




Can the Apgar Score be Used for International Comparisons of Newborn Health?

Ayesha Siddiqui,^a  Marina Cuttini,^b Rachel Wood,^c Petr Velebil,^d Marie Delnord,^a Irisa Zile,^e Henrique Barros,^f Mika Gissler,^g Ashna D Hindori-Mohango,^{h,i} Béatrice Blondel,^a Jennifer Zeitlin^a for the Euro-Peristat Scientific Committee¹

^aINSERM UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Center for Epidemiology and Biostatistics, Paris-Descartes University, Paris, France

^bClinical Care and Management Innovation Research Area, Bambino Gesù Children's Hospital, IRCCS, Roma, Italy

^cInformation Services Division, NHS National Services Scotland, Edinburgh, Scotland, UK

^dInstitute for the Care of Mother and Child, Prague, Czech Republic

^eCentre for Disease Prevention and Control of Latvia, Riga, Latvia

^fISPUP-EPI Unit, Universidade do Porto, Porto, Portugal

^gTHL National Institute for Health and Welfare, Helsinki, Finland

^hDepartment Child Health, Netherlands Organisation for Applied Scientific Research, TNO Healthy Living, Leiden, the Netherlands

ⁱFaculty of Medical Sciences, Department Public Health, Anton de Kom University of Suriname, Paramaribo, Suriname

Abstract

Background: The Apgar score has been shown to be predictive of neonatal mortality in clinical and population studies, but has not been used for international comparisons. We examined population-level distributions in Apgar scores and associations with neonatal mortality in Europe.

Methods: Aggregate data on the 5 minute Apgar score for live births and neonatal mortality rates from countries participating in the Euro-Peristat project in 2004 and 2010 were analysed. Country level associations between the Apgar score and neonatal mortality were assessed using the Spearman rank correlation coefficient.

Results: Twenty-three countries or regions provided data on Apgar at 5 minutes, covering 2 183 472 live births. Scores <7 ranged from 0.3% to 2.4% across countries in 2004 and 2010 and were correlated over time ($\rho = 0.88$, $P < 0.01$). There were large differences in healthy baby scores: scores of 10 ranged from 8.8% to 92.7% whereas scores of 9 or 10 ranged from 72.9% to 96.8%. Countries more likely to score 10 s, as opposed to 9 s, for healthy babies had lower proportions of Apgar <7 ($\rho = -0.43$, $P = 0.04$). Neonatal mortality rates were weakly correlated with Apgar score <7 ($\rho = -0.06$, $P = 0.61$), but differences over time in these two indicators were correlated ($\rho = 0.56$, $P = 0.02$).

Conclusions: Large variations in the distribution of Apgar scores likely due to national scoring practices make the Apgar score an unsuitable indicator for benchmarking newborn health across countries. However, country-level trends over time in the Apgar score may reflect real changes and merit further investigation.

Keywords: Apgar Score, neonatal morbidity, neonatal mortality, health indicators.

Since its introduction in 1952, the Apgar score has been used worldwide for the rapid and standardised assessment of neonates after delivery to determine the need for prompt resuscitative intervention.^{1,2} The Apgar score consists of five clinical signs which are each given a score of 0, 1, or 2 (Table S1). The

overall Apgar score is the sum of these components and is defined as reassuring for a score of 7–10, moderately abnormal for a score of 4–6, and low for a score of 0–3.³ The score is reported at 1 and 5 minutes after birth for all neonates, and subsequently at five-minute intervals until 20 minutes after birth for neonates with a non-reassuring score. Over half a century since the Apgar score became integrated in routine clinical practice, it remains a standardised, effective, and convenient tool for neonatal assessment.^{4–6}

Although the Apgar score was never intended for the prediction of mortality or long-term disability, its

Correspondence:

Ayesha Siddiqui, Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Center for Epidemiology and Statistics Sorbonne Paris Cité, INSERM UMR 1153, 75014 Paris, France.
E-mail: ayasha.siddiqui@inserm.fr

¹The Euro-Peristat Scientific Committee authors are listed in Appendix section.

