PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	1-minute and 5-minute Apgar scores and child developmental health at 5 years of age, a population-based cohort study in British Columbia, Canada
AUTHORS	Razaz, Neda; Cnattingius, Sven; Persson, Martina; Tedroff, Kristina; Lisonkova, Sarka; Joseph, K

VERSION 1 - REVIEW

REVIEWER	Jeffrey Roth University of Florida, USA
REVIEW RETURNED	26-Nov-2018

GENERAL COMMENTS	This paper examined Apgar scores of 150,081 full-term infants born in British Columbia between 1993 and 2009 who were subsequently evaluated by their kindergarten teacher with the Early Development Instrument [EDI] between 1999 and 2014. The study employed multivariable log-linear regression to estimate the association between 1- and 5-minute Apgar scores and "developmental vulnerability" [a score below the 10th percentile on any one of EDI's five domains] and "special needs" [a chronic medically, physically, or intellectually disabling condition, determination by whom or by what assessment measure not stipulated]. The authors found that in this sample of healthy children [infants born with major congenital anomalies were excluded [number not given] as were 470 infants with Apgar scores of 0 at 1 or 5 minutes who did not have diagnostic codes of birth asphyxia, resuscitation, or ventilation] 30.2% were judged developmentally vulnerable and 2.5% required "special assistance" [not otherwise defined]. On the basis of finding that "compared with children who had Apgar scores of 10 at both 1 and 5 minutes, children whose 1 minute Apgar score decreased from 10 to a 5-minute Apgar score of <10, had 1.53 times higher rates of developmental vulnerability" [P9,L3-8], the authors recommend "monitoring all infants with 1 minute Apgar scores of 10 (to identify those whose scores may be declining)" [P11,L52 . . .]. This study extends, by enlarging time frame and sample size, work previously done also with a healthy full-term infant sample from another Canadian province [Manitoba] using the EDI (Razaz N et al. Five-minute Apgar score as a marker for developmental vulnerability at 5 years of age. Arch Dis Child Fetal Neonatal Ed 2016;101:F114–F120). Not surprisingly, the findings and implications of the two studies are similar, namely, that "the Apgar score might well serve as a population-level indicator of developmental risk" [F119]. I have no objection to similarity in methods and results of past and present study, given how

	scrupulously both studies have been executed. In comments below, I suggest several additions that might further distinguish the contribution of this latest analysis. • The authors cite only one study that questioned the value of Apgar scores in predicting outcomes [Bharati, 2005]. To update the skepticism about using the Apgar score beyond reporting the status of the newborn infant immediately after birth and the response to resuscitation if needed, please consider citing this passage from a recent Policy Statement by the American Academy of Pediatrics: "The Apgar . . . does not predict individual neurologic outcome; and should not be used for that purpose" [p. 821]. American Academy of Pediatrics Committee on Fetus and Newborn. The Apgar Score. Pediatrics. 2015 Oct;136(4):819-22. • While not equivalent to "developmental vulnerability," developmental delay in children under 5 has recently been estimated globally at between 1 and 3%. [Mithyantha R, Kneen R, McCann E, Gladstone M. Current evidence-based recommendations on investigating children with global developmental delay. Arch Dis Child. 2017 Nov;102(11):1071-1076]. Is it possible that the elevated rate of 30.2% for all children in Kindergarten born full term judged to be developmentally vulnerable may be the result of defining it as scoring under the 10th percentile in any one of the EDI's five domains? • the Methods section state: "the study population included all singleton term (≥37 weeks' gestation) infants born between April 1, 1993 and December 31, 2009, who had documented 1-minute and 5-minute Apgar scores as well as a completed EDI assessment" [P6, L22-26]. If that is the case, why do Tables 2 and 3 give different numbers for children with 1- and 5-minute Apgar scores? Similarly, why is the total number of children with developmental vulnerability 150081 [Table 2] whereas the total number of children with special needs is 148699 [Table 3]? • Please comment on the different percentages of developmental vulnerability in Tables 2 and 3 in Apgar Score at 1-minute (consistently declines) and 5-minute (bounces up and down) • It appears that the 5-minute Apgar score was a more stable predictor of developmental vulnerability than the 1-minute score: 6 Adjusted Rate Ratios were significant at 5 minutes compared to 1 at 1 minute. Likewise for Special Needs: 7 Adjusted Rate Ratios were significant at 5 minutes vs. 3 at 1 minute. Fact worth adding to Discussion section? • Wasn't it surprising that children with a 5-minute Apgar score of 2 had just a 57% higher rate of developmental vulnerability compared with children with an Apgar score of 10 at 5-minute [P9, L6-8], whereas children with 5-minute Apgar score of 2 had a 517% higher rate of Special Needs • Please describe the kind of support that ought to be provided in the delivery room to an infant whose Apgar score is 10 at 1 minute and is 9 at 5 minutes. • Since your database contains the clinical information, why not supply the chief diagnoses that accompanied the three Apgar score levels: 0-3 critically low; 4-6 abnormal; 7-10 normal? • Please explain why "there are no plans to disseminate the results of the research to study participants or the relevant patient community" [P8, L22-24].

