

1 **The clinical performance and population health impact of birth weight-for-gestational age**  
2 **indices with regard to adverse neonatal outcomes in term infants**

3

4 Sid John,\*<sup>1</sup> K S Joseph,\*<sup>†1,2</sup> John Fahey,<sup>3</sup> Shiliang Liu,<sup>4</sup> Michael S Kramer<sup>5</sup>

5 \* These authors contributed equally to the work.

6 **Short title:** Performance and impact of birth weight-for-gestational age indices

7

8 1 Department of Obstetrics and Gynaecology University of British Columbia, and the Children's  
9 and Women's Hospital of British Columbia, Vancouver, British Columbia, Canada

10 2 School of Population and Public Health, University of British Columbia, Vancouver, British  
11 Columbia, Canada

12 3 Reproductive Care Program of Nova Scotia, Halifax, Nova Scotia, Canada

13 4 Centre for Surveillance and Applied Research, Public Health Agency of Canada, and the  
14 School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

15 5 Departments of Epidemiology and Occupation Health and of Pediatrics, McGill University,  
16 Montréal, Quebec, Canada

17

18 Word count: Abstract 324; Text 3,130

19 † ksjoseph@bcchr.ca

20

21 **Abbreviations**

22 AGA Appropriate for gestational age

23 LGA Large for gestational age

24 PAF Population attributable fraction

25 SGA Small for gestational age

26 SNMM Serious neonatal morbidity or neonatal mortality

27

28 **Abstract**

29 **Background**

30 Despite the recent creation of several birth weight-for-gestational age references and standards,  
31 none has proven superior. We identified birth weight-for-gestational age cut-offs, and  
32 corresponding United States population-based, Intergrowth 21<sup>st</sup> and World Health Organization  
33 centiles associated with higher risks of adverse neonatal outcomes, and evaluated their ability to  
34 predict serious neonatal morbidity and neonatal mortality (SNMM).

35 **Methods and findings**

36 The study population comprised singleton live births at 37-41 weeks' gestation in the United  
37 States, 2003-2017. Birth weight-specific SNMM, which included 5-minute Apgar score<4,  
38 neonatal seizures, assisted ventilation and neonatal death, was modeled by gestational week  
39 using penalized B-splines. We estimated the birth weights at which SNMM odds was minimized  
40 (and higher by 10%, 50% and 100%), and identified the corresponding population, Intergrowth  
41 21<sup>st</sup> and World Health Organization (WHO) centiles. We then evaluated the individual- and  
42 population-level performance of these cut-offs for predicting SNMM. The study included  
43 40,179,663 live births at 37-41 weeks' gestation and 991,486 SNMM cases. Among female  
44 singletons at 39 weeks' gestation, SNMM odds was lowest at 3,203 g birth weight (population,  
45 Intergrowth and WHO centiles 40, 52 and 46, respectively), and 10% higher at 2,835 g and 3,685  
46 g (population centiles 11<sup>th</sup> and 82<sup>nd</sup>, Intergrowth centiles 17<sup>th</sup> and 88<sup>th</sup> and WHO centiles 15<sup>th</sup> and  
47 85<sup>th</sup>). SNMM odds were 50% higher at 2,495 g and 4,224 g and 100% higher at 2,268 g and  
48 4,593 g. Birth weight cut-offs were poor predictors of SNMM. For example, the birth weight cut-  
49 off associated with 10% higher odds of SNMM among female singletons at 39 weeks' gestation  
50 resulted in a sensitivity of 12.5%, specificity of 89.4% and population attributable fraction of

51 2.1%, while the cut-off associated with 50% higher odds resulted in a sensitivity of 2.9%,  
52 specificity of 98.4% and population attributable fraction of 1.3%.

### 53 **Conclusions**

54 Birth weight-for-gestational age cut-offs and centiles perform poorly when used  
55 to predict adverse neonatal outcomes in individual infants, and the population impact associated  
56 with these cut-offs is also small.

57 **Funding** Canadian Institutes of Health Research (MOP-67125 and PJT153439).

58

## 59 **Author summary**

### 60 Why was this study done

- 61 • Despite the recent creation of several birth weight-for-gestational age references and  
62 standards, no method has proved superior for identifying small-for-gestational age  
63 (SGA), appropriate-for-gestational age (AGA) and large-for-gestational age (LGA)  
64 infants.
- 65 • For instance, infants classified as AGA by the Intergrowth Project 21<sup>st</sup> standard and SGA  
66 by national references have a higher risk of perinatal death compared with infants  
67 deemed AGA by both.

68

### 69 What did the researchers do and find?

- 70 • Our study identified the birth weights at each gestational week at which the risk of  
71 serious neonatal morbidity and neonatal mortality (SNMM) was lowest and elevated to  
72 varying degrees, and showed that the corresponding Intergrowth and WHO centiles were  
73 right-shifted compared with population centiles.
- 74 • Outcome-based birth weight and centile cutoffs performed poorly for predicting serious  
75 neonatal morbidity and neonatal mortality (SNMM) at the individual level.
- 76 • The population attributable fractions associated with these Outcome-based birth weight  
77 and centile cutoffs cut-offs were also small.
- 78 • The birth weight distributions of live births and SNMM cases (at each gestational week)  
79 overlapped substantially, showing that birth weight-for-gestational age in isolation cannot  
80 serve as an accurate predictor of adverse neonatal outcomes, irrespective of the cut-off  
81 used to identify SGA and LGA infants.

82 What do these findings mean?

- 83       • Using birth weight-for-gestational age cutoffs to identify SGA, AGA and LGA infants
- 84       does not add significantly to individual- or population-level prediction of adverse
- 85       neonatal outcomes.
- 86       • Birth weight-for-gestational age centiles are best suited for use in multivariable
- 87       prognostic functions, in conjunction with other prognostic indicators of adverse perinatal
- 88       outcomes.
- 89

## 90 **Introduction**

91 Birth weight for gestational age reflects in utero (fetal) growth, and fetuses and infants are  
92 routinely categorized as small-for-gestational age (SGA), appropriate-for-gestational age (AGA)  
93 and large-for-gestational age (LGA). Such SGA and LGA categorization at birth is often used to  
94 predict adverse neonatal outcomes at the individual level, while at the population level, high  
95 SGA and LGA rates can help to delineate disease mechanisms (e.g., abnormal fetal growth  
96 among women with hypertension or diabetes), and identify subpopulations at increased risk for  
97 adverse perinatal outcomes (e.g., women of low socioeconomic status). Nevertheless, the  
98 assessment of birth weight for gestational age and the identification of SGA and LGA infants  
99 remain contentious. Although the newly created normative/prescriptive Intergrowth 21<sup>st</sup> Project  
100 and World Health Organization (WHO) birth weight-for-gestational age standards<sup>1-4</sup> represent a  
101 conceptual departure from previous descriptive, population-based references<sup>5-11</sup> and offer several  
102 theoretical advantages, studies comparing existing references and the new standards have not  
103 succeeded in establishing a preferred method for identifying SGA and LGA infants.<sup>12-26</sup>

104

105 One study,<sup>22</sup> which contrasted the Intergrowth standard with population-based references from  
106 15 European countries, showed that infants classified as AGA by the Intergrowth standard and  
107 SGA by the 15 national references had a 2.7-fold higher risk of perinatal death (compared with  
108 infants deemed AGA by both). Conversely, infants categorized as LGA by the Intergrowth  
109 standard but AGA by the national references were at significantly reduced risk of perinatal  
110 death.<sup>22</sup> Another study<sup>25</sup> assessed three estimated fetal weight-for-gestational age and three birth  
111 weight-for-gestational-age references and standards and observed “marked variation in  
112 classification and .... similarly poor performance” among them.