utility beyond triage for immediate resuscitation is of significant interest. A low Apgar score at 1 minute can often be the result of a transient depression in the component parameters and has not been shown to be associated with clinical outcomes.⁷ A low or moderately abnormal 5 minute Apgar score is, however, of greater predictive value, as evidenced by its correlation with neonatal mortality in large population and clinical studies.^{5–10} A non-reassuring 5 minute Apgar score has also been shown to be predictive of infant mortality.^{8,10,11} Furthermore, large national registry studies with long-term follow up indicate an association of a 5 minute Apgar score less than 7 with neurological and developmental disabilities later in life, such as cerebral palsy,^{9,10} epilepsy,¹⁰ mental retardation,^{10,12} and lower academic achievement in secondary school.¹³ Created for the assessment of term neonates, the physiological applicability of Apgar scores in preterm births is questioned since immaturity itself results in lower scores in the absence of neonatal distress.¹⁴ However, population studies suggest that low Apgar scores are equally prognostic of mortality in preterm neonates.^{8,11,15}

The aforementioned studies suggest that, despite significant advances in perinatal care and neonatology since the Apgar score was introduced, it remains not only a relevant tool in the clinical management of neonates, but is also associated with mortality and morbidity at the population level. An analysis of population level trends in Apgar scores may therefore potentially serve as an internationally comparative indicator reflecting the burden of mortality and morbidity. While neonatal mortality is a key indicator of newborn health, its low incidence – under 2 per 1000 live births in many countries in Europe and under 3 per 1000 live births in the majority of European countries¹⁶ – may reduce its usefulness for monitoring outcomes, in particular for sub-populations with a fewer number of births. The Apgar score is routinely collected by many European national health systems and could add to the tool kit for evaluating obstetrical and early neonatal quality of care. To our knowledge estimates of national aggregate Apgar scores have not previously been tested towards such an application.

We examined population level rates and trends in Apgar scores to determine if they relate to neonatal mortality in European cross-country comparisons. Evidence suggests that there is significant inter-observer variability in the 5 minute Apgar score,^{17–20} possibly due to its subjective components. Therefore,

we also tested for differences between countries in the distribution of extreme and intermediate Apgar scores, to determine if norms vary nationally.

Methods

Data for this study were collected by the Euro-Peristat project which is an EU-funded network of clinicians, statisticians, and researchers from EU Member States, Norway, Iceland, and Switzerland. Scientific Committee members in each participating country are responsible for the data collection of 10 core and 20 recommended indicators essential for the surveillance of perinatal health. We used data collected on births in 2004 and 2010, published as part of the European Perinatal Health Reports.^{21,22} Full details of the Euro-Peristat indicators are available on the Euro-Peristat website.^{23,24}

The Euro-Peristat indicators are compiled from aggregated data provided from routine data sources.^{23,24} In most countries, perinatal health data were held in vital statistics, civil registers, and medical birth registers but data also came from nationally representative surveys of births in Cyprus and in France. Data were required to be population-based and should cover the whole country, but if countries could not provide information at the national-level, regional data were also accepted. For example, data for Belgium were provided separately from the regions of Brussels, Wallonia and Flanders, and data from the United Kingdom were provided separately by the UK's constituent countries: England and Wales combined, Northern Ireland, and Scotland. Finally, some countries could not provide data for the years requested, and data from the most recent year available were accepted. Data from Cyprus were from 2007 instead of 2010, France provided data from 2003 instead of 2004, and England and Wales provided neonatal mortality data from 2005 instead of 2004.

Data were collected on all births and deaths at or after 22 weeks; the 22-week gestational age threshold is recommended by Euro-Peristat for the collection of all data on births in Europe.²⁴ Gestational age was defined as the best obstetric estimate. When gestational age was missing, we asked countries to include births if birth weight was 500 grams or more.

We used data from two indicators: the distribution of the 5 minute Apgar score and the neonatal mortality rate. We calculated percentages of live births with an Apgar score <4, 4–6 7–10. The neonatal mortality

rate is defined as the number of neonatal deaths (day 0 through 27) after live birth and expressed per 1000 live births. Indicators from 2004 and 2010 were used.

For this ecological study, the unit of inference was the country. We first described the distributions of Apgar at 5 minutes across countries and examined the association of low Apgar scores across the two time periods to assess consistency and changes over time. We then correlated the proportions of low Apgar scores (<7), and very high Apgar scores of 9 or 10 across countries in order to assess whether coding practices for healthy infants affected the classification of high risk infants. To measure differences in coding practices we computed the proportion of neonates with a score of 10 among those with scores of 9 or 10 to measure the tendency to score more highly, independent of the Apgar score distribution overall. To investigate the association with neonatal mortality, we correlated low scores with neonatal mortality in 2004 and 2010, and studied temporal trends by assessing correlations between rate differences in the Apgar score and neonatal mortality between 2004 and 2010.

