Author’s response to reviews

Title: Customized versus population birth weight charts for identification of newborns at risk of long-term adverse cardio-metabolic and respiratory outcomes: A population-based prospective cohort study

Authors:

Jan Erkamp (j.erkamp@erasmusmc.nl)
Vincent Jaddoe (v.jaddoe@erasmusmc.nl)
Annemarie Mulders (a.mulders@erasmusmc.nl)
Eric Steegers (e.a.p.steegers@erasmusmc.nl)
Irwin Reiss (i.reiss@erasmusmc.nl)
Liesbeth Duijts (l.duijts@erasmusmc.nl)
Romy Gaillard (r.gaillard@erasmusmc.nl)

Version: 1 Date: 23 Aug 2019

Author’s response to reviews:

Please refer to the uploaded "Response letter" for a word version of below response to the reviewers.

Dear Dr. Diana Samuel,

Thank you for reviewing our manuscript entitled ‘Customized versus population birth weight charts for identification of newborns at risk of long-term adverse cardio-metabolic and respiratory outcomes: A population-based prospective cohort study’ (BMED-D-19-00939).

We are grateful for the detailed comments of the editors and reviewers. In response to those comments, we have made major changes in our manuscript. Most importantly, we improved the results section by reporting on screening performance of customized and population charts more accurately, and we further improved the discussion based on suggestions by the reviewers.

In the response to the reviewer’s document, you will find our specific responses to all the questions or comments from you and your reviewers, including changes in the manuscript and the figures. Textual changes are highlighted by track changes in the revised manuscript and are underlined in this rebuttal letter.
We hope you will consider the revised manuscript acceptable for publication in BMC Medicine.

Yours sincerely,
On behalf of all authors,
Jan Erkamp, MD, and Romy Gaillard, MD, PhD

Response to the Editorial comments

Editorial comments
Comment 1: Please change the main text headings 'Introduction' to 'Background' and 'Subjects and methods' to 'Methods'.
Response: We changed the main text headings as requested.

Comment 2: Please remove the figures from the main manuscript file and upload them as individual figure files.
Response: We removed the figures from the main manuscript file and uploaded them as individual figure files.

Comment 3: If you intend to include the STROBE checklist as supplementary material, please cite it in the main text and add its title to the list of supplementary figures/tables.
Response: We added the STROBE Checklist to the Supplementary Files and cited the STROBE Checklist in the main manuscript.

Comment 4: Please reverse the order of the in-text references to the supplementary material so that 'Additional file 1' comes first.
Response: We reversed the order of the in-text references to the supplementary material.

Comment 5: Author Irwin KM Reiss still needs to confirm authorship for this manuscript. We have resent the email requesting this, but would appreciate it if you could encourage this author to complete this.
Response: Irwin KM Reiss confirmed authorship for this manuscript.

Response to the reviewers
Reviewer 1
This is a well-conducted and novel study on an important topic. The results are presented in detail and the manuscript is well-written. I have written some specific comments below.
Response: We thank the reviewer for the comments. Please see our detailed response to each comment below.

Response: We thank the reviewer for this suggestion and incorporated this publication within the Discussion Section: “A recent population-based linkage study among 979,912 singleton pregnancies in the United Kingdom between 1992 and 2010 assessed the predictive ability of non-customized versus partially customized birth weight centiles for the prediction of the risks of stillbirth, infant death and neonatal morbidity. This study showed that partial customization of birth weight charts does not improve prediction of these perinatal complications [1]. For the partial customization, maternal height,
parity and fetal sex were used.” (Page, 15 lines 285 to 290).