113

114 We carried out a study to assess the predictive performance of categorizing singleton, term  
115 infants in the United States as SGA or LGA. Instead of identifying SGA and LGA categories  
116 based on traditional centiles, we first determined critical risk levels for serious neonatal  
117 morbidity or neonatal mortality at different values of birth weight for gestational age in this  
118 population.<sup>27</sup> These values were then used as cut-offs to i) identify the corresponding U.S.  
119 population, Intergrowth and WHO centiles; ii) group infants into SGA and LGA categories; and  
120 iii) assess the clinical (individual-level) and the epidemiologic (population-level) ability of such  
121 SGA and LGA categorization to predict adverse neonatal outcomes.

122

## 123 **Methods**

124 Our study population comprised all singleton, term, live births and infant deaths in the United  
125 States from 2003 to 2017. Data on these live births and infant deaths were obtained from the  
126 period linked birth-infant death files of the National Center for Health Statistics, which included  
127 information on all live birth and infant death registrations in the United States.<sup>28</sup> Gestational age  
128 was based on the ‘clinical estimate of gestation’ variable in these data files. Information in this  
129 data base has been validated<sup>29</sup> and previously used in epidemiologic studies, including studies  
130 that developed birth weight-for-gestational age references for the United States.<sup>6,9</sup>

131

132 We excluded stillbirths, because the birth weight of stillborn fetuses is affected by postmortem  
133 changes, and the gestational age at stillbirth typically exceeds the gestational age at fetal death by  
134 a few days and sometimes by a week or more.<sup>30,31</sup> Live births were restricted to those at 37-41  
135 weeks’ gestation, as the large numbers of live births at these gestational ages permits precise  
136 modeling of birth weight-specific adverse neonatal outcomes. We also excluded infants



137 diagnosed with congenital anomalies at birth, neonatal deaths due to homicide, accident or  
138 congenital anomaly, and live births with a missing or implausible birth weight for gestational age  
139 (e.g., birth weight <1,000 g at 37-41 weeks' gestation<sup>6</sup>) or missing information on sex or adverse  
140 neonatal outcomes.

141  
142 The primary outcome was a composite of serious neonatal morbidity or neonatal mortality  
143 (SNMM), which included one or more of the following: 5-minute Apgar score <4, neonatal  
144 seizures, need for assisted ventilation and neonatal death. Birth weight was grouped into 28-gram  
145 categories in order to address ounce and digit preference issues.<sup>32</sup> A noniterative penalized B-  
146 spline transformation<sup>33</sup> was used to fit a smooth curve though a scatter plot of birth weight (at  
147 each gestational week) and SNMM rate. The cubic splines had 100 evenly spaced interior knots,  
148 3 evenly spaced exterior knots and a difference matrix of order 3 that penalized smoothness and  
149 selected the non-negative smoothing parameter automatically based on minimizing the  
150 Schwarz's Bayesian criterion.

151  
152 The gestational age-specific (28 g) birth weight category at which the predicted SNMM odds  
153 was minimized was identified separately for female and male infants, and this served as the  
154 SNMM reference odds used to calculate SNMM odds ratios for other birth weight categories.<sup>27</sup>  
155 The gestational week-specific birth weight centiles of the study population (referred to  
156 henceforth as population centiles), and the corresponding centiles of the Intergrowth and WHO  
157 newborn growth standards<sup>3,4</sup> that corresponded to the mid-point of the 28 g birth weight category  
158 were then ascertained. The gestational age-specific birth weight categories at which SNMM odds  
159 were higher by 10%, 50% and 100% (relative to the minimized odds) were also determined

160 along with the corresponding population, Intergrowth and WHO centiles.

161

162 We then identified SGA and LGA infants using birth weight cutoffs/centiles at which SNMM

163 odds were increased by 10%, 50% and 100%, and evaluated the performance of such

164 categorization for predicting SNMM. Birth weight cut-offs for categorizing SGA and LGA

165 infants were based on the upper (for SGA categorization) or the lower (for LGA categorization)

166 boundary of the 28 g birth weight category at which SNMM odds were increased by 10%, 50%

167 and 100%. Sensitivity, specificity and positive and negative likelihood ratios were calculated.

168 The pretest and posttest probabilities of SNMM were estimated to determine the information

169 gained (marginal informativeness<sup>34</sup>) in predicting SNMM in individual infants by identifying

170 them as SGA or LGA. The epidemiologic performance of such SGA and LGA categorization

171 was assessed by estimating measures of association (odds ratios, rate ratios and rate differences)

172 between SGA and LGA status and SNMM and also population attributable fractions (PAF)<sup>35</sup>

173 (which quantify the proportion of SNMM cases that would be prevented if SGA or LGA was

174 eliminated from the population).

175

176 All analyses were carried out using SAS 9.2 (Cary, NC). Ethics approval was not sought, as the

177 data were anonymized and are publicly available.

178

## 179 **Results**

180 The study population included 40,179,663 singleton live births at 37 to 41 weeks' gestation and

181 991,486 cases of SNMM. The overall rate of SNMM was 24.7 per 1,000 livebirths, with female

182 infants having a lower SNMM rate than male infants (22.8 vs 26.5 per 1,000 live births,

183 respectively). SNMM rates were highest at 37 weeks' gestation (34.4 per 1,000 live births),  
184 substantially lower at 38, 39 and 40 weeks (24.2, 21.8 and 24.0 per 1,000 live births,  
185 respectively) and 30.7 per 1,000 live births at 41 weeks' gestation. The frequencies of SNMM  
186 components were: assisted ventilation, 19.6 per,1000 live births; 5-minute Apgar<4, 1.8 per  
187 1,000 live births; seizures, 0.26 per 1,000 live births; and neonatal deaths, 0.28 per 1,000 live  
188 births. Rates of 5-minute Apgar<4, seizures and neonatal deaths among infants requiring assisted  
189 ventilation were 40.7, 6.2 and 4.7 per 1,000 live births, respectively.

190

191 Figure 1 shows the observed and modeled birth weight-specific rates of SNMM among  
192 singleton males and singleton females born at 39 weeks' gestation (Appendix Figures 1-5 show  
193 rates at 37-41 weeks). Among singleton females at 39 weeks' gestation, the odds of SNMM was  
194 lowest at 3,203 g, while the odds of SNMM was 10% higher at both 2,835 g and 3,685 g, 50%  
195 higher at 2,495 g and 4,224 g and 100% higher at 2,268 g and 4,593 g. The corresponding  
196 SNMM odds in males was lowest at 3,374 g, 10% higher at both 2,920 g and 3,884 g, 50%  
197 higher at 2,580 g and 4,479 g and 100% higher at 2,353 g and 4,791 g. The SNMM frequency at  
198 which the SNMM odds was minimized was lower among singleton females compared with  
199 singletons males (e.g., 18.7 vs 21.9 per 1,000 live births at 39 weeks' gestation). Appendix  
200 Tables 1 and 2 show the birth weights at which SNMM odds were lowest (and 10%, 50% and  
201 100% higher) among female and male singletons at 37-41 weeks' gestation.