Spearman's non-parametric correlation was used for all analyses. While the focus was on low Apgar scores <7, we also assessed other clinically meaningful cut-offs: <4, representing the most critically abnormal scores and <9, representing anything besides completely healthy babies. To take into consideration differences in country population size, we integrated data on annual volume of births into graphical presentations to illustrate how countries with small numbers of annual births might affect the observed associations. However, we did not weight for country size as the country was our observation. Weighting by population size would effectively change the inference away from the ecological towards a study of the total population of the included countries. We also carried out sensitivity analyses by removing countries with low population sizes.

There were less than 3% missing data on the Apgar score in participating countries in 2004 and 2010. Exceptions included Finland (missing 19% in 2004 and 15% in 2010) and Wales (missing 22% in 2004 and 8% in 2010). For Finland, most children missing data at 5 minute had Apgar scores at 1 minute, since 5 minute scores may remain unrecorded if the 1 minute scores are high. For these births, we assigned the mean value at 5 minute corresponding to 1 minute Apgar scores as observed in the sample with non-missing values. Other missing data were

excluded from the analyses. All analyses were performed using Stata, V.14.0 SE (Stata Corporation, College Station, Texas, USA).

Results

Twenty-three countries or regions covering a total of 2 192 632 live births provided data on Apgar scores in 2010, of which 6550 or 0.3% had scores under 4 with a range from 0.1% to 0.7% and a further 17 468 or 0.8% between 4 and 6, with a range from 0.2% to 1.7% (Table 1). The percent of all Apgar scores <7 across countries ranged between 0.3–2.4%. Overall 93.2% of reported live births received Apgar scores of 9 or 10 with a range from 72.9% to 96.8%. Within this group there were large variations in the scoring of 9s vs. 10s. For example, in France 92.7% of the scores were 10 and only 3.9% were 9, whereas in Scotland 18.7% of scores were 10 and 76.3% were 9. Certain countries were outliers across the range of scores, most notably Estonia, Iceland, and Latvia. Apgar score distributions for 2004 are presented in Table S2.

Apgar scores <7 in 2010 were highly correlated with scores <7 in 2004, as shown in Figure 1 ($\rho = 0.88$, $P < 0.01$). Scores of 10 ranged from 4% to 88% of total scores 9 and 10. Countries that were more likely to assign 10 had a lower prevalence of Apgar scores <7 in 2010 (Spearman's $\rho = -0.43$, $P = 0.04$, 23 countries). A negative correlation was observed in 2004, but it was not statistically significant (Spearman's $\rho = -0.38$, $P = 0.10$, 20 countries).

Neonatal mortality rates across countries were not associated with the proportion of abnormal Apgar scores <7 for both study years investigated (Figure 2) nor were they correlated with scores <4 (2010: $\rho = -0.05$, $P = 0.83$, 22 countries; 2004: $\rho = -0.16$, $P = 0.52$, 19 countries) or scores <9 (2010: $\rho = -0.02$, $P = 0.94$, 22 countries; 2004: $\rho = 0.20$, $P = 0.41$, 19 countries). These correlations were not affected by sensitivity analyses which involved removing the countries with smallest sample sizes (Iceland, Malta, and Luxembourg). Data on the number of neonatal deaths and mortality rates in 2004 and 2010 are presented in Table S3.

In countries with data on Apgar scores and neonatal mortality in both 2004 and 2010, the average decrease in neonatal mortality was 23% (an average rate difference of -0.77 per 1000), whereas Apgar scores <7 increased slightly by 2.8% (a rate difference of 0.14). Rate differences in the Apgar score <7 between these two periods

