Obstetrical and Pediatric Anesthesia

Best evidence in anesthetic practice

Prognosis: the Apgar score predicts 28-day neonatal mortality

Article appraised

Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med* 2001; 344: 467–71.

Structured abstract

Question: In infants born before term (26–36 weeks of gestation) and infants born at term (≥ 37 weeks of gestation), does the Apgar score, at five minutes of age, predict 28-day neonatal death?

Design: Prospective cohort.

Setting: Academic hospital with a level III neonatal intensive care unit.

Patients: 151,891 live-born singleton infants, with gestational age ≥ 26 weeks, who were delivered between January 1988 and December 1998. Infants with missing umbilical blood gas results ($n=6264$) were excluded.

Assessment of prognostic factors: The cohort was divided by general anesthesia and five-minute Apgar score. Umbilical blood pH were also recorded. Gestational age was based on the estimate used for clinical care. Apgar score was measured by nursing staff or third-year pediatric house staff. Obstetrical data sheets and hospital records were used to extract neonatal data.

Main outcomes: Neonatal death during the first 28 days after birth.

Main results: Preterm infants had higher neonatal death rates than term infants (Table I). Apgar scores of 0–3 and 4–6 are associated with increased risk of death compared to Apgar scores of 7–10 (Table I). Severity of metabolic acidemia did not appreciably modify the relative risk of death associated with Apgar scores ≤ 3 ; however, infants with Apgar scores ≤ 3 and umbilical blood pH ≤ 7.0 were at significant risk for death (preterm, relative risk (RR) 102, 95% confidence interval (CI) 65–160; term, RR 3204, 95% CI 1864–5508).

Conclusion: 28-day neonatal mortality was associated with lower five-minute Apgar scores in infants born at ≥ 26 weeks of gestation.

Funding: Not described.

Correspondence: Dr. Brian M. Casey, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA. Email: brian.casey@utsouthwestern.edu

TABLE Risk of 28-day mortality in singleton infants born at 26 weeks of gestation in relation to Apgar scores at five minutes of age

Five-minute Apgar scores	n	Preterm (26 to 36 weeks of gestation)			n	Term (≥ 37 weeks of gestation)		
		Deaths (rate/1000 births) ratio†	Relative risk (95% CI)* ratio†	Positive likelihood		Deaths (rate/1000 births)	Relative risk (95% CI)*	Positive likelihood
0 to 3	92	29 (315)	59 (40–87)	0.2	86	1 (244)	1460 (835–2555)	0.4
4 to 6	56	40 (72)	13 (9–20)	0.3	561	5 (9)	53 (20–140)	0.1
7 to 10	12,751	68 (5)	1	11.3	13,158	22 (0.2)	1	91.6

*Relative risks for 28-day mortality are calculated relative to infants with five-minute Apgar scores between 7 and 10.

†The positive likelihood ratio (calculated by sensitivity / [1-specificity]) reflects the likelihood that a neonate with a five-minute Apgar score within the specific range will live. A value greater than 10 indicates a high likelihood of survival. For sensitivity and specificity, see Table in the first commentary.

Commentary by A. Macarthur

Casey *et al.*'s review of neonatal death rates in their ten-year (1988 to 1998) birth cohort should be commended. The data provided information on survival of singleton infants born between 26 to 43 weeks gestation and assessed the relationship between the Apgar score and umbilical artery blood pH.

The cohort design attempted to evaluate the relationship between Apgar score or umbilical artery gas pH and neonatal death (during the first 28 days after birth). Casey *et al.* calculated the relative risk (RR), a statistic commonly used to describe the relationship between a suspected risk factor and disease. In this study, the "risk" factor was low Apgar score (0–3 or 4–6); the disease was neonatal death. The RR is the quotient obtained from dividing the proportion of neonatal death in each low Apgar score group by the proportion of neonatal death in the control group (those with Apgar score 7 to 10). RR of one indicates no difference in risk between the two groups; RR less than or greater than one indicates increased risk for one of the groups.

As hoped, the Apgar scale, which was developed to measure neonatal well being, indicated an increasing risk for neonatal death with decreasing Apgar score. However, these results existed only at the five-minute assessment. The RR for neonatal death was greatest for infants whose five-minute Apgar score was less than 4 compared to those with Apgar scores greater than 6 (preterm infants 59 times more likely to die; term infants 1,460 times more likely to die). Important in understanding the precision of RR is its 95% confidence interval (CI). The CI represents the range with 95% probability of containing the true value of the RR. None of the 95% CI for RR of death of preterm or term infants included one.