202

203 Among singleton females at 39 weeks' gestation, the SNMM odds was lowest at population  
204 centile 40, Intergrowth centile 52 and WHO centile 46, and 10% higher at the 11<sup>th</sup> and 82<sup>nd</sup>  
205 population centiles, the 17<sup>th</sup> and 88<sup>th</sup> Intergrowth centiles and the 15<sup>th</sup> and 85<sup>th</sup> WHO centiles

206 (Figure 2). Similarly, Intergrowth centiles associated with minimal and elevated odds of SNMM  
207 among singleton males at 39 weeks' gestation were markedly right-shifted, while WHO centiles  
208 were moderately right-shifted compared with population centiles (Figure 2).

209

210 Table 1 shows the ability of different birth weight cut-offs (based on elevated SNMM odds) for  
211 categorizing SGA and LGA infants in order to predict SNMM at the individual level. Using the  
212 birth weight cut-off  $\leq 2,849$  g (upper boundary of the 28 g birth weight category at which the  
213 SNMM odds was 10% higher) for identifying SGA infants among female singletons at 39  
214 weeks' gestation resulted in a sensitivity of 12.5%, specificity of 89.4% and modest likelihood  
215 ratios (positive and negative likelihood ratios of 1.18 and 0.98). The "pretest" probability of  
216 SNMM among singleton females at 39 weeks' gestation was 20.2 per 1,000 live births, and the  
217 posttest probability of SNMM increased to 23.8 per 1,000 live births given a positive "test"  
218 (using the  $\leq 2,849$  g birth weight cut-off for defining SGA infants). Conversely, the pretest  
219 probability of not having SNMM increased only slightly from 979.8 to 980.3 per 1,000 live  
220 births if the test was negative. Labeling infants using cut-offs at which the SNMM odds was  
221 elevated by 50% or 100% as SGA or LGA increased the specificity and positive predictive value  
222 but at the cost of reduced sensitivity, with likelihood ratios remaining modest (Table 1). For  
223 example, the birth weight cut-off at which SNMM odds were elevated by 50% among female  
224 singletons at 39 weeks resulted in a sensitivity of 2.9%, a specificity of 98.4%, and positive and  
225 negative likelihood ratios of 1.82 and 0.99, respectively. Similar findings were obtained among  
226 males (Table 1) and among males and females at other gestational ages (Appendix Tables 3-7).

227

228 Table 2 shows the population-level performance of such SGA and LGA categorization. Using

229 the birth weight cut-off  $\leq 2,849$  g for identifying SGA infants among female singletons at 39  
230 weeks' gestation resulted in a rate difference of 2.8 per 1000 live births and a PAF of 2.1%.  
231 Using cut-offs at which the SNMM odds was 50% or 100% higher increased the rate difference  
232 but decreased the PAF (Table 2). Individual- and population-level performance of SGA and  
233 LGA categorization using SNMM components instead of composite SNMM yielded similar  
234 results, although likelihood ratios and PAFs for neonatal death were larger than those for the  
235 other components of SNMM. (Appendix Tables 8-15).

236

237 Figure 3 shows the birth weight distribution of live births and SNMM cases for singletons at 39  
238 weeks' gestation. The distributions overlapped substantially, with most live births and SNMM  
239 cases occupying the central part of the birth weight distribution. For example, among male  
240 singletons at 39 weeks' gestation, 73.5% of live births (5,736,492 of 7,800,660, Table 1) and  
241 69.0% of SNMM cases (125,545 of 181,940, Table 1) occurred in the central, lowest risk birth  
242 weight category, namely, between 2,935 and 3,869 g. Similarly, substantial overlap was  
243 observed in the birth weight distributions of live births and components of composite SNMM  
244 (Appendix Figures 6-9). For example, 59.2% of neonatal deaths among female singletons and  
245 66.7% of neonatal deaths among male singletons at 39 weeks' gestation occurred in the central,  
246 low-risk part of the birth weight distribution (Appendix Figure 9 and Table 11).

247

## 248 **Discussion**

249 We identified the birth weights at each term gestational week at which SNMM risk was lowest  
250 and elevated to varying degrees, and the corresponding centiles of the U.S. study population, the  
251 Intergrowth standard and the WHO standard. Intergrowth and WHO centiles at which SNMM

252 risk was lowest and elevated were right-shifted compared with U.S. population centiles. For  
253 instance, SNMM risk was lowest at 3,203 g birth weight among singleton females at 39 weeks'  
254 gestation, which corresponded to population centile 40, Intergrowth centile 52, and WHO centile  
255 46. Evaluation of outcome-based birth weight and centile cutoffs for identifying SGA and LGA  
256 infants showed a poor ability to predict adverse neonatal outcomes at the individual level. The  
257 population-level performance, i.e., the PAF associated with SGA and LGA categories, was also  
258 modest. The birth weight distributions of live births, SNMM cases and neonatal deaths (at each  
259 gestational week) overlapped substantially, showing that birth weight for gestational age used in  
260 isolation cannot serve as an accurate individual- or population-level predictor of adverse neonatal  
261 outcomes, irrespective of the birthweight or centile cut-offs used to identify SGA and LGA  
262 infants.

263  
264 Populations used to create standards include pregnancies based on strict demographic, clinical,  
265 social, educational and other criteria,<sup>1-4</sup> while reference populations include fetuses from higher  
266 risk pregnancies and those with congenital malformations. The many overt and subtle differences  
267 that characterize standard vs reference populations make it challenging to anticipate how  
268 reference centiles will map on to standard population centiles. Studies showing differences  
269 between standards and references with regard to the prediction of adverse perinatal outcomes  
270 have typically use traditional centiles (such as the 10<sup>th</sup> and 90<sup>th</sup> centile) for categorizing SGA and  
271 LGA infants. These centiles tend to favour references over standards, because the centiles were  
272 developed for use with references. Differences between standards and references  
273 notwithstanding, our study illustrates that it is possible to achieve a similar categorization of

274 SGA and LGA infants using references or standards if chart-specific, empirically identified  
275 centiles are used.

276

277 The diagnosis of SGA or LGA in a fetus implies a high probability of restricted or excessive  
278 growth. Theoretically, such diagnoses could be used as inputs for improving obstetric and  
279 neonatal management in conjunction with other markers and pregnancy complications. However,  
280 as previous studies have shown<sup>14,15,23-26,36</sup> and our study confirms, SGA or LGA diagnosis, when  
281 used in isolation without other predictors, performs poorly for predicting SNMM; in other words,  
282 the probability of SNMM (or neonatal death) is not materially changed by such a diagnosis. In  
283 our study, the expected risk of SNMM changed marginally from 20.2 per 1,000 live births to  
284 23.8, 36.1 and 54.4 per 1,000 live births using a mild (10% increase in odds), moderate (50%) or  
285 extreme (100%) definition of SGA among female singletons at 39 weeks' gestation (Table 1).