Table 1. Distribution of Apgar scores at 5 minute by country in 2010

Country	Code	Total live births ^a N	5-minute Apgar score						
			<7			≥7			
			<4%	4–6%	All <7%	7%	8%	9%	10%
Austria	AT	78 609	0.2	0.6	0.8	0.8	3.2	12.0	83.2
BE ^b : Brussels	BE:BR	24 742	0.2	1.4	1.6	2.0	5.1	16.9	74.5
BE ^b : Flanders	BE:FL	69 575	0.3	1.3	1.5	1.9	6.1	37.2	53.2
BE ^b : Wallonia	BE:WA	38 083	0.3	1.2	1.4	1.8	5.1	20.7	71.0
Cyprus ^c	CY	8529	0.1	0.4	0.5	1.3	4.3	13.4	80.6
Czech Republic	CZ	116 399	0.2	0.8	1.0	1.1	3.5	14.4	80.0
Denmark	DK	62 902	0.3	0.6	0.8	0.6	1.7	4.7	92.1
Estonia	EE	15 774	0.2	1.0	1.2	2.6	16.5	62.2	17.6
Finland ^c	FI	61 080	0.4	1.7	2.1	2.1	6.6	66.4	22.8
France	FR	14 602	0.2	0.6	0.8	0.8	1.7	3.9	92.7
Germany	DE	632 780	0.2	0.8	1.0	1.3	4.4	16.4	77.0
Iceland	IS	4903	0.7	1.7	2.4	3.0	8.3	32.3	54.1
Italy	IT	538 177	0.3	0.5	0.8	0.8	5.0	31.8	61.6
Latvia	LV	19 043	0.2	1.4	1.6	4.6	20.9	64.1	8.8
Lithuania	LT	30 763	0.1	0.2	0.3	0.6	6.4	33.8	58.9
Luxembourg	LU	6493	0.1	0.8	0.9	1.0	4.4	15.8	77.9
Malta	MT	4013	0.2	0.7	0.9	0.7	1.7	82.7	14.0
Netherlands	NL	177 649	0.4	1.2	1.6	1.4	4.0	17.1	76.0
Norway	NO	62 345	0.4	1.0	1.4	1.2	3.9	31.9	61.6
Slovenia	SI	22 292	0.2	0.7	0.9	0.8	3.7	59.5	35.1
Sweden	SE	113 950	0.3	1.0	1.3	1.1	2.4	10.2	85.0
UK: Scotland	UK:S	56 756	0.5	1.0	1.5	0.9	2.6	76.3	18.7
UK: Wales	UK:W	33 173	0.3	1.0	1.2	0.9	2.4	43.9	51.6
All countries		2 192 632	0.3	0.8	1.1	1.2	4.5	25.3	67.9

^aFor which Apgar stated.

^bRegion of Belgium; note data from Brussels includes women delivering in maternity units in Brussels and includes some women transferred into hospitals from outside the region.

^cMissing data at 5 minute imputed from data at 1 minute (see methods).

were positively correlated with mortality rate differences (Figure 3). In sensitivity analyses removing the two smallest countries (Malta and Luxembourg) the estimation of the correlation coefficient remained positive, but the significance level declined ($\rho=0.44$, $P = 0.09$, 16 countries). The difference in the neonatal mortality rate was similarly correlated with rate differences in the proportions of Apgar scores <4 ($\rho = 0.46$, $P = 0.06$, $N = 18$) and <9 ($\rho = 0.51$, $P = 0.03$, 18 countries).

Comment

Principal findings

There were wide variations in the prevalence of newborns with low 5 minute Apgar score <7 across European countries, but our results suggest that some of this variability reflects heterogeneity in clinical

scoring practices as opposed to true differences in biomedical outcomes. While there was no correlation between neonatal mortality rates and low Apgar scores in both years included in this study, changes at the country-level in proportions of newborns with low Apgar scores over time were associated with changes in neonatal mortality rates.

Strengths and limitations

This investigation had several strengths. We included data covering a broad geographical area representative of all regions of Europe. Data were collected for the same years using consistent data sources within countries. In addition, common data collection protocols and definitions of indicators (i.e. Apgar score and neonatal mortality rate) were employed. The use of aggregate data was a limitation of the study as it

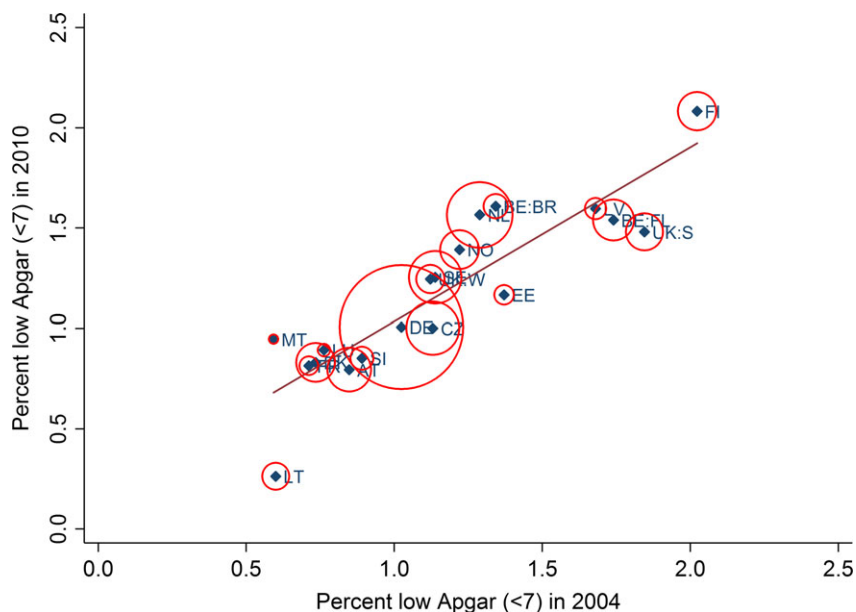


Figure 1. Correlation of proportions of low Apgar scores <7 in 2004 and 2010
 NOTES: red circles proportional to annual number of births within each country; Spearman’s rho 0.88 ($p \leq 0.01$), $n=19$.