Comment 2: The authors cite reference 15 (Verburg et al, 2008) which they used for population charts but this reference only gives reference curves for individual ultrasound measurements (biparietal diameter, head circumference, abdominal circumference and femur length). Could the authors give some more detail on how these were used to calculate population birth weight percentiles? Do the two methods (customized and non-customised) compare, i.e. was the same gestational age adjusted (fetal) weight reference used in both methods with and without customization, so that any difference between them would be explained only through customization for all other variables except for gestational age?
Response: We have clarified our approach for the calculation of the population birth weight percentiles within our study. The reviewer is correct that the same gestational-age adjusted fetal weight reference chart was used in both methods with and without customization. So any difference between the classification by the two charts is explained through customization for all other variables except for gestational age. We added the following sentences to the Methods Section: “Customized charts have been developed within our study cohort as described previously, and include gestational age, fetal sex, maternal parity, age, height, weight, ethnicity, and smoking[2]. The pathological determinant maternal smoking was also used for the development of the customized charts because it has a substantial effect on fetal growth and birth weight and led to a more accurate regression model[2]. For the construction of a customized growth chart, the term for smoking was set to zero, whether the pregnant woman smoked or not. Hereby non-smoking was used as reference category within our customized models. To calculate the customized birth weight percentile, we entered the maternal characteristics, fetal sex and gestational age at birth for each newborn within our customized charts model and compared actual birth weight to the expected weight. For the population charts, we used gestational age adjusted weight charts modeled on the same population[2]. We calculated the birth weight percentile, by entering gestational age at birth for each newborn within our population charts model and compared actual birth weight to the expected weight. The population chart only included gestational age and no other characteristics, which allows for the optimal comparison between the population charts and customized charts in which any difference in outcome would only be explained by the process of customization. The formulas for both the customized charts and population charts have been published previously (ref Gaillard)” (Page 5-6, lines 99-113)

Comment 3: Individual customization of charts in Generation R included smoking. Customization for a pathological variable such as smoking may prevent identification of growth restriction in infants whose growth is affected by it. The authors could discuss their choice of variables chosen for customization.
Response: We agree with the reviewer that addition of a pathological variable to the customized charts model needs to be considered carefully. We have clarified our approach in the Methods Section: “Customized charts have been developed within our study cohort as described previously, and include gestational age, fetal sex, maternal parity, age, height, weight, ethnicity, and smoking. The pathological determinant maternal smoking was also used for the development of the customized charts because it has a substantial effect on fetal growth and birth weight and led to a more accurate regression model[2]. For the construction of a customized growth chart, the term for smoking was set to zero, whether the pregnant woman smoked or not. Hereby non-smoking was used as reference category within our customized models. (Page 5, lines 99-104).

We also addressed this issue in the Discussion section: “Which maternal factors should be included in the customized charts also remains debatable. We included the pathological variable maternal smoking in the construction of the model to obtain a better fitted model. For the construction of a customized growth chart the term for smoking was set to zero, whether the pregnant woman smoked or not, and thereby non-smoking was used as reference category within our customized model for all women. This
approach still allowed us to detect pathological fetal growth restriction due to maternal smoking during pregnancy. A similar approach may also be used for other pathological variables and further improve customized charts. Further studies are needed to explore whether customized charts which consider more maternal factors improve the classification of size at birth.” (Page 17, lines 363-371).

Comment 4: In the tables, the p-values currently presented as 0.000 should be presented as <0.001, since a p-value is never exactly zero. There is no need to dichotomise p-values into "significant" and "non-significant" or mark the "significant" ones with asterisk (*) since all p-values are given with an adequate accuracy (see Colquhoun, 2017. R Soc Open Sci. 2017 Dec; 4(12): 171085 p. 16 points (3) and (4)).
Response: We made the requested changes.

Comment 5: For clarity, Figure 1 title could be changed to "Prevalence of birth weight classifications and their association with infant growth patterns and cardio-metabolic and respiratory outcomes at age 10". The Figure S2 title could also be changed similarly.
Response: We agree with the reviewer and changed the titles of the figures as suggested.

Comment 6: The figures should be readable on their own, e.g. “Clustering” should be explained in the footnote.
Response: We agree with the reviewer and explained Clustering in the footnote of Figure 1 and Figure 2 in the revised manuscript: “Clustering of cardio-metabolic risk factors is defined as having three or more of the following components: visceral fat mass >75th percentile; systolic or diastolic blood pressure >75th percentile; HDL-cholesterol <25th percentile or triglycerides >75th percentile; and insulin level >75th percentile of our study population.”