286

287 The poor performance of SGA and LGA diagnoses in predicting SNMM and neonatal death  
288 weakens arguments regarding the superiority of references vs standards.<sup>13,17,19-24,37</sup> For instance,  
289 diagnosis of SGA among term births using national references was associated with positive and  
290 negative likelihood ratios for perinatal death of 3.2 and 0.77, respectively; this changed the  
291 pretest probability of perinatal death to an inconsequential extent (from 1.7 per 1,000 total births  
292 to a posttest probability of 5.4 per 1,000 total births).<sup>22</sup> Similarly, the sensitivity, specificity,  
293 likelihood ratios and PAFs for neonatal morbidity associated with an SGA diagnosis based on  
294 customized charts was of limited utility at both the individual and population level (sensitivity  
295 and PAF for neonatal seizures 14.1% and 4.3%, respectively; sensitivity and PAF for glucose  
296 instability 22.9% and 14.0%, respectively).<sup>24</sup>

297 Treating SGA and LGA as adverse outcomes in themselves is fraught with problems,  
298 irrespective of the reference/standard and birth weight-for-gestational age centiles used for such  
299 diagnosis. Diagnosing and treating ‘diseases’ based on a categorization of continuous biological  
300 measures, which represent anomalous states associated with illness (as opposed to identifying a  
301 specific underlying somatic anomaly<sup>38</sup>), has gained favour in medicine. A well-known example  
302 is the diagnosis and management of hypertension among individuals with elevated systolic or  
303 diastolic blood pressure. However, the success of this strategy is dependent on the distributional  
304 relationships between the continuous measure and illness complications. Unfortunately, birth  
305 weight for gestational age does not satisfy the requirements for such a biological measure, as a  
306 large proportion of live births, SNMM cases, and neonatal deaths occur at the central, low-risk  
307 part of the birth weight-for-gestational age distribution – not at the extremes (Figure 3). This is  
308 unlike the case of hypertension, where a majority of cerebrovascular and cardiovascular events  
309 occurs among hypertensive individuals.<sup>39,40</sup>

310  
311 Estimated fetal weight for gestational age and birth weight for gestational age are ideally treated  
312 as determinants that could contribute to a multivariable prognostic function for predicting the  
313 probability of specific adverse perinatal and neonatal events. Such noncausal functions would  
314 need to include a range of physiologic (e.g., infant sex, maternal race, and multi-fetal pregnancy)  
315 and non-physiologic factors (e.g., advanced maternal age, maternal hypertension, diabetes, and  
316 smoking status), in addition to biochemical and ultrasound indicators of risk, to be clinically  
317 useful for predicting adverse outcomes. Representation of birth weight for gestational age in such  
318 models could take the form of (uncategorized) centiles based on any reference or standard,  
319 although use of a universal standard may be preferable for facilitating communication. Repeated



320 measures of birth weight for gestational age that quantify growth trajectory could also contribute  
321 valuable information to such a prediction equation. Although current clinical practice does  
322 consider birth weight for gestational age as a predictor of adverse perinatal outcomes, formal  
323 inclusion of this index and its longitudinal trajectory as one of several determinants in a  
324 multivariable prediction function should greatly enhance clinical utility.

325

326 Quantifying SGA and LGA frequencies in non-normal (vs normal) populations (such as women  
327 with hypertension or diabetes) can provide insights into disease mechanisms by highlighting  
328 patterns of restricted or excessive growth. However, such relative comparisons between normal  
329 and high-risk groups can be based on birth weight-for-gestational age centile cutoffs from any  
330 reference or standard. This also applies to contrasts between population subgroups (e.g., by  
331 socioeconomic status), although our results strongly suggest that the SGA or LGA frequency in  
332 subpopulations does not adequately predict overall neonatal morbidity and mortality, as the  
333 PAFs associated with these indices are small.

334

335 Strengths of our study included an evaluation of birth weight-for-gestational age references and  
336 standards based on outcome-based criteria, rather than arbitrary centiles such as the 10<sup>th</sup> and 90<sup>th</sup>  
337 centiles. Our use of a composite outcome including both serious neonatal morbidity and neonatal  
338 mortality is another important strength, because neonatal mortality is a relatively rare adverse  
339 outcome. The large numbers of live births and SNMM cases in our study is a third strength. One  
340 limitation of our study is the restriction to term birth; our findings may not apply to preterm  
341 births, whose risk of SNMM may be mediated by distinct mechanisms such as immaturity.

342 Another study limitation is the potential data quality weakness inherent in large perinatal  
343 databases.

344

345 Using birth weight-for-gestational-age cutoffs to identify SGA, AGA, and LGA infants does not  
346 add substantially to individual- or population-level prediction of adverse neonatal outcomes.

347 Birth weight-for-gestational-age centiles are best suited for use in multivariable prognostic  
348 functions, and could help improve obstetric and neonatal care in conjunction with other  
349 prognostic indicators of adverse perinatal outcomes.

350

351 **Acknowledgements** KSJ's work is supported by an Investigator award from the BC Children's  
352 Hospital Research Institute.

353 **Data sharing** Data used in this study are available from the National Center for Health Statistics,  
354 Centers for Disease Control and Prevention ([https://www.cdc.gov/nchs/data\\_access/  
355 vitalstatsonline.htm](https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm))

356

## 357 **References**

- 358 1. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Cheikh Ismail L, Lambert A, et  
359 al.; International Fetal and Newborn Growth Consortium for the 21st Century  
360 (INTERGROWTH-21st). Inter-national standards for fetal growth based on serial  
361 ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-  
362 21st Project. *Lancet* 2014;384:869-79.
- 363 2. Villar J, Papageorghiou AT, Pang R, Ohuma EO, Cheikh Ismail L, Barros FC, et al.;  
364 International Fetal and Newborn Growth Consortium for the 21st Century  
365 (INTERGROWTH-21st). The likeness of fetal growth and newborn size across non-  
366 isolated populations in the INTERGROWTH-21st Project: the Fetal Growth Longitudinal  
367 Study and Newborn Cross-Sectional Study. *Lancet Diabetes Endocrinol* 2014;2:781-92.
- 368 3. Villar J, Cheikh Ismail L, Victora CG, Ohuma EO, Bertino E, Altman DG, et al.;  
369 International Fetal and Newborn Growth Consortium for the 21st Century  
370 (INTERGROWTH-21st). International standards for newborn weight, length, and head  
371 circumference by gestational age and sex: the Newborn Cross-Sectional Study of the  
372 INTERGROWTH-21st Project. *Lancet* 2014;384:857-68.
- 373 4. Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Neerup Jensen L, et al. The  
374 World Health Organization Fetal Growth Charts: A Multinational Longitudinal Study of  
375 Ultrasound Biometric Measurements and Estimated Fetal Weight. *PLoS Med* 2017;14:  
376 e1002220.
- 377 5. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth  
378 charts. *Lancet* 1992;339:283-7.
- 379 6. Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. A United States national  
380 reference for fetal growth. *Obstet Gynecol* 1996;87:163-8.
- 381 7. Gardosi J. Customized growth curves. *Clin Obstet Gynecol* 1997;40:715-22.
- 382 8. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, et al. A new  
383 and improved population-based reference for birth weight for gestational age. *Pediatrics*  
384 2001;108:e3.
- 385 9. Duryea EL, Hawkins JS, McIntire DD, Casey BM, Leveno KJ. A revised birth weight  
386 reference for the United States. *Obstet Gynecol* 2014;124:16-22.
- 387 10. Norris T, Seaton SE, Manktelow BN, Baker PN, Kurinczuk JJ, Field D, et al. Updated  
388 birth weight centiles for England and Wales. *Arch Dis Child Fetal Neonatal Ed* 2018;  
389 103:F577-F582.
- 390 11. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards:  
391 Growth velocity based on weight, length and head circumference: Methods and  
392 development. Geneva: World Health Organization, 2009.
- 393 12. Liu S, Metcalfe A, León JA, Sauve R, Kramer MS, Joseph KS; Canadian Perinatal  
394 Surveillance System (Public Health Agency of Canada). Evaluation of the  
395 INTERGROWTH-21st project newborn standard for use in Canada. *PLoS One* 2017;  
396 12:e0172910.
- 397 13. Francis A, Hugh O, Gardosi J. Customized vs INTERGROWTH-21<sup>st</sup> standards for the  
398 assessment of birthweight and stillbirth risk at term. *Am J Obstet Gynecol* 2018;218:  
399 S692-S699.
- 400

- 401 14. Hua X, Shen M, Reddy UM, Louis GB, Souza JP, Gülmezoglu AM, Zhang  
402 J. Comparison of the INTERGROWTH-21st, National Institute of Child Health and  
403 Human Development, and WHO fetal growth standards. *Int J Gynaecol Obstet* 2018;  
404 143:156-163.
- 405 15. Boghossian NS, Geraci M, Edwards EM, Horbar JD. Neonatal and fetal growth charts to  
406 identify preterm infants <30 weeks gestation at risk of adverse outcomes. *Am J Obstet*  
407 *Gynecol* 2018;219:195.e1-195.e14.
- 408 16. Lavin T, Nedkoff L, Preen D, Theron G, Pattinson RC. INTERGROWTH-21st v. local  
409 South African growth standards (Theron-Thompson) for identification of small-for-  
410 gestational-age fetuses in stillbirths: A closer look at variation across pregnancy. *S Afr*  
411 *Med J* 2019;109:519-525.
- 412 17. Pritchard NL, Hiscock RJ, Lockie E, Permezel M, McGauren MFG, Kennedy AL, et al.  
413 Identification of the optimal growth charts for use in a preterm population: An Australian  
414 state-wide retrospective cohort study. *PLoS Med* 2019;16:e1002923.
- 415 18. Savirón-Cornudella R, Esteban LM, Tajada-Duaso M, Castán-Mateo S, Dieste-Pérez P,  
416 Cotaina-Gracia L, et al. Detection of adverse perinatal outcomes at term delivery  
417 using ultrasound estimated percentile weight at 35 weeks of gestation: Comparison of  
418 five fetal growth standards. *Fetal Diagn Ther* 2020;47:104-114.
- 419 19. Kabiri D, Romero R, Gudicha DW, Hernandez-Andrade E, Pacora P, Benshalom-Tirosh  
420 N, et al. Prediction of adverse perinatal outcome by fetal biometry: comparison of  
421 customized and population-based standards. *Ultrasound Obstet Gynecol* 2020;55:177-  
422 188.
- 423 20. Pritchard N, Lindquist A, Siqueira IDA, Walker SP, Permezel M. INTERGROWTH-21st  
424 compared with GROW customized centiles in the detection of adverse perinatal  
425 outcomes at term. *J Matern Fetal Neonatal Med* 2020;33:961-966.
- 426 21. Hirsch L, Lipworth H, Kingdom J, Barrett J, Melamed N. Identification of the optimal  
427 growth chart and threshold for the prediction of antepartum stillbirth. *Arch Gynecol*  
428 *Obstet* 2021 Feb;303(2):381-390.
- 429 22. Hocquette A, Durox M, Wood R, Klungsøyr K, Szamotulska K, Berrut S, et al.  
430 International versus national growth charts for identifying small and large-for-gestational  
431 age newborns: A population-based study in 15 European countries. *Lancet Reg Health*  
432 *Eur* 2021;8:100167.
- 433 23. Monier I, Ego A, Benachi A, Hocquette A, Blondel B, Goffinet F, et al. Comparison of  
434 the performance of estimated fetal weight charts for the detection of small- and large-for-  
435 gestational age newborns with adverse outcomes: a French population-based study.  
436 *BJOG*. 2021 Nov 19. doi: 10.1111/1471-0528.17021.
- 437 24. Fay E, Hugh O, Francis A, Katz R, Sitcov K, Souter V, Gardosi J. Customized GROW vs  
438 INTERGROWTH-21<sup>st</sup> birthweight standards to identify small for gestational age  
439 associated perinatal outcomes at term. *Am J Obstet Gynecol MFM* 2021 Dec 4:100545.
- 440 25. Choi SKY, Gordon A, Hilder L, Henry A, Hyett JA, Brew BK, et al. Performance of six  
441 birth-weight and estimated-fetal-weight standards for predicting adverse perinatal  
442 outcome: a 10-year nationwide population-based study. *Ultrasound Obstet Gynecol* 2021;  
443 58:264-277.
- 444 26. Liauw J, Mayer C, Albert A, Fernandez A, Hutcheon JA. Which chart and which cut-  
445 point: deciding on the INTERGROWTH, World Health Organization, or Hadlock fetal  
446 growth chart. *BMC Pregnancy Childbirth* 2022;22:25.

- 447 27. Joseph KS, Fahey J, Platt RW, Liston RM, Lee SK, Sauve R, et al. An outcome-based  
448 approach for the creation of fetal growth standards: do singletons and twins need separate  
449 standards? *Am J Epidemiol* 2009;169:616-24.
- 450 28. CDC/National Center for Health Statistics. Vital Statistics Online Data Portal 2018  
451 [updated January 5, 2022]. Available from: [https://www.cdc.gov/nchs/data\\_access/  
452 vitalstatsonline.htm](https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm)
- 453 29. Martin JA, Wilson EC, Osterman MJ, Saadi EW, Sutton SR, Hamilton BE. Assessing the  
454 quality of medical and health data from the 2003 birth certificate revision: results from  
455 two states. *Natl Vital Stat Rep* 2013; 62:1–19.
- 456 30. Yudkin PL, Wood L, Redman CWG. Risk of unexplained stillbirth at different  
457 gestational ages. *Lancet*. 1987;1(8543): 1192–1194.
- 458 31. Hovatta O, Lipasti A, Rapola J, Karjalainen O. Causes of stillbirth: a clinicopathological  
459 study of 243 patients. *Br J Obstet Gynaecol* 1983;90(8):691–696.
- 460 32. Joseph KS. A compelling symmetry: The extended fetuses-at-risk perspective on modal,  
461 optimal and relative birthweight and gestational age. *PLoS One* 2020;15:e0238673.
- 462 33. Eilers PHC, Marx BD. Flexible Smoothing with B-Splines and Penalties. *Statistical  
463 Science* 1996;11:89–121.
- 464 34. Miettinen OS. The modern scientific physician: 4. The useful property of a diagnostic.  
465 *CMAJ* 2001;165:910-1.
- 466 35. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions.  
467 *Am J Public Health* 1998;88:15-9.
- 468 36. Wilcox AJ, Cortese M, McConaughy DR, Moster D, Basso O. The limits of small-for-  
469 gestational-age as a high-risk category. *Eur J Epidemiol* 2021;36:985-991.
- 470 37. Urquia ML, Berger H, Ray JG, Canadian Curves Consortium. Risk of adverse outcomes  
471 among infants of immigrant women according to birth-weight curves tailored to maternal  
472 world region of origin. *CMAJ* 2015;187:E32-E40.
- 473 38. Miettinen OS. *Medicine as a scholarly field: An introduction*. Springer, New York. 2015.
- 474 39. Li C, Engström G, Hedblad B, Berglund G, Janzon L. Risk factors for stroke in subjects  
475 with normal blood pressure: a prospective cohort study. *Stroke* 2005;36:234-8.
- 476 40. He J, Gu D, Chen J, Wu X, Kelly TN, Huang JF, et al. Premature deaths attributable to  
477 blood pressure in China: a prospective cohort study. *Lancet* 2009;374(9703):1765-72.  
478  
479