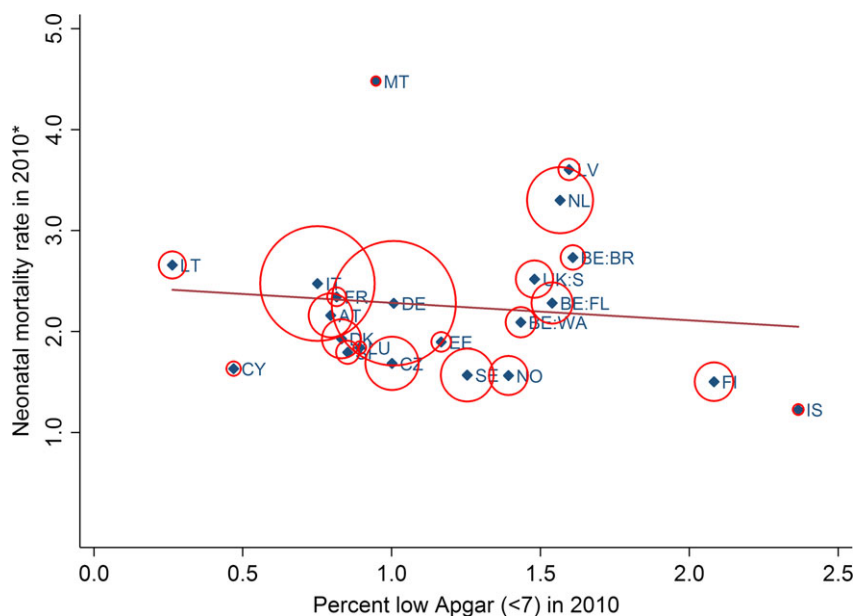


Figure 2. Correlation of Neonatal Mortality Rates with Proportions of Apgar Scores <7 in 2010. NOTES: *Neonatal deaths per 1,000 live births; red circles proportional to annual number of births within each country; Spearman’s rho: -0.06 ($p=0.78$), $N=22$.

prevented the ability to explore medical and social covariates of interest or allow for more detailed statistical analyses within subgroups, for instance, assessing mortality rates by individual 5 minute Apgar scores or analysing Apgar scores by gestational age week. We were not able to identify cause-specific mortality which may be linked to the Apgar score, such as asphyxia or deaths due to very preterm birth. Another potential limitation was differences in data sources between countries which could lead to differences in data quality or other parameters.

Comparison with other studies and interpretation

The range of European 5 minute Apgar scores <7 identified in this study is consistent with those from other developed countries for which national data are available. In Australia, approximately 0.9% of all scores were less than 7 during 2004–09, however, the rates were higher than the national average among indigenous populations (1.4%) and babies delivered in public hospitals (1.1%).²⁵ In the United States, the overall proportion of scores less than 7 was 1.9% in 2012, however, the

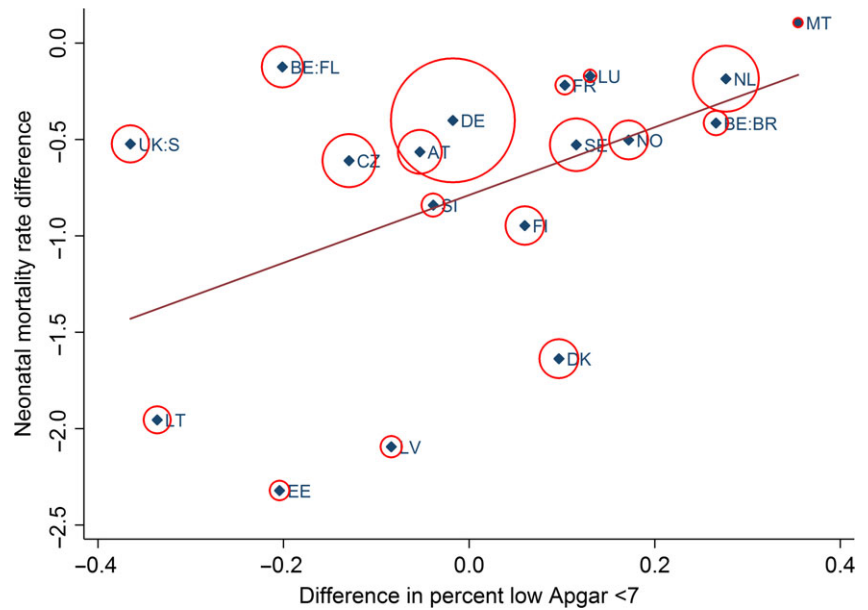


Figure 3. Correlation of Neonatal Mortality Rate Differences with Differences in Percent of Apgar Scores <7 in 2010 and 2004. NOTES: red circles proportional to annual number of births within each country; Spearman's rho: 0.56 ($p=0.02$), $N=19$.

percentage was notably higher for Black babies (3.2%).²⁶

Unlike our findings, prior clinical research as well as population based studies using routinely collected data found strong associations between neonatal mortality and low Apgar scores.^{6–11,27} However, our study differed from these previous analyses in that it used cross-national ecological correlations of low Apgar scores and neonatal mortality rates. We hypothesise that some of the observed variation in Apgar scores reflect country-specific variations in scoring practices which may be a result of clinical training or convention. To investigate the effect of coding, we examined the more frequent use of a score of 10 as opposed to 9 for healthy babies (coded 9 or 10), which we considered to primarily reflect scoring practices because of the very large differences observed across countries in the attribution of the scores 10 vs. 9. The fact that a higher proportion of 10s was correlated with low scores suggests that scoring differences for healthy babies may cause a shift across the whole distribution and less stringent application of the score to high risk babies. Some of the variation in the prevalence of 9 vs. 10 scores could also reflect real differences between countries; one Canadian study found higher risk of developmental vulnerability at 5 years of age among the 82% of babies with an Apgar score of 9 vs. the 6% with an Apgar score of 10.²⁸ However, our study illustrates that findings such as these

must be translated with caution into other contexts, as we observed a range from 9% to 93% in scores of 10. The subjective nature of the Apgar score's various components as well as differences in clinical conditions, such as resuscitation, have previously been posited as reasons for inter-observer variability within countries.^{17–20}

We could not take into consideration the effect of early neonatal interventions such as resuscitation and mechanical ventilation. The uncertainty of assigning Apgar scores in these circumstances could be one of the reasons for the differences in cross-national scoring. 'Expanded' or 'combined' Apgar scores with additional objective clinical components describing resuscitative measures have been proposed which may be better suited to modern healthcare settings.^{4,14,29,30} These have been shown in some instances to be superior in predicting outcomes as compared to the traditional Apgar score.³¹

While these results call into question the external validity of the Apgar score for comparing the prevalence of high risk newborns at birth across countries, we found correlations between changes in Apgar scores and neonatal mortality over time. By using rate differences, country specific errors in measurement of indicators would be minimised. These results suggest that this indicator could provide useful information about national trends in newborn health, however, they should be replicated using more complete time series data and more countries.

Implications for future research, policy, and practice

This study suggests that comparisons of Apgar scores across countries should not be undertaken without first assessing the external validity of this indicator. Despite the existence of a common definition and established guidelines for assigning Apgar scores, its translation into practice was highly heterogeneous across countries. Although there is wide agreement on the <7 threshold for defining low scores, this threshold would potentially select different subsets of children depending on a specific country's coding practices. Higher thresholds may be even more vulnerable to variations in scoring practices, as shown by large differences in the prevalence of newborns with Apgar scores of 9 vs. 10. However, integrating more information on delivery, cause-specific mortality, gestational age, as well as early neonatal interventions could make it possible to improve post hoc standardisation of the scoring system across different contexts by, for instance, comparing Apgar scores among term infants or taking into consideration the proportion of infants who are ventilated.

More generally, this study highlights the difficulty of conducting valid cross-national comparisons of health and the importance of examining the external validity of indicators used for these comparisons. This may be true for other health indicators which presume a common definition and shows the need for research to validate indicators used to benchmark health and care internationally.

Conclusion

Internationally comparative perinatal health indicators make it possible to assess differences in practices and outcomes and to assess interventions towards the creation of best practices. However, while the Apgar score is available in many routine databases and may have good internal validity, as shown by many previous studies linking it to perinatal and child health outcomes, our analysis suggests that the external validity in cross-national comparisons is limited. These results do not support the use of the Apgar score to benchmark newborn health across countries without more work to standardise coding and reporting; however, monitoring trends over time in Apgar scores could provide additional information about newborn health and quality of obstetrical and early neonatal care.