Comment 7: The authors commented on the faster decrease in stillbirths in England and Wales in the areas that implemented customized charts (reference 28: Gardosi et al, 2013). Causality is hard to establish without a trial, and to put these findings in context, the authors could comment on the results from across Europe where much faster decreases in stillbirth rates have been seen in other countries such as Denmark and the Netherlands (Zeitlin et al, 2016. Declines in stillbirth and neonatal mortality rates in Europe between 2004 and 2010: results from the Euro-Peristat project. J Epidemiol Community Health. 2016 Jun;70(6):609-15).
Response: We agree with the reviewer and have clarified this issue in the Discussion section: “Previous studies mainly focused on the effects of customization on selecting newborns at risk for adverse perinatal outcomes. A meta-analysis including 20 studies comparing the effectiveness of customized versus population charts for prediction of adverse perinatal outcomes has shown similar effect estimates for associations of abnormal size at birth with intra-uterine fetal demise, neonatal intensive care unit admission, and neonatal and perinatal death[3]. A recent population-based linkage study among 979,912 singleton pregnancies in the United Kingdom between 1992 and 2010 assessed the predictive ability of non-customized versus partially customized birth weight centiles for the prediction of the risks of stillbirth, infant death and neonatal morbidity. This study showed that partial customization of birth weight charts does not improve prediction of these perinatal complications [1]. For the partial customization, maternal height, parity and fetal sex were used. Contrary, analysis of data on live births and stillbirths in England and Wales between 2007 and 2012 from the Office of National Statistics, suggested in areas that implemented customized charts, a decline in stillbirth rates of 19% occurred, while stillbirth rates remained the same in areas that did not implement customized charts[4, 5]. However, these findings need to be interpreted carefully and causality cannot be established from these observational studies. Recently, a study across different countries in Europe, including the UK, performed between 2004 and 2010 showed that rates of stillbirths declined by an
average of 17%. A large number of these countries did not implement the use of customized charts. Thus, in comparison by the overall decline in stillbirth rates in Europe, the difference in decline in stillbirth rates in areas with and without implementation of customized charts may be relatively small [5, 6].” (Page 15, lines 285-300).

Reviewer 3
In this paper, the authors examined the association between SGA and LGA defined by using population or customized charts and adverse outcomes at 10 years. The authors report that the direction and strength of the associations were similar when using either population or customized charts. Overall the paper is well written. Below are some comments:
Response: We thank the reviewer for the comments. Please see our response to each comment below.

Comment 1: The majority of the studies that use population based charts are sex-specific birth weight for gestational age. Are the population based charts used in this study also sex-specific? If not why did not the authors consider examining sex specific charts?
Response: We did not use sex-specific population birth weight charts, as we specifically aimed to assess the effect of customization. Customized charts take into account important physiological determinants of fetal growth, of which fetal sex is a major determinant together with several maternal characteristics. By including fetal sex in the population charts, we would underestimate the effect of customization and we therefore consider this inappropriate in our analyses. If we had included fetal sex in the population charts, we hypothesize that the differences between the classifications of size at birth according to the customized and population charts may even have been smaller. We agree with the reviewer that population charts in clinical practice often include fetal sex and we therefore addressed this issue in the Discussion Section: “In clinical practice, often sex-specific population charts are used to classify abnormal size at birth weight. Given the aim of our study to specifically assess the effect of customization by major determinants of fetal growth, we constructed a population chart which included gestational age only to enable the most optimal comparison. By including fetal sex in the population chart, we could underestimate the effect of customized charts, as fetal sex is one of the major physiological determinants of fetal growth. If we had included fetal sex in our population charts, we expect similar or even weaker differences between the associations of abnormal size at birth with the risk of long-term adverse outcomes according to customized charts and population charts.” (Page 17, lines 355-363).

Comment 2: Prevalence of adverse outcomes within each gestational age adjusted birth weight category by dividing the cases by the number of newborns in each birth weight category. Do the authors mean here the prevalence of outcomes among SGA, AGA, and LGA infants as reported in Figure 1?
Response: The reviewer is correct. We have clarified this in the Methods Section: “For categorical outcomes, we calculated prevalences of adverse outcomes among SGA, AGA and LGA newborns, by dividing the number of cases by the number of newborns in each birth weight category.” (Page 8, lines 168-170).

Comment 3: Table 3 can be presented as a supplementary table and Fig S2 as a main figure.
Response: As suggested by the reviewer, we moved Additional file 1, Figure S2 to the main manuscript and described the findings of this figure more clearly. For the detailed textual changes, please see our response at comment 5 of reviewer 3. We prefer to keep Table 3 within the main manuscript, as this table shows important differences in population characteristics for newborns classified SGA and LGA according to customized or population charts. If the Editors and Reviewers
wish, we are more than willing to move Table 3 to the supplemental material.