Table 1. Clinical performance of birth weight cut-offs/categories/centiles for identifying small- and large-for-gestational infants and associated sensitivity (Se), specificity (Sp), likelihood ratios (LR+ and LR-), and positive and negative predictive values (PPV and NPV) for composite serious neonatal morbidity and neonatal mortality (SNMM), singletons at 39 weeks' gestation, United States.

Birth weight cut-offs, categories and corresponding centiles					No. of live births†	SNMM		Se†	Sp†	LR+†	LR-†	PPV†	NPV†
Cut-offs (g)*	Categories (g)*	Population centile	Intergrowth centile	WHO centile		Number	Rate/1,000						
Singleton females at 39 weeks													
2,268	≤2,282	0.4	0.8	1.0	29,028	1,580	54.4	1.0	99.6	2.80	0.99	54.4	980.0
2,495	2,283-2,509	1.6	3.3	3.1	93,191	2,830	30.4	2.9	98.4	1.82	0.99	36.1	980.1
2,835	2,510-2,849	10.6	17.3	15.2	684,441	14,749	21.5	12.5	89.4	1.18	0.98	23.8	980.3
3,203	2,850-3,670	39.5	51.7	46.1	5,254,261	98,291	18.7	-	-	-	-	-	-
3,685	3,671-4,209	81.7	88.4	84.8	1,328,715	29,086	21.9	23.1	80.0	1.16	0.96	23.3	980.6
4,224	4,210-4,577	97.8	99.0	99.8	154,974	4,666	30.1	4.1	97.5	1.65	0.98	32.8	980.2
4,593	≥4,578	99.6	99.9	99.9	36,537	1,625	44.5	1.1	99.5	2.26	0.99	44.5	980.0
All	All	-	-	-	7,581,147	152,827	20.2	-	-	-	-	-	-
Singleton males at 39 weeks													
2,353	≤2,367	0.4	1.0	0.9	33,423	1,981	59.3	1.1	99.6	2.64	0.99	59.3	976.8
2,580	2,368-2,594	1.6	3.5	1.8	89,499	3,113	34.8	2.8	98.5	1.81	0.99	41.4	977.0
2,920	2,595-2,934	9.5	16.4	12.0	619,727	15,770	25.4	11.5	90.5	1.21	0.98	28.1	977.2
3,374	2,935-3,869	43.6	56.9	45.4	5,736,492	125,545	21.9	-	-	-	-	-	-
3,884	3,870-4,464	84.5	91.1	85.8	1,191,010	30,409	25.5	19.5	83.1	1.16	0.97	26.9	977.4
4,479	4,465-4,776	98.5	99.4	99.9	98,240	3,409	34.7	2.8	98.4	1.71	0.99	39.2	976.9
4,791	≥4,777	99.6	99.9	99.9	32,269	1,713	53.1	0.9	99.6	2.35	0.99	53.1	976.8
All	All	-	-	-	7,800,660	181,940	23.3	-	-	-	-	-	-

\* Birth weight cut-offs identify the mid-point of 28 g subcategories at which odds of SNMM was minimized and elevated by 10%, 50% and 100%. Centiles correspond to these birth weight cut-offs.

† Small- and large-for-gestational age (SGA and LGA) defined using the upper or lower bound of 28 g birth weight categories. Sensitivity of using birth weight ≤2,282 g for defining SGA =  $1580 \times 100 / 152,827 = 1.0\%$ ; Sensitivity of using birth weight ≤2,849 g for defining SGA =  $(1580 + 2,830 + 14,749) \times 100 / 152,827 = 12.5\%$ , etc.

Table 2. Epidemiologic (population-level) performance of birth weight cut-offs/categories for identifying small- and large-for-gestational infants and associated odds ratios, rate ratios, risk differences (per 1,000 live births) and population attributable fractions (PAF) for composite serious neonatal morbidity and neonatal mortality (SNMM), singletons at 39 weeks' gestation, United States.

Birth weight		No. of live births <sup>†</sup>	SNMM		Odds ratio	Rate ratio	Rate difference/1000	Population attributable fraction (%)
Cutoffs (g)*	Categories (g)*		Number	Rate/1,000				
Singleton females at 39 weeks								
2,268	≤2,282	29,028	1,580	54.4	3.0	2.9	35.7	0.7
2,495	2,283-2,509	93,191	2,830	30.4	1.6	1.6	11.7	1.3
2,835	2,510-2,849	684,441	14,749	21.5	1.2	1.1	2.8	2.1
3,203	2,850-3,670	5,254,261	98,291	18.7	1.0	1.0	0.0	0.0
3,685	3,671-4,209	1,328,715	29,086	21.9	1.2	1.2	3.2	3.9
4,224	4,210-4,577	154,974	4,666	30.1	1.6	1.6	11.4	1.6
4,593	≥4,578	36,537	1,625	44.5	2.4	2.4	25.8	0.7
All	All	7,581,147	152,827	20.2	-	-	-	-
Singleton males at 39 weeks								
2,353	≤2,367	33,423	1,981	59.3	2.8	2.7	37.4	0.8
2,580	2,368-2,594	89,499	3,113	34.8	1.6	1.6	12.9	1.2
2,920	2,595-2,934	619,727	15,770	25.4	1.2	1.2	3.5	2.2
3,374	2,935-3,869	5,736,492	125,545	21.9	1.0	1.0	0.0	0.0
3,884	3,870-4,464	1,191,010	30,409	25.5	1.2	1.2	3.6	3.1
4,479	4,465-4,776	98,240	3,409	34.7	1.6	1.6	12.8	1.2
4,791	≥4,777	32,269	1,713	53.1	2.5	2.4	31.2	0.5
All	All	7,800,660	181,940	23.3	-	-	-	-

\* Birth weight cut-offs identify the mid-point of 28 g subcategories at which odds of SNMM were minimized and elevated by 10%, 50% and 100%. Small- and large-for-gestational age (SGA and LGA) defined using the upper or lower bound of 28 g birth weight categories.

† Measures of association and impact assessing the performance of birth weight cut-offs used birth weight 2,850 g to 3,670 g as the reference category e.g., odds ratio associated with using the birth weight ≤2,282 g for defining SGA =  $(1,580 * (5,254,261 - 98,291)) / ((29,028 - 1,580) * 98,291)$ .

