difference associated with the level of the center with regard to any of these features. (Table IV).

Finally, the authors investigated separately the 224 centers that found the Apgar score not useful for all newborns. The opinions were cross tabulated with regards to usefulness in newborns born in good overall condition and typical bias in scores in infants with perinatal hypoxemia. The common opinion (66.4%) was that the Apgar score was useful in newborns born in good overall condition but was overestimated in infants with perinatal hypoxemia. 13.4% of the centers found the score useful in newborns born in good overall condition but underestimated in infants with perinatal hypoxemia and 20.2% believed it to be useless for newborns born in good overall condition. (Table V).

Discussion

Apgar score has played a crucial role in the delivery room but it is not without limitations [6, 7]. The present study determined the attitudes of neonatologists in Poland towards the reliability of the Apgar score. There was a general agreement that the score was reliable in infants born without hypoxemia. However, in infants with hypoxemia there was no strong consensus about a particular direction of bias in the score. Three-quarters of the respondents felt the Apgar score was inappropriately high, but nearly a quarter felt it was inappropriately low. Previously, the 3- or under 5-minute Apgar score was considered vital for the diagnosis of perinatal asphyxia [1]. Polish neonatologists did not find it to be a specific indicator of hypoxia. Also, most neonatologists felt the Apgar

score was unreliable in preterm infants, regardless of their weight, and only about half of them found the Apgar score reliable in full term infants. These results may be reflecting visible lack of Apgar reliability in hypoxemic infants and the likelihood of hypoxemia in these two groups. It is also supported by the fact that level III centers found the score less reliable in full term infants, and they would be more likely to see hypoxemic full term infants. There are numerous data indicating poor correlation between the 5-minute Apgar score in full term infants and future neurologic outcomes. However, data confirming that Apgar is less reliable in infants with hypoxia are scarce [1]. Connolly did not observe statistically significant effect on the score in term infants born to mothers with intra-amniotic infection and meconium stained fluid [8]. Leuthner and Das noted that fetal heart rate patterns, prolonged labor, meconium-stained fluid, a low 1-minute Apgar score, and mild to moderate acidemia have no predictive value for long-term neurological injuries [9]. In 12 neonatal units in Germany, Austria and Switzerland 44 residents and 52 specialists determined the Apgar score for the same babies. The discrepancy reached five points and was especially remarkable in the group of specialists [10]. Similar discrepancies were reported by O'Donnell CP and coworkers [11]. These reports are consistent with the general opinion of lack of Apgar reliability reported in the present study. The authors found that neonatologists feel the Apgar score is unreliable in preterm infants and attributed it primarily to their likelihood of being hypoxemic. Also, it is common knowledge that healthy premature babies may obtain very low score solely because they are not fully developed. The main difficulty is connected with the subjective assessment of the skin color, muscle tension and reaction to stimuli [12] as those elements depend largely on the maturity of a newborn [1, 3]. Therefore, in the evaluation of a full term as well as premature babies, a more complex assessment of their clinical and biochemical state is vital [13, 14]. In spite of the issues of reliability, recent outcome research has confirmed the accuracy of the Apgar score. Recently published Swedish and English research on the relationship between IQ of full term born people (not suffering from hypoxic-ischemic encephalopathy) at the age of 18 and their post-natal clinical condition in the first minutes of life, claims a correlation between the low short-lasting (till 5

Usefulness of the Apgar score: a national survey of Polish neonatal centers

minutes of life) and longer-lasting (longer than 5 minutes of life) score in Apgar scale and disorders in their further development. A connection between the low short-lasting (till 5 minutes of life) and longer-lasting (longer than 5 minutes of life) score in the Apgar scale and the disorders in their further development [15] has been found. Thorngren-Jerneck and Herbst also found that high Apgar scores were associated with the likelihood of high morbidity and further developmental disorders [16]. It seems clear that the Apgar score, regardless of its limitations, accurately predicts risks in the general population. Opinions of Polish neonatal centers reflect the reality: the Apgar score, though useful in indicating risk in a large group, is not clinically useful for management of a particular patient, especially at the NICU.