REVIEWER	Leif Nelin Nationwide Children's Hospital, Columbus, Ohio, USA
REVIEW RETURNED	06-Dec-2018

GENERAL COMMENTS	This is a potentially interesting study examining how Apgar scores are associated with behavioral outcomes at school age using an assessment tool (EDI) filled out by teachers when the children are in kindergarten. The premise is that even so called normal Apgar scores may be associated with poorer outcomes at school age. The manuscript is well written, however the data presentation is dense and difficult to follow, and the conclusions drawn from this observational study are far too strong. My specific comments follow. 1. The study cohort spans nearly 16 years and includes 150,000 children born at term without major malformations and complete Apgar and EDI data in British Columbia. However, the British Columbia website suggests that there were at least 40,000 births annually during this time period of which a little less than 10% were preterm, so a conservative estimate would be about 600,000 term births during the study period. Thus, the question arises are the 150,000 term births studied really a representative sample? Did the exclusion of certain children result in bias, particularly when only about a quarter to half of children were included? This draws into question the generalizability of this cohort. A much more thorough characterization of the cohort is necessary to fully understand the presented results. 2. The authors report that 30% of their cohort had EDI scores below the 10th percentile on the EDI and that as recommended the 10th percentile was used as the cutoff for developmental vulnerability. Does this also suggest that there is something special about their cohort compared to the Canadian population at large? Again suggesting the potential for bias. 3. This is an observational study examining associations, therefore discussing potential changes in practice based on these data in the discussion section is not appropriate. Nor for that matter is the use of the terms suggesting causality like the word “resulted”. 4. Furthermore, although 150,000 subjects sounds like a lot, close examination of the data tables reveals that the numbers are small in many of the groups examined. 5. The authors allude to differences in groups that may not be present, i.e. if the 95% CI of the aRR crosses one, as it does in many cases, then the groups aren’t different. Focusing on those comparisons that are actually statistically different would help to focus the results section so that the reader could understand the findings. 6. Table 1 is very difficult to interpret, particularly around the description in the first paragraph of the results section. 7. The authors should comment further on the potential subjectivity of the Apgar score under weaknesses of the study. There are so many things that affect the 1 and 5 minute Apgar scores that strong conclusions regarding practice are not possible. The conclusion here is really that, provided the population is representative and there is no significant bias, the association described between Apgar score and developmental outcome could be used to design either a randomized controlled trial or a prospective rigorous data collection to actually examine the utility of the Apgar score in improving neurodevelopmetnal outcomes.

VERSION 1 – AUTHOR RESPONSE

REVIEWER COMMENTS:

Reviewer: 1

Reviewer Name: Jeffrey Roth

Institution and Country: University of Florida, USA

Please state any competing interests or state 'None declared': None declared.