Comment 4: Results page 13 line 222. All the reported ORs are non-significant. The authors mentioned higher risk of childhood overweight (OR 1.24, 95% CI: 0.95-1.60), hyperlipidemia (OR 1.25; 95% CI: 0.88-1.79) and liver steatosis (OR 1.77; 95% CI 0.88-3.54) but these are non-significant ORs. Please revise this section.
Response: We have rewritten the following sentences in Results section: “They also tended to have higher risk of childhood overweight (OR 1.24 (95%CI 0.95 to 1.60)), hyperlipidaemia (OR 1.25 (95% CI 0.88 to 1.79)) and liver steatosis (OR 1.77 (95% CI 0.88 to 3.54)), but these findings did not reach statistical significance (Figure 1C-G).” (Page 13, lines 237-240).

Comment 5: The authors mention in the results section that repeating the analyses among newborns classified SGA or LGA by customized or population charts only (page 13 line 227) results in similar findings. However, in the discussion section lines 268-270, they report that infants classified by customized charts only did have higher risks...Please correct this discrepancy when reporting the results.
Response: We have rewritten the following sentences in the Results Section: “When we repeated the analyses among newborns classified SGA or LGA by customized or population charts only, largely similar findings were observed. We only observed a slightly higher risk of high blood pressure (OR 2.14 (95% CI 1.28 to 3.58) among newborns classified SGA by customized charts only compared to those classified SGA by population charts only (Figure 2).” (Page 13, lines 246- 250).

Comment 6: Discussion section page 15, line 277, please provide the reference.
Response: We provided the references as requested by the reviewer.
Discussion section: “First, current customized birth weight charts have been criticized as they might not yet capture growth potential well enough to truly differentiate between pathologically and constitutionally SGA and LGA newborns[7, 8].” (Pages 16, lines 320-323).

Comment 7: The main idea behind this study is whether one chart is more predictive of adverse outcomes compared to the other chart used. However, the approach used in this study, i.e. examining strength of associations, does not provide the full picture to suggest using one chart versus the other. Additional results need to be presented, examining ROC curve analyses to assess the predictive power of either chart for long term adverse outcomes. Please refer to this reference: Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882-90. The conclusions are likely to be similar given the small reported ORs.
Response: We agree with the reviewer. We did not perform these additional analyses as we only observed small differences in the ORs. Based on the comment of this reviewer, we performed additional analyses focused on the predictive performance of the classification of abnormal size at birth according to customized and population charts for identification of newborns at risk of long-term adverse health outcomes. We calculated Receiver Operating Characteristic curves for either charts, obtained the Area Under the Curve (AUC)s and determined corresponding sensitivity at a 90% specificity for the risk of each adverse outcome among newborns born SGA or LGA according to customized or population charts. These results were added to Additional file 1, Table S2. We added the following sentences to the manuscript:
Statistical methods section: “Finally, we assessed the predictive performance of both classifications for the prediction of the risk of long term adverse health outcomes among SGA and LGA newborns by calculating Receiver Operating Characteristic(ROC)-curves, the corresponding Area Under the Curve and sensitivity at a 90% specificity.”(Page 8, lines 172-175).
Results section: “Additional file 1, Table S2 shows AUC’s and derived sensitivities at a 90% specificity for both classifications for the risk of each long-term adverse health outcome. Both classifications had a poor to moderate ability to discriminate between those with and those without long-term adverse health outcomes with AUCs (95% CI) ranging from 0.51 (95% CI 0.48-0.54) and 0.51 (95% CI 0.48-0.54) for risk of childhood asthma diagnosis to 0.66 (95% CI 0.64-0.69) and 0.63 (95% CI 0.61-0.65) for risk of infant catch up growth for customized and population charts, respectively” (Page 13, lines 250-256).

Discussion section: “When we determined the accuracy of both classification methods for the prediction of individual risk of adverse outcomes, we observed a poor to moderate performance for both customized and population charts. This suggests that the neither classification can be used for individual prediction of the risk for long-term adverse health outcomes based on classifying size at birth. However, the apparent increased risk of long-term adverse health outcomes among the group of SGA newborns, classified using either classification, suggest that on a population level this characteristic can be used for screening or prevention strategies, especially in combination with other prognostic factors.” (Page 15-16, lines 311-318).

References: