ESJ Natural/Life/Medical Sciences Peer-reviewed

Assessing Biometric Predictors of Gestational Age Among Pregnant Women in the Upper East Region of Ghana: The Quadratic Classifier Approach

Abukari Alhassan, PhD
Department of Statistics, Faculty of Physical Sciences
University for Development Studies, Ghana

Doi:10.19044/esj.2025.v21n30p1

Submitted: 12 September 2025
Accepted: 12 October 2025
Published: 31 October 2025

Cite As:
Alhassan, A. (2025). *Assessing Biometric Predictors of Gestational Age Among Pregnant Women in the Upper East Region of Ghana: The Quadratic Classifier Approach*. European Scientific Journal, ESJ, 21 (30), 1. https://doi.org/10.19044/esj.2025.v21n30p1

Abstract

Aims/Objectives: Gestational outcomes are influenced by a variety of maternal factors, including diabetes, yet not all pregnant women with gestational diabetes experience abnormal variations. It is, therefore, essential to examine additional predictors of gestational variations amongst women. **Subject/Methods**: Data were obtained from the Biostatistics Department of the War Memorial Hospital in Navrongo, Ghana. Records of 1085 mothers and their children were collected between January 2014 and January 2017, and analysed using the quadratic discriminant analysis to evaluate the impact of maternal and neonatal characteristics on gestational variation. **Results:** Maternal parity, age, and the weight of the newborn were the principal discriminating variables. Of these, parity was the most significant factor in distinguishing between deliveries below the Estimated Date of Confinement (EDC) within EDC, and above EDC. **Conclusion:** Parity was identified as the leading factor influencing gestational variation. The study recommends further research into the biochemical and physiological mechanisms linking parity to gestational outcomes.

Keywords: Estimated Date of Confinement, Gestation, Discriminant Analysis, Quadratic Classification, Preterm Birth

Introduction

Gestation refers to the period between conception and birth. Clinically, gestational age is calculated from the last day of the mother's menstrual cycle, with a typical duration of 40weeks (280 days). The due date is termed the estimated date of confinement (EDC), yet only about 4% of women deliver within the EDC (Ohuma E. *et al*., 2023). Deviation from EDC results in preterm and post-term births, both of which are associated with elevated risks of maternal and neonatal morbidity and mortality. Because of the scale of these risks and their contribution to global health burdens, the World Health Organization (WHO) proposed some measures in 2010 to minimize mortality related to premature deliveries by 50% between 2010 and 2015.

Globally, preterm birth (PTB) affects over 13 million infants annually, representing more than 10% of all births (WHO, 2021). Sub-Saharan Africa and South Asia account for approximately 60% of these cases. In Ghana, preterm birth is a leading cause of neonatal mortality, contributing significantly to the estimated 29,000 newborn deaths each year (UNICEF-WHO, 2015). Post-term pregnancies increase risks of macrosomia, maternal injuries, and perinatal complications (Sam, 2021).

Previous research has associated PTB and post-term outcomes with factors such as maternal age, parity, birth weight, and maternal height (Bakhteyor *et al.,* 2012; Yamoah, 2014; Derraik et al., 2016). However, limited studies have applied advanced classification methods to evaluate these predictors in Ghanaian populations. This study addresses this gap by employing quadratic discriminant analysis (QDA) to classify gestational outcomes among mothers in Navrongo, Ghana.

Methods

Sample

Data were sourced from the War Memorial Hospital, Navrongo, covering January 2014 to January 2017. The sample consisted of a total of 1085 mother-infant pairs. Variables included the maternal age, height, parity, complications, gestational period, and infant birth weight.

Study Area

The study was conducted in Kasena-Nankana Municipality, Upper East Region, Ghana. The municipality had a population of approximately 163,620 (Ghana Statistical Service, 2010) and relies primarily on agriculture. The War Memorial Hospital serves as the principal referral centre.

Analytical Approach

The Quadratic Discriminant Analysis (QDA) was employed to classify gestational outcomes into three categories: below EDC ($\leq$ 37weeks), within

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

EDC (38–40 weeks), and above EDC (≥ 41weeks). QDA was selected following diagnostic tests confirming unequal covariance matrices across groups. Cross-validation was applied to estimate error rates.