Please leave your comments for the authors below

This paper examined Apgar scores of 150,081 full-term infants born in British Columbia between 1993 and 2009 who were subsequently evaluated by their kindergarten teacher with the Early Development Instrument [EDI] between 1999 and 2014. The study employed multivariable log-linear regression to estimate the association between 1- and 5-minute Apgar scores and "developmental vulnerability" [a score below the 10th percentile on any one of EDI's five domains] and "special needs" [a chronic medically, physically, or intellectually disabling condition, determination by whom or by what assessment measure not stipulated]. The authors found that in this sample of healthy children [infants born with major congenital anomalies were excluded [number not given] as were 470 infants with Apgar scores of 0 at 1 or 5 minutes who did not have diagnostic codes of birth asphyxia, resuscitation, or ventilation] 30.2% were judged developmentally vulnerable and 2.5% required "special assistance" [not otherwise defined]. On the basis of finding that "compared with children who had Apgar scores of 10 at both 1 and 5 minutes, children whose 1 minute Apgar score decreased from 10 to a 5-minute Apgar score of <10, had 1.53 times higher rates of developmental vulnerability" [P9,L3-8], the authors recommend "monitoring all infants with 1 minute Apgar scores of 10 (to identify those whose scores may be declining)" [P11,L52 . . .].

This study extends, by enlarging time frame and sample size, work previously done also with a healthy full-term infant sample from another Canadian province [Manitoba] using the EDI (Razaz N et al. Five-minute Apgar score as a marker for developmental vulnerability at 5 years of age. Arch Dis Child Fetal Neonatal Ed 2016;101:F114–F120). Not surprisingly, the findings and implications of the two studies are similar, namely, that "the Apgar score might well serve as a population-level indicator of developmental risk" [F119]. I have no objection to similarity in methods and results of past and present study, given how scrupulously both studies have been executed. In comments below, I suggest several additions that might further distinguish the contribution of this latest analysis.

The authors cite only one study that questioned the value of Apgar scores in predicting outcomes [Bharati, 2005]. To update the skepticism about using the Apgar score beyond reporting the status of the newborn infant immediately after birth and the response to resuscitation if needed, please consider citing this passage from a recent Policy Statement by the American Academy of Pediatrics: "The Apgar . . . does not predict individual neurologic outcome; and should not be used for that purpose" [p. 821]. American Academy of Pediatrics Committee on Fetus and Newborn. The Apgar Score. Pediatrics. 2015 Oct;136(4):819-22.

Response: Thank you for your comments; we added this reference in the introduction section of the manuscript.

Introduction (Page 4): Although the value of a low Apgar score for accurately predicting adverse neurologic outcomes at the individual level has been questioned,1 2 low Apgar scores are well correlated with both short-term3 and long-term outcomes, in both preterm and term infants.4-10

While not equivalent to “developmental vulnerability,” developmental delay in children under 5 has recently been estimated globally at between 1 and 3%. [Mithyantha R, Kneen R, McCann E, Gladstone M. Current evidence-based recommendations on investigating children with global developmental delay. Arch Dis Child. 2017 Nov;102(11):1071-1076]. Is it possible that the elevated rate of 30.2% for all children in Kindergarten born full term judged to be developmentally vulnerable may be the result of defining it as scoring under the 10th percentile in any one of the EDI’s five domains?

Response: Yes. The 10th percentile cut-off for “developmental vulnerability” has been recommended 11 because it usually leads to a higher prevalence than clinical cut-offs with a 1-3% prevalence for developmental delay. The ‘vulnerability’ label thus includes children who may be more difficult to diagnose and refers to the portion of the population which, without additional support and care, may experience future challenges in school and society. We have emphasized this point in the Methods section of the manuscript.