TABLE Predictive qualities of Apgar score for neonatal death

Predictive measure	Preterm neonates	Term neonates
<i>Sensitivity</i>		
Apgar score 0 – 3	21.2%	43.8%
Apgar score 4 – 6	29.2%	10.4%
Apgar score 7 – 10	49.6%	45.8%
<i>Specificity</i>		
Apgar score 0 – 3	0.5%	0.05%
Apgar score 4 – 6	3.9%	0.4%
Apgar score 7 – 10	95.6%	99.5%
<i>Positive predictive value</i>		
Apgar score 0 – 3	31.5%	24.4%
Apgar score 4 – 6	7.2%	0.9%
Apgar score 7 – 10	0.5%	0.02%
<i>Negative predictive value</i>		
Apgar score 0 – 3	68.5%	75.6%
Apgar score 4 – 6	92.8%	99.1%
Apgar score 7 – 10	99.5%	100%

The collection of birth data required formal commitment by nursing, obstetrical and pediatric services. Description of the identification of neonatal death was limited and I was not assured that all neonatal deaths within the cohort were captured. Missing data would have an effect on the estimated risk if large numbers of deaths, especially those within the first 28 days, were missed.

A greater concern with the study was its use of the measure, relative risk, to evaluate the predictive abilities of the Apgar scale or umbilical pH. Usual measurements to assess prediction include: sensitivity (the number of deaths with the Apgar scores of interest, divided by all deaths); specificity (the number of survivors without the Apgar scores of interest, divided by all survivors); positive predictive value (the proportion of all neonates, with the Apgar scores of interest, who died); and the negative predictive value (the proportion of all neonates, with the Apgar scores of interest, who did survive; Table). The only useful predictive result from the five-minute Apgar score is for scores between 7 and 10. With specificity and negative predictive values greater than 90%, most acute care providers and parents can be reassured that a child with an Apgar score in this range is unlikely to die. For low Apgar scores, the prediction of death is very poor. Rather than prediction, I believe the Apgar score should be evaluated for its utility in directing acute resuscitative efforts in the beginning minutes of a neonate's life.

Alison Macarthur MD FRCPC MSc
Toronto, Ontario

Reference

- 1 Last JM. A Dictionary of Epidemiology, 2nd ed. Toronto: Oxford University Press, 1988.

Commentary by S. Halpern

The Apgar score was developed by Dr. Virginia Apgar in 1953¹ for the purpose of providing a "simple, clear classification or 'grading' of newborn infants which can be used as a basis for discussion and comparison of the results of obstetric practices, types of maternal pain relief and the effects of resuscitation." Since that time, this scoring system has been adopted by delivery room staff all over the world.

Of interest, the Apgar score was developed in an era when obstetrical anesthesia was practiced very differently than it is today. Cyclopropane and oxygen, sometimes given with d-tubocurarine, was the most common general anesthetic technique. The anesthetic was administered by mask to the spontaneously

breathing parturient. The potential benefit of uterine displacement had not yet been discovered and hypotension, particularly after spinal anesthesia, was extremely common. The scientific basis of newborn resuscitation had not yet been described. Intensive care for the premature infant did not exist.

In this milieu, Dr. Apgar devised her scale. She selected the items using face validity and convenience of administration as the most important considerations. She assessed the content validity by noting that all the important physiologic systems (cardiovascular, respiratory, musculoskeletal, and central nervous system) were represented and assessed. She described the interpretation of each item in terms that allowed the reader to use the scale. She tested the scale for inter-rater reliability. She then tested the construct validity retrospectively on 1,021 liveborn infants born in her hospital over a 7.5-month period using neonatal mortality as the primary endpoint. In addition she correlated the score with type of maternal anesthesia and mode of delivery. In other words, her method of scale design was similar to the usual approach seen today.²

Five years later, Apgar *et al.* published a larger report on their experience with the scale in 15,348 infants.³ They found that the death rate in infants with an Apgar score of <3 (at one minute) was 15%, but only 0.13% if the score was 10. In the premature infant (500 to 2,500 g), they found that the neonatal death rate was 44% and 4.8% respectively. Further, they found relatively poor correlation between umbilical vein acid-base biochemistry and the Apgar score, although there was better correlation between the score and umbilical artery or aortic values.

Casey *et al.* have confirmed that Dr. Apgar's work still has application today. The anesthetic techniques and care of the newborn have changed dramatically in the intervening time. We have abandoned the use of the one-minute Apgar score (although it is still recorded in most institutions) in favour of the five-minute score. Casey *et al.* correctly point out that the Apgar score was not intended to predict neurological outcomes such as cerebral palsy or learning disabilities and should not be used for this purpose. However, the Apgar score was carefully crafted and validated at the time of development. It remains the gold standard for prediction of neonatal mortality and requirement for resuscitation.

Stephen Halpern MD FRCPC MSc
Toronto, Ontario

References

- 1 Apgar V A proposal for a new method of evaluation of the newborn infant. *Anesth Analg* 1953; 32: 260–7.
- 2 Streiner DL, Norman GR Health Measurement Scales: A Practical Guide to Their Development and Use, 2nd ed. Oxford: Oxford University Press, 1995.
- 3 Apgar V, Holaday DA, James LS, Weisbrot IM, Berrien C. Evaluation of the newborn infant – second report. *JAMA* 1958; 168: 1985–8.