Further research, integrating additional information on pregnancy and newborn characteristics and interventions would be useful to confirm these findings.

Acknowledgements

This study was funded by grants from the European Commission for the Euro-Peristat project: 2010 13 01 and for the Bridge Health project: 664691. The funding agency was not involved in the study. Marie Delnord received doctoral funding from Paris Descartes University, Paris, France.

Project coordination: Jennifer Zeitlin, Marie Delnord, Ashna Hindori-Mohangoo.

References

- 1 Apgar V. A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia & Analgesia* 1953; 32:260.
- 2 Apgar V, Holaday DA, James LS, Weisbrot IM, Berrien C. Evaluation of the newborn infant-second report. *Journal of the American Medical Association* 1958; 168:1985–1988.
- 3 Watterberg KL, Aucott S, Benitz WE, Cummings JJ, Eichenwald EC, Goldsmith J, et al. The Apgar Score. *Pediatrics* 2015; 136:819–822.
- 4 ACOG. Committee Opinion No. 644: The Apgar Score. *Obstetrics & Gynecology* 2015; 126:e52–e55.
- 5 Papile L-A. The Apgar score in the 21st century. *New England Journal of Medicine* 2001; 344:519–520.
- 6 Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *New England Journal of Medicine* 2001; 344:467–471.
- 7 Drage J, Kennedy C, Schwarz B. The Apgar score as an index of neonatal mortality: a report from the Collaborative Study of Cerebral Palsy. *Obstetrics & Gynecology* 1964; 24:222–230.
- 8 Iliodromiti S, Mackay DF, Smith GC, Pell JP, Nelson SM. Apgar score and the risk of cause-specific infant mortality: a population-based cohort study. *The Lancet* 2014; 384:1749–1755.
- 9 Moster D, Lie RT, Irgens LM, Bjerkedal T, Markestad T. The association of Apgar score with subsequent death and cerebral palsy: a population-based study in term infants. *The Journal of Pediatrics* 2001; 138:798–803.
- 10 Thorngren-Jerneck K, Herbst A. Low 5-Minute Apgar Score: a population-based register study of 1 million term births. *Obstetrics & Gynecology* 2001; 98:65–70.
- 11 Li F, Wu T, Lei X, Zhang H, Mao M, Zhang J. The apgar score and infant mortality. *PLoS One* 2013; 8:e69072.
- 12 Ehrenstein V, Pedersen L, Grijota M, Nielsen GL, Rothman KJ, Sørensen HT. Association of Apgar score at five minutes with long-term neurologic disability and cognitive function in a prevalence study of Danish conscripts. *BMC Pregnancy and Childbirth* 2009; 9:1.