Methods (Page 7): The 10th percentile cut-off has been recommended because it is usually higher than clinical cut-off points of 3% or 5% for diagnosing developmental delay. 12 Developmentally vulnerable children may not manifest developmental delays but are at risk to experience challenges in school and society without additional support and care.13

the Methods section state: “the study population included all singleton term (≥37 weeks’ gestation) infants born between April 1, 1993 and December 31, 2009, who had documented 1-minute and 5-minute Apgar scores as well as a completed EDI assessment” [P6, L22-26]. If that is the case, why do Tables 2 and 3 give different numbers for children with 1- and 5-minute Apgar scores? Similarly, why is the total number of children with developmental vulnerability 150081 [Table 2] whereas the total number of children with special needs is 148699 [Table 3]?

Response: Information on special needs was available in 148699 (99.1%) children; [n=1382 had missing information for this variable and were not part of the analyses examining the relationship between the Apgar score and having special needs]. We have now included this information in the revised Results section.

Results (Page 8): There were 150,081 children (mean age = 5.7 years) with a gestational age at birth of ≥37 weeks, without major malformations and complete Apgar and EDI data included in the study. Information on special needs was available in 148,699 (99.1%) children.

Please comment on the different percentages of developmental vulnerability in Tables 2 and 3 in Apgar Score at 1-minute (consistently declines) and 5-minute (bounces up and down)

Response: We agree with the Reviewer that rates of developmental vulnerability did not decline as consistently as the rates for special needs. Nevertheless, overall there was a decreasing trend in the rate of developmental vulnerability with increasing 1 minute and 5-minute Apgar scores (P for trend <0.001). For example, the rate of vulnerability declined from 40.6% for children with 5-min Apgar scores of 4 to 32.8% for children with an Apgar score of 6, and to 29.6% for children with a 5-min Apgar score of 10. We have added a sentence to the revised manuscript commenting on this.

Discussion (page 12): Nevertheless, measurement errors inherent in routinely recorded Apgar scores (and possibly the EDI) may potentially explain the lack of an evident dose-response relationship between Apgar scores and developmental vulnerability.

It appears that the 5-minute Apgar score was a more stable predictor of developmental vulnerability than the 1-minute score: 6 Adjusted Rate Ratios were significant at 5 minutes compared to 1 at 1 minute. Likewise for Special Needs: 7 Adjusted Rate Ratios were significant at 5 minutes vs. 3 at 1 minute. Fact worth adding to Discussion section?

Response: Thank you for your comments; we have added a statement in the discussion highlighting this point.

Results (Page 9): There was an increasing trend in the rate of developmental vulnerability with decreasing 1-minute and 5-minute Apgar scores (P for trend <0.001; Table 2). However, this association was much more pronounced for the 5-minute Apgar score.

Discussion (Page 10): In this population-based study, we found graded, continuously increasing risks of developmental vulnerability and special needs at 5 years of age with decreasing 1- and 5-minute Apgar scores. A low Apgar score at 5-minutes was more strongly associated with developmental vulnerability and special needs than a low Apgar score at 1-minute.

Wasn't it surprising that children with a 5-minute Apgar score of 2 had just a 57% higher rate of developmental vulnerability compared with children with an Apgar score of 10 at 5-minute [P9, L6-8], whereas children with 5-minute Apgar score of 2 had a 517% higher rate of Special Needs

Response: Thank you for your comment, special needs designation follows an already diagnosed relatively serious disabling condition, whereas EDI-defined developmental vulnerability includes conditions ranging from a possible risk of experiencing challenges at school (e.g., a score <10% in only one domain) to a potentially serious developmental disability (e.g., minimal scores in all domains). As such, the proportion of children with Special Needs is very low among those with a 5 minute Apgar score of 10 (reference category) as compared with children who had an Apgar score of 2, yielding a large rate ratio. The proportion of EDI-defined vulnerable children is much higher than the proportion of children with Special Needs (30% vs 2%) among children with an Apgar score of 10 at 5 minutes. An 177% increase (RR 2.77, the upper limit of the 95% CI) in the proportion of EDI yields a proportion of almost 90% (of children). I

Please describe the kind of support that ought to be provided in the delivery room to an infant whose Apgar score is 10 at 1 minute and is 9 at 5 minutes.