The question remains whether CRIB, SNAP, or perhaps some other approaches, would be more useful and reliable. The abovementioned scores estimate the complex of pre-, peri- and postnatal factors during a prolonged period of time (a few or 12 hours after the delivery) while the Apgar score concerns only the first few minutes after the delivery and estimates the influence of factors which play the role at birth and just after the delivery. Numerous doubts and inquiries remain. First of all, whether there exists at least one score which can play the role of such a fast, although subjective, score as the one created by Virginia Apgar or if it is better to assess the state of full term babies born in good overall condition according to the "good old" Apgar score. Also, whether children who were delivered prematurely in bad state should be evaluated by means of a different score that has yet to be designed and what should happen to the low and very low scores given mainly to these babies. Finally, whether a new improved Apgar score should also reflect a variety of resuscitating procedures applied to the baby, thus supplementing the old traditional score or replace it altogether.

Conclusions

1. The Apgar score remains to be a good tool for assessing full term newborns born in good overall condition.
2. The Apgar score is useless in assessing the state of very low body weight and extremely low body weight babies.
3. The Apgar score is not reliable in the assessment of the degree of neonates hypoxia and in these cases it is usually overstated by doctors;
4. The percentage of opinions expressing the limited reliability of Apgar score rises together with the referentiality of neonatal centers.

References

1. American Academy of Pediatrics. The Apgar Score. *Pediatrics*. 2006, 117, 1444-1447.
2. Drage J, Kennedy C, Schwarz B. The Apgar score as an index of neonatal mortality. A report from collaborative study of cerebral palsy. *Obstet Gynecol*. 1964, 24, 222-230.
3. Goldenberg R, Huddleston J, Nelson K. Apgar score and umbilical arterial pH in preterm newborn infants. *Am J Obstet Gynecol*. 1984, 149, 651-654.
4. Nelson K, Ellenberg J. Apgar Scores as Predictors of Chronic Neurologic Disability *Pediatrics*. 1981, 68, 36-44.
5. Casey B, McIntire D, Leveno K. The continuing value of the Apgar Score for the assessment of Newborn Infants. *NEJM*. 2001, 344, 467-471.

6. Lopriore E, van Burk G, Walther F, [et al.]. Correct use of the Apgar score for resuscitated and intubated newborn babies: questionnaire study. *BMJ*. 2004, 329, 143-144.
7. Berner-Trąbska M, Kowalska-Koprek U, Karowicz-Bilińska A, [et al.]. Wpływ nadwagi i otyłości u ciężarnych na przebieg ciąży i okresu okołoporodowego z uwzględnieniem stanu urodzeniowego noworodka w oparciu o materiał kliniczny. *Ginekol Pol*. 2009, 80, 845-850.
8. Connolly T. Clinical intra-amniotic infection with meconium-stained at term: an analysis of Apgar score. *JAOA*. 2000, 100, 492-496.
9. Leuthner S, Das U. Low Apgar scores and the definition of birth asphyxia. *Pediatr Clin North Am*. 2004, 51, 737-745.
10. Rudiger M, Wauer R, Schmidt R, [et al.]. The Apgar Score. *Pediatrics*. 2006, 118, 1314-1315.
11. O'Donnell C, Kamlin C, Davis P, [et al.]. Clinical assessment of infant color at delivery. *Arch Dis Child Fetal Neonatal Ed*. 2007, 92, F465-467.
12. Martin G, Stark A. The Apgar Score. In reply: *Pediatrics*. 2006, 118, 1315-1316.
13. Stark C, Gibbs R, Freedman W. Comparison of umbilical artery pH and 5-minute Apgar score in the low-birth-weight and very low-birth-weight infants. *Am J Obstet Gynecol*. 1990, 163, 818-823.
14. Silverman F, Suidan J, Wasserman J, [et al.]. The Apgar Score: It is Enough? *Obstet Gynecol*. 1985, 66, 331-336.
15. Odd D, Rasmussen F, Gunnell D, [et al.]. A cohort study of low Apgar scores and cognitive outcomes. *Arch Dis Child Fetal Neonatal Ed*. 2008, 93, F115-F120.
16. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar Score: A population-based register study of 1 million term birth. *Obstet Gynecol*. 2001, 98, 65-70.