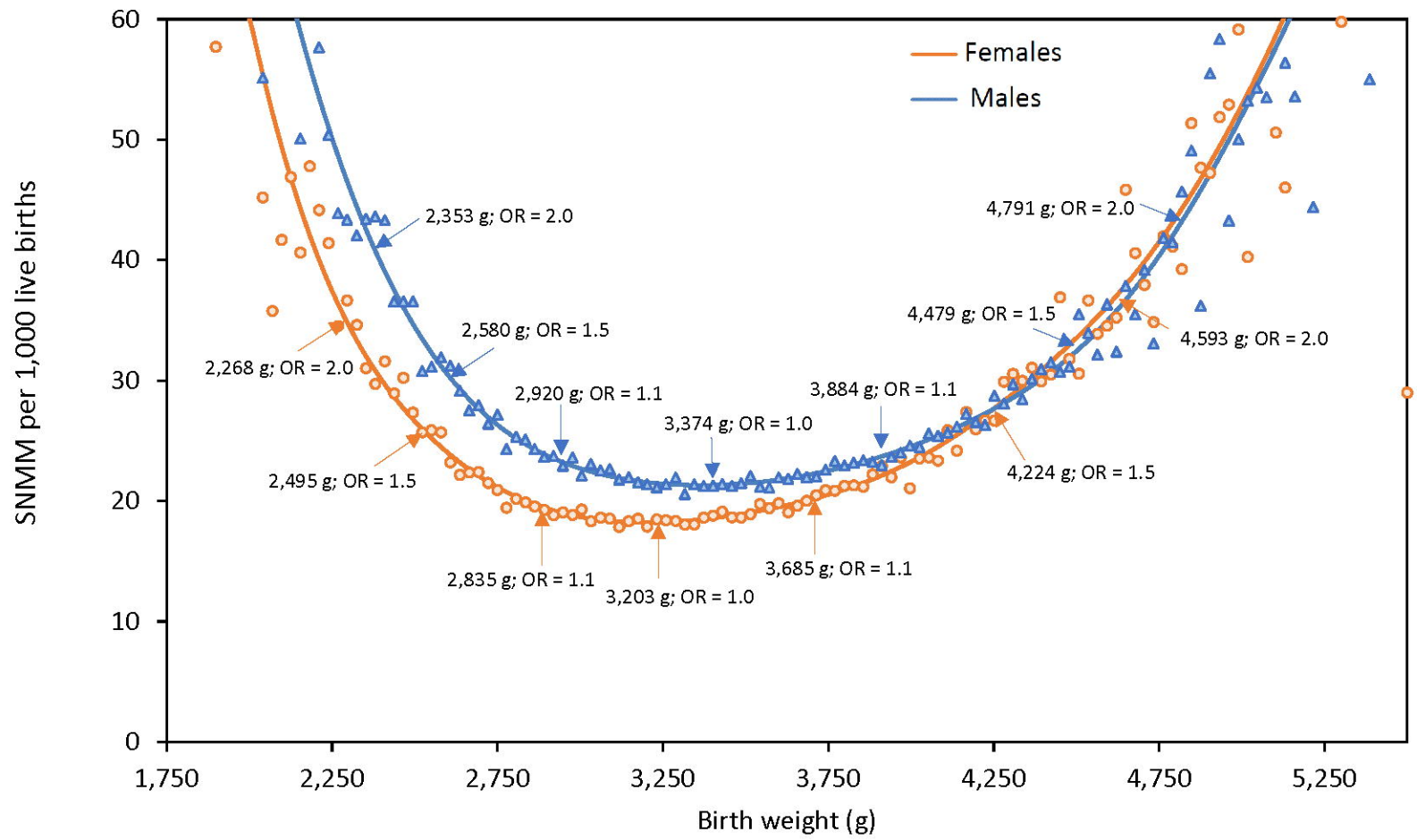


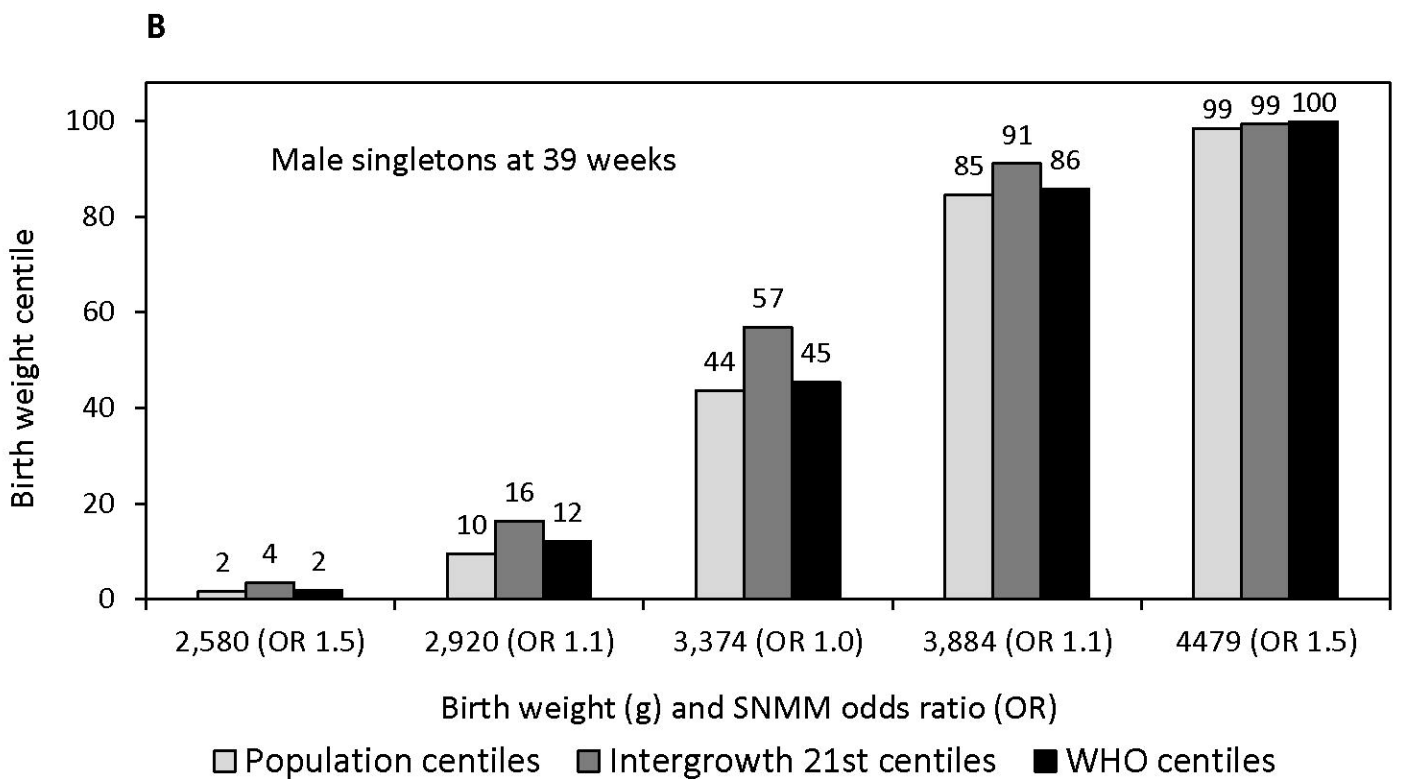
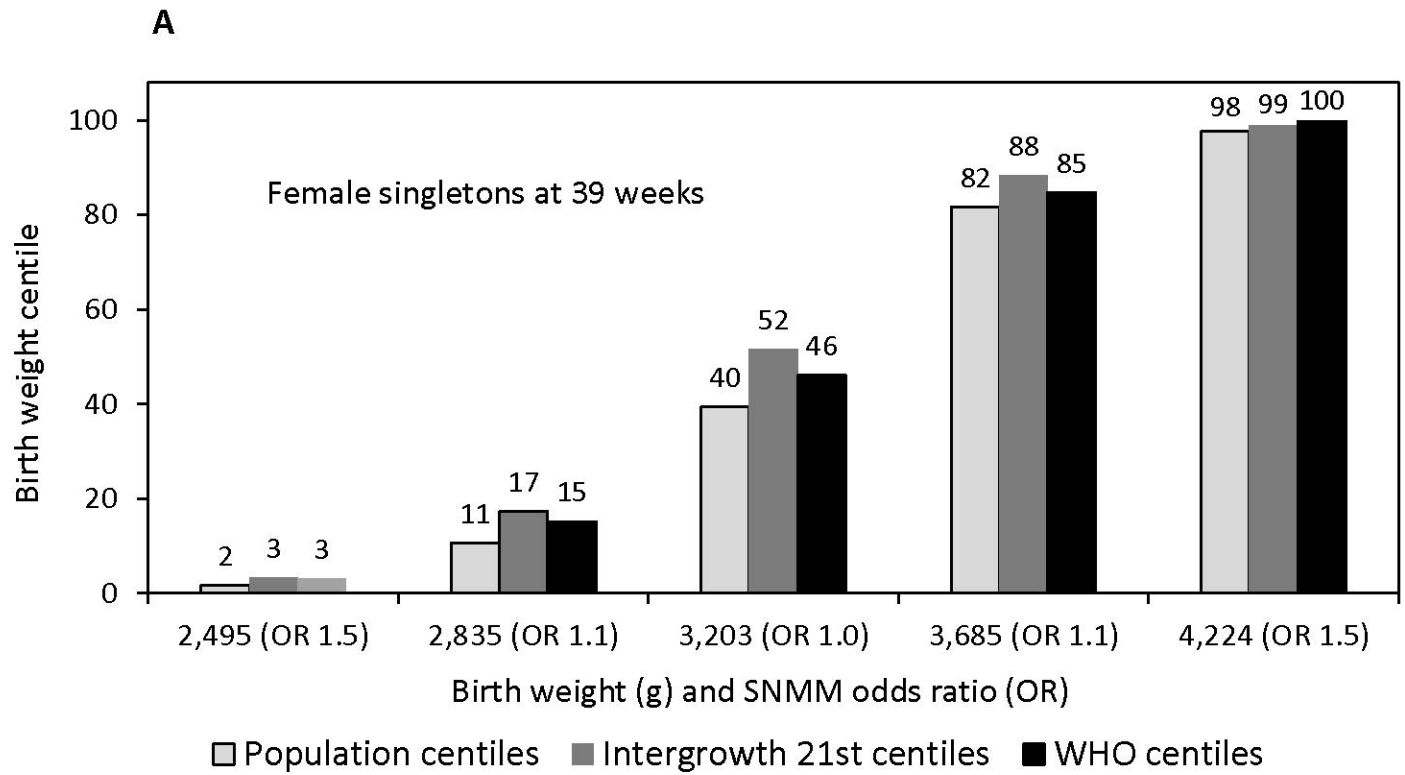
1 **Figure Legends**

2 Figure 1. Observed and modeled birth weight-specific rates of serious neonatal morbidity and  
3 neonatal mortality (SNMM) among singleton females and males at 39 weeks' gestation, United  
4 States 2003-2017.

5 Figure 2. Birth weights and centiles of the study population, Intergrowth 21<sup>st</sup> standard and World  
6 Health Organization (WHO) standard at which the odds of composite serious neonatal morbidity  