Response: It is not clear what support can be provided. At the very least, a deteriorating Apgar score is indicative of the need to re-evaluate the infant for problems adapting to life after birth. We have added a couple of sentences to the revised manuscript.

Discussion (Page 11-12): The higher developmental vulnerability observed among infants whose optimal Apgar score (of 10) at 1 minute falls with time after birth may be important clinically; such a progression may indicate problems with circulatory, respiratory or central nervous system changes that are associated with birth. Deterioration in the Apgar score immediately after birth, therefore, warrants re-evaluation of the infant and close clinical scrutiny in order to exclude congenital abnormalities and drug induced depression of the central nervous system.

Since your database contains the clinical information, why not supply the chief diagnoses that accompanied the three Apgar score levels: 0-3 critically low; 4-6 abnormal; 7-10 normal?

Response: Thank you for your comment. Most studies examining the long-term prognostic value of the Apgar score on child outcomes have focused on the association between extremely low 5-minute Apgar scores, such as those linked to severe birth asphyxia, and distinct neurocognitive profiles such as IQ and academic ability. However, the aim of our study was to examine Apgar scores at 1 and 5 minutes across the entire range of score values rather than focusing on the conventional categories of 0-3, 4-6 and 7-10.

Please explain why "there are no plans to disseminate the results of the research to study participants or the relevant patient community" [P8, L22-24].

Response: This statement was included in the manuscript as it is a requirement for the BMJ Open journal. We have plans for dissemination of findings to physicians, related health-professionals and researchers via publication and through presentations at local, national and international conferences. We have now removed the above-mentioned sentence from the manuscript.

Methods (Page 8): Patient and public involvement: No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for or implementation of the study. No patients were asked to advise on interpretation of the findings.

Reviewer: 2

Reviewer Name: Leif Nelin

Institution and Country: Nationwide Children's Hospital, Columbus, Ohio, USA

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

This is a potentially interesting study examining how Apgar scores are associated with behavioral outcomes at school age using an assessment tool (EDI) filled out by teachers when the children are in kindergarten. The premise is that even so called normal Apgar scores may be associated with poorer outcomes at school age. The manuscript is well written, however the data presentation is dense and difficult to follow, and the conclusions drawn from this observational study are far too strong. My specific comments follow.

The study cohort spans nearly 16 years and includes 150,000 children born at term without major malformations and complete Apgar and EDI data in British Columbia. However, the British Columbia website suggests that there were at least 40,000 births annually during this time period of which a little less than 10% were preterm, so a conservative estimate would be about 600,000 term births during the study period. Thus, the question arises are the 150,000 term births studied really a representative sample? Did the exclusion of certain children result in bias, particularly when only about a quarter to half of children were included? This draws into question the generalizability of this cohort. A much more thorough characterization of the cohort is necessary to fully understand the presented results.

Response: Thank you for your comments; The EDI data were collected in waves (every one to three years) to cover school districts across the entire province (minimum 80% participation required in each school district). Furthermore, although the EDI had broad coverage across BC, it was collected less frequently in independent schools (30% coverage). Parents who enroll their children in independent schools are often more affluent, therefore this public school sample may have under-represented families at higher income levels. As per the Reviewer’s comments this information has been added to the Methods section of the manuscript.

Methods (Page 6): The EDI has been routinely administered province-wide in British Columbia every one to three years since the 1999/2000 school year, achieving at least 85% participation of kindergarten children from each school district.

Discussion (Page 12-13): Furthermore, although the EDI has broad coverage across British Columbia, it is collected less frequently in independent schools (30% coverage). Since parents who enroll their children in independent schools tend to be more affluent, our study population may have under-represented families at higher income.

The authors report that 30% of their cohort had EDI scores below the 10th percentile on the EDI and that as recommended the 10th percentile was used as the cutoff for developmental vulnerability. Does this also suggest that there is something special about their cohort compared to the Canadian population at large? Again suggesting the potential for bias.