- 13 Stuart A, Olausson PO, Källen K. Apgar scores at 5 minutes after birth in relation to school performance at 16 years of age. *Obstetrics & Gynecology* 2011; 118:201–208.
- 14 American Academy of Pediatrics, American College of Obstetricians and Gynecologists, and Committee on Obstetric Practice. The Apgar score. *Pediatrics* 2006; 117: 1444–1447.
- 15 Weinberger B, Anwar M, Hegyi T, Hiatt M, Koons A, Paneth N. Antecedents and neonatal consequences of low Apgar scores in preterm newborns: a population study. *Archives of Pediatrics & Adolescent Medicine* 2000; 154:294–300.
- 16 Zeitlin J, Mortensen L, Cuttini M, Lack N, Nijhuis J, Haidinger G, *et al.* Declines in stillbirth and neonatal mortality rates in Europe between 2004 and 2010: results from the Euro-Peristat project. *Journal of Epidemiology and Community Health* 2016; 70:609–615. doi: 10.1136/jech-2015-207013.
- 17 O'Donnell CP, Kamlin COF, Davis PG, Carlin JB, Morley CJ. Interobserver variability of the 5-minute Apgar score. *The Journal of Pediatrics* 2006; 149:486–489.
- 18 O'Donnell CP, Kamlin COF, Davis PG, Carlin JB, Morley CJ. Clinical assessment of infant colour at delivery. *Archives of Disease in Childhood-Fetal and Neonatal Edition* 2007; 92:F465–F467.
- 19 Rüdiger M, Küster H, Herting E, Berger A, Müller C, Urlesberger B, *et al.* Variations of Apgar score of very low birth weight infants in different neonatal intensive care units. *Acta Paediatrica* 2009; 98:1433–1436.
- 20 Bashambu MT, Whitehead H, Hibbs AM, Martin RJ, Bhola M. Evaluation of interobserver agreement of apgar scoring in preterm infants. *Pediatrics* 2012; 130:e982–e987.
- 21 Zeitlin J, Mohangoo A, Cuttini M. The European Perinatal Health Report: comparing the health and care of pregnant women and newborn babies in Europe. *Journal of Epidemiology and Community Health* 2009; 63:681–682.
- 22 Zeitlin J, Mohangoo A, Delnord M, Cuttini M, Haidinger G, Alexander S, *et al.* The second European Perinatal Health Report: documenting changes over 6 years in the health of mothers and babies in Europe. *Journal of Epidemiology and Community Health* 2013; 67:983–985.
- 23 Euro-Peristat project with SCPE EUROCAT and EURONEOSTAT. *European Perinatal Health Report. Better Statistics for Better Health for Pregnant Women and their Babies in 2004*, 2008. <http://www.europeristat.com/images/doc/EPHR/european-perinatal-health-report.pdf> [last accessed 29 March 2017].
- 24 Euro-Peristat project with SCPE and EUROCAT. *European Perinatal Health Report. The Health and Care of Pregnant Women and Babies in Europe in 2010*. 2013. http://www.europeristat.com/images/doc/EPHR2010_w_disclaimer.pdf [last accessed 29 March 2017].
- 25 AIHW National Perinatal Epidemiology and Statistics Unit and AIHW 2013. *National Core Maternity Indicators*. Cat. no. PER 58. Canberra: AIHW.
- 26 Martin JA, Hamilton BE, Ventura SJ, and Mathews TJ. *Births: Final Data for 2011. National vital statistics reports*, 2013; 62:1–69.
- 27 Ma XL, Xu XF, Chen C, Yan CY, Liu YM, Liu L, *et al.* Epidemiology of respiratory distress and the illness severity in late preterm or term infants: a prospective multi-center study. *Chinese Medical Journal (English Edition)* 2010; 123:2776.
- 28 Razaz N, Boyce WT, Brownell M, Jutte D, Tremlett H, Marrie RA, *et al.* Five-minute Apgar score as a marker for developmental vulnerability at 5 years of age. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 2016; 101: F114–F120.
- 29 Dalili H, Nili F, Sheikh M, Hardani AK, Shariat M, Nayeri F. Comparison of the four proposed Apgar scoring systems in the assessment of birth asphyxia and adverse early neurologic outcomes. *PLoS One* 2015; 10:e0122116.
- 30 Bharti B, Bharti S. A review of the Apgar score indicated that contextualization was required within the contemporary perinatal and neonatal care framework in different settings. *Journal of Clinical Epidemiology* 2005; 58:121–129.
- 31 Dalili H, Sheikh M, Hardani AK, Nili F, Shariat M, Nayeri F. Comparison of the Combined versus Conventional Apgar Scores in Predicting Adverse Neonatal Outcomes. *PLoS One* 2016; 11:e0149464.

Appendix

The Euro-Peristat Scientific Committee

Gerald Haidinger (Austria), Sophie Alexander (Belgium), Pavlos Pavlou (Cyprus), Petr Velebil (Czech Republic), Laust Mortensen (Denmark), Luule Sakkeus (Estonia), Mika Gissler (Finland), Béatrice Blondel (France), Nicholas Lack (Germany), Aris Antsaklis (Greece), István Berbik (Hungary), Helga Sól Ólafsdóttir (Iceland), Sheelagh Bonham (Ireland), Marina Cuttini (Italy), Janis Misins (Latvia), Jone Jaselioniene (Lithuania), Yolande Wagener (Luxembourg), Miriam Gatt (Malta), Jan Nijhuis (Netherlands), Kari Klungsoyr (Norway), Katarzyna Szamotulska (Poland), Henrique Barros (Portugal), Mihai Horga (Romania), Jan Cap (Slovakia), Natasa Tul Mandić (Slovenia), Francisco Bolúmar (Spain), Karin Gottvall (Sweden), Sylvie Berrut (Switzerland), Alison Macfarlane (United Kingdom).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Apgar score guidelines (ACOG 2015).

Table S2 Distribution of Apgar scores at 5 minute by country in 2004.

Table S3 Neonatal mortality rates in 2004 and 2010 and rate differences 2010–2004 for countries with data in at least 1 year.