7 and neonatal mortality (SNMM) were lowest (odds ratio 1.0) and elevated by 10% and 50%  
8 (odds ratios 1.1 and 1.5), singleton females (Panel A) and males (Panel B) at 39 weeks'  
9 gestation, United States 2003-2017.

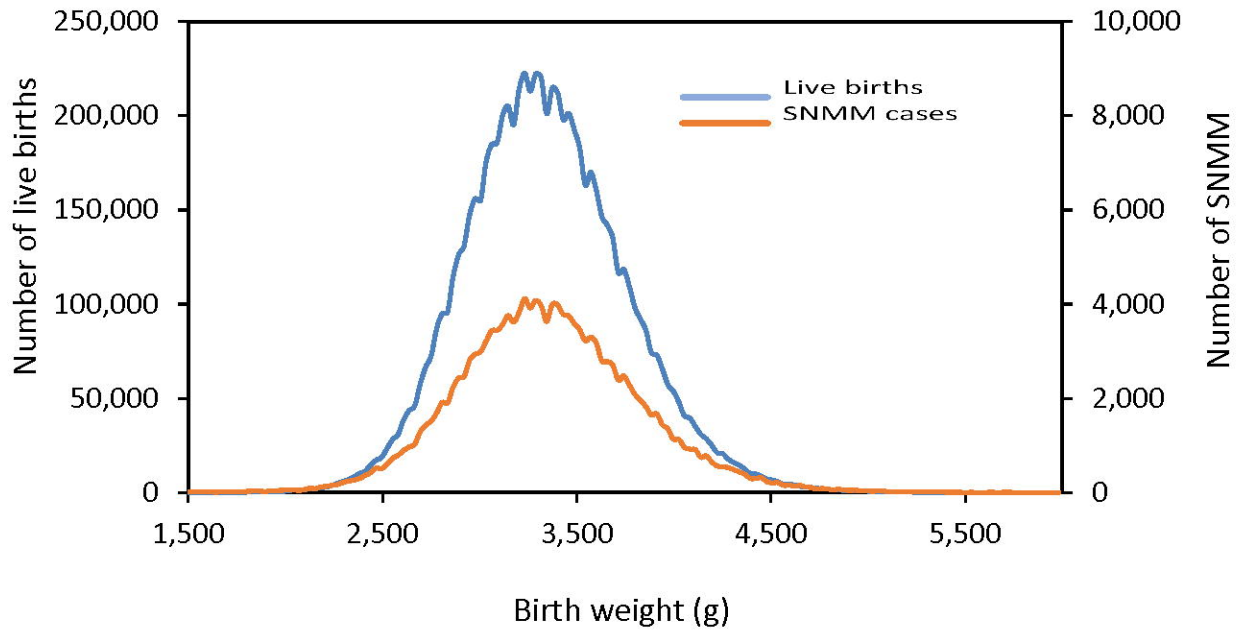
10 Figure 3. Distributions of live births and composite serious neonatal morbidity (SNMM) by birth  
11 weight, singleton females (Panel A) and males (Panel B) at 39 weeks' gestation, United States,  
12 2003-2017.





**A**

Female singletons at 39 week's gestation



**B**

Male singletons at 39 week's gestation