Response: The 10th percentile value was based on the national EDI cut-off scores,14 and hence we believe our sample was representative of the Canadian population.

This is an observational study examining associations, therefore discussing potential changes in practice based on these data in the discussion section is not appropriate. Nor for that matter is the use of the terms suggesting causality like the word “resulted”.

Response: Thank you for your comments; we have omitted any terms suggesting causality throughout the manuscript.

Discussion (Page 13): Furthermore, improvements in Apgar scores between 1 and 5 minutes among children with a 1-minute Apgar score of 7 or 9 was associated with lower risk of developmental vulnerability. These results provide clinicians with valuable prognostic information and the justification to carefully monitor infants who are even mildly compromised at 1 and 5 minutes.

Discussion (Page 12): However, our results reveal that lower scores within the normal range (7-9) and even a slight reduction in score from 10 at 1 minute to 9 at 5 minutes are both associated with a significant increase in the risk of developmental vulnerability.

Furthermore, although 150,000 subjects sounds like a lot, close examination of the data tables reveals that the numbers are small in many of the groups examined.

Response: We agree that numbers are small for lower Apgar scores, in particular Apgar scores in the 0-3 range. We have added a statement in the Results highlighting this issue.

Results (Page 9-10): The adjusted risk ratios for having special needs among children with 1 and 5 minutes Apgar scores in the 0-3 range had wide 95% confidence intervals because of small numbers of children in these categories.

The authors allude to differences in groups that may not be present, i.e. if the 95% CI of the aRR crosses one, as it does in many cases, then the groups aren’t different. Focusing on those comparisons that are actually statistically different would help to focus the results section so that the reader could understand the findings.

Response: We have ensured that comparisons that are highlighted in the Results section are either significant or part of a pattern which had a significant p-value for trend.

Table 1 is very difficult to interpret, particularly around the description in the first paragraph of the results section.

Response: We have simplified Table 1.

The authors should comment further on the potential subjectivity of the Apgar score under weaknesses of the study. There are so many things that affect the 1 and 5 minute Apgar scores that strong conclusions regarding practice are not possible. The conclusion here is really that, provided the population is representative and there is no significant bias, the association described between Apgar score and developmental outcome could be used to design either a randomized controlled trial or a prospective rigorous data collection to actually examine the utility of the Apgar score in improving neurodevelopmental outcomes.

Response: We have highlighted the potential subjectivity of the Apgar score in the limitation section of the manuscript. As per the reviewer's comment we have also added a statement regarding future directions for research.

Discussion (Page 12-13): We recognize that the Apgar score as recorded in medical charts represents routine clinical practice, and is prone to interobserver variability, specifically in intubated newborn babies. However, the quality of Apgar score values should not differ between children with and without subsequently diagnosed developmental vulnerability. Nevertheless, measurement errors inherent in routinely recorded Apgar scores (and possibly the EDI) may potentially explain the lack of an evident dose-response relationship between Apgar scores and developmental vulnerability.

Conclusion: (Page 13): Future studies should examine the underlying mechanism by which Apgar score in the normal ranges could influence long-term neurodevelopmental outcomes.

VERSION 2 – REVIEW

REVIEWER	Jeffrey Roth University of Florida, USA
REVIEW RETURNED	27-Jan-2019

GENERAL COMMENTS	In this revision, the authors have made 12 changes to their previously submitted manuscript. These additions do improve the paper but there remains one issue that needs further clarification. Reviewer 2 was concerned about potential bias in sample selection. My concern is actual bias in the interpretation of results. The study's objective was to investigate associations between Apgar score and child development health at 5 years of age "across the entire range of score values" [P2, L8-10]. Commitment to this objective involved including children with critically low scores (0-3) in the sample, even though this group comprised only 2% of the data (3313/150081). Children with an Apgar score of 2 showed much higher risks of developmental vulnerability (aRR 1.57 at 5 minutes) and special needs (aRR (1.72 at 1 minute). Neither of these results was highlighted in the Abstract. Instead we are told that children with a score of 9 at 1 and 5 minutes had an aRR of 1.03. Yes, far more children in this healthy sample scored a 9 but a 3% elevated risk is hardly grounds to "carefully monitor infants who are even mildly compromised," especially in light of the fact that the p value for these analyses was relaxed to 0.10. As Reviewer 2 pointed out, the confidence levels for one minute scores "across the entire range of score values" in Tables 2 and 3 are mostly non-significant (9 of 10 for developmental vulnerability; 7 of 10 for special needs). Clearly the authors see their main novel finding as "improvements in Apgar scores between 1 and 5 minutes among children with a 1- minute Apgar score of 7 to 9 were associated with a lower risk of developmental vulnerability" (P14, L10-13]. Commitment to the "full range of score values" [P12, L5] has generated an unbalanced sample such that findings about the much higher elevated risks of the critically low group are downplayed for statistical reasons: "1 and 5 minutes Apgar scores in the 0-3 range had wide 95% confidence intervals because of small numbers of children in these categories" [P10, L40-44]. One

	possibly extravagant approach to remedy this mixed message would be to restrict the analysis to the generally normal group (7-10). In fact, the authors put forth this very rationale in their Introduction: “the long-term significance of changes in such Apgar scores within the “normal” range (i.e., 7-10) is not clear” [P6, L5-8]. Virginia Apgar (and many others since) said the test was not designed to predict long term health outcomes; its purpose is to quickly determine 1) the effects of obstetric anesthesia on infants and 2) whether immediate medical care is needed. That said, the authors cite 7 studies (References 5-11) that have looked at Apgar scores and long term outcomes. Though the number of References is at the journal’s limit of 35, a recent article (Tweed EJ, et al. Five-minute Apgar score and educational outcomes: retrospective cohort study of 751 369 children. Arch Dis Child Fetal Neonatal Ed 2016;101:F121–F126) merits being substituted since it concludes: “Apgar scores are therefore associated with long-term as well as short term prognoses and with educational as well as clinical outcomes at the population level” (p. F125). Minor matters • Since each independent variable lists its constituent categories, the parenthesis after socioeconomic status should state “quintiles” • “problems with circulatory, respiratory or central nervous system changes . . . associated with delivery” (rather than “birth”) [P13, L5] • It does not appear that the “study population may have under-represented families at higher income” [P13, L35-37] since 2/5 of the sample were from the two highest SES quintiles

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Jeffrey Roth

Institution and Country: University of Florida, USA

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

In this revision, the authors have made 12 changes to their previously submitted manuscript. These additions do improve the paper but there remains one issue that needs further clarification.

Reviewer 2 was concerned about potential bias in sample selection. My concern is actual bias in the interpretation of results. The study’s objective was to investigate associations between Apgar score and child development health at 5 years of age “across the entire range of score values” [P2, L8-10]. Commitment to this objective involved including children with critically low scores (0-3) in the sample, even though this group comprised only 2% of the data (3313/150081). Children with an Apgar score of 2 showed much higher risks of developmental vulnerability (aRR 1.57 at 5 minutes) and special needs (aRR (1.72 at 1 minute). Neither of these results was highlighted in the Abstract. Instead we are told that children with a score of 9 at 1 and 5 minutes had an aRR of 1.03. Yes, far more children in this healthy sample scored a 9 but a 3% elevated risk is hardly grounds to “carefully monitor infants who are even mildly compromised,” especially in light of the fact that the p value for these analyses was relaxed to 0.10.

Response: Thank you for your comment, we have now highlighted the higher risk of developmental vulnerability for the 5-minute Apgar of 2 in the Abstract. Due to the word limit for the Abstract we focused the Results section on Apgar score between 7 and 9, given that the relation between low Apgar score of less than 7 or less than 4 and adverse outcomes is well-established in the literature. Our study demonstrates that even children with a 5-minute Apgar scores in the normal range are at increased risk of developmental vulnerability.

Note: The more liberal P value criterion of <0.10 was used for retaining potentially confounding variables in the regression model (in line with standard regression modeling procedures). This did not apply to the main exposure variable, namely, the Apgar score, for which the traditional P value cut-off of 0.05 was used to determine statistical significance.

Abstract (Results): Compared with children with an Apgar score of 10 at 5-minute, the aRR for developmental vulnerability increased steadily with decreasing Apgar score from 1.02 (95% CI 1.00-1.04) for an Apgar score of 9 to 1.57 (95% CI 1.03-2.39) for an Apgar score of two.

As Reviewer 2 pointed out, the confidence levels for one minute scores “across the entire range of score values” in Tables 2 and 3 are mostly non-significant (9 of 10 for developmental vulnerability; 7 of 10 for special needs). Clearly the authors see their main novel finding as “improvements in Apgar scores between 1 and 5 minutes among children with a 1- minute Apgar score of 7 to 9 were associated with a lower risk of developmental vulnerability” (P14, L10-13]. Commitment to the “full range of score values” [P12, L5] has generated an unbalanced sample such that findings about the much higher elevated risks of the critically low group are downplayed for statistical reasons: “1 and 5 minutes Apgar scores in the 0-3 range had wide 95% confidence intervals because of small numbers of children in these categories” [P10, L40-44]. One possibly extravagant approach to remedy this mixed message would be to restrict the analysis to the generally normal group (7-10). In fact, the authors put forth this very rationale in their Introduction: “the long-term significance of changes in such Apgar scores within the “normal” range (i.e., 7-10) is not clear” [P6, L5-8].

Response: Our primary objective was to demonstrate the associations between Apgar scores at 1 and 5 minutes across the entire range of score values and developmental health, and the secondary objective was to examine the effect of s change in score for children with 1-minute Apgar scores in the 7-9 range. We believe that it is important for the readers to see the increasing trend in the rate of developmental vulnerability and special needs with decreasing 1-minute and 5-minute Apgar scores across the entire range, and restricting our study sample to children with Apgar scores of 7-9 will lead to loss of valuable information.

Virginia Apgar (and many others since) said the test was not designed to predict long term health outcomes; its purpose is to quickly determine 1) the effects of obstetric anesthesia on infants and 2) whether immediate medical care is needed. That said, the authors cite 7 studies (References 5-11) that have looked at Apgar scores and long term outcomes. Though the number of References is at the journal’s limit of 35, a recent article (Tweed EJ, et al. Five-minute Apgar score and educational outcomes: retrospective cohort study of 751 369 children. Arch Dis Child Fetal Neonatal Ed 2016;101:F121–F126) merits being substituted since it concludes: “Apgar scores are therefore associated with long-term as well as short term prognoses and with educational as well as clinical outcomes at the population level” (p. F125).

Response: Thank you for your comment, we added the Tweet EJ et al study to the list of references.

Minor matters

• Since each independent variable lists its constituent categories, the parenthesis after socioeconomic status should state “quintiles”

Response: We added quintiles in parenthesis after socioeconomic status in the list.

• “problems with circulatory, respiratory or central nervous system changes . . . associated with delivery” (rather than “birth”) [P13, L5]

Response: We revised the sentence to “problems with physiologic circulatory, respiratory or central nervous system changes . . . associated with delivery”

• It does not appear that the “study population may have under-represented families at higher income” [P13, L35-37] since 2/5 of the sample were from the two highest SES quintiles

Response: Thank you for pointing this out, we removed this statement from the Discussion section.

VERSION 3 - REVIEW

REVIEWER	Jeffrey Roth University of Florida USA
REVIEW RETURNED	01-Mar-2019

GENERAL COMMENTS	This revision--and the explanations in the accompanying Author’s Response--successfully address each of the issues raised in my previous review. The manuscript is extremely well written and its findings deserve to be added to the growing literature on the use of 1- and 5-minute Apgar scores to anticipate the developmental health status of children attending kindergarten.