Discriminant Analysis

Discriminant Analysis (DA) is a multivariate statistical technique used to determine which variables discriminate between two or more naturally occurring groups. Through DA, one may classify women into two or more mutually exclusive and exhaustive groups on the basis of a set of independent variables.

Linear Discriminant/Classification Model ($\Sigma i = \Sigma j = \Sigma$)

Assume that the two populations $\pi_1$and π_2 have multivariate normal densities $X' = [x_1, x_2, \dots, x_p]$ and that their respective mean vectors and covariance matrices are, μ_1, Σ_1 and μ_2, Σ_2 correspondingly given by

$$f_i(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x - \mu_i)'\Sigma^{-1}(x - \mu_i)\right] \; for\ i = 1, 2 \tag{1}$$

The allocation rule that minimizes the expected cost of misclassification (ECM) is given by: Allocate x_0 to π_1 if:

$$(\mu_1 - \mu_2)'\Sigma^{-1}x_0 - \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \geq ln \ln\left[\left(\frac{c\left(\frac{1}{2}\right)}{c\left(\frac{2}{1}\right)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{2}$$

Allocate x_0 to π_2 otherwise (Johnson and Wichern 2007).

The population parameters in (2) can be replaced by its sample estimates; $\underline{x}_1, \underline{x}_2$ and S_{pooled} . Given a special case when there are equal prior probabilities and equal misclassification cost, then we assign x_0 to π_1 if:

$$\left(\underline{x}_1 - \underline{x}_2\right)' S_{pooled}{}^{-1}\, x - \frac{1}{2}\left(\underline{x}_1 - \underline{x}_2\right)' S_{pooled}{}^{-1} \left(\underline{x}_1 + \underline{x}_2\right)' \geq 0 \tag{3}$$

We can estimate additional discriminant functions, such as the one shown above, when there are more than two groups. When there are three groups, for instance, we could estimate two functions: one to distinguish between group 1 and the combination of groups 2 and 3, and another to distinguish between group 2 and group 3. We could, for instance, have a function that distinguishes

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

between high school graduates who attend college and those who do not (rather, go on to get a job or attend a professional school), and another function that distinguishes between graduates who attend professional schools and those who do not. The coefficients in those discriminant functions could then be interpreted as before.

The Quadratic Classification Model ($\Sigma i \neq \Sigma j$)

The density ratio serves as the decision boundary or the minimum estimated cost of misclassification $f_1(x)/f_2(x)$. Substituting multivariate normal densities with different covariance matrices into (1) after taking natural logarithms and simplifying, the resulting classification regions are:

$$R_1: -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu'_1\Sigma_1^{-1} - \mu'_2\Sigma_2^{-1})x - K \geq ln\ ln\ \left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

$$R_2: -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu'_1\Sigma_1^{-1} - \mu'_2\Sigma_2^{-1})x - K \geq ln\ \ln\left[\left(\frac{c\left(\frac{1}{2}\right)}{c\left(\frac{2}{1}\right)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{4}$$

By substituting sample estimates for population parameters, the allocation function that reduces the expected cost of misclassification is obtained, and the minimal ECM is given as follows:

Allocate x_0 to π_1 if:

$$-\frac{1}{2}x'_0(S_1^{-1} - S_2^{-1})x_0 + \left(\underline{x'}_1S_1^{-1} - S_2^{-1}\underline{x'}_2\right)x_0 - K \geq ln\ \ln\left[\left(\frac{c\left(\frac{1}{2}\right)}{c\left(\frac{2}{1}\right)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{5}$$

Allocate x_0 to π_2 otherwise (Johnson and Wichern 2007).

Where,

$$K = \frac{1}{2}ln\ \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}\left(\underline{x'}_1S_1^{-1}\underline{x_1} - \underline{x'}_2S_2^{-1}\underline{x_2}\right) \tag{6}$$

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

$\left(\frac{c(1/2)}{c(2/1)}\right)$ is the expected cost ratio and $\left(\frac{p_2}{p_1}\right)$ is the prior probability ratio. If we assume that there are equal prior probability and misclassification costs for each population, the allocation rule becomes,

$$-\frac{1}{2}x'_0(S_1^{-1} - S_2^{-1})x_0 + \left(\underline{x'}_1 S_1^{-1} - S_2^{-1}\underline{x'}_2\right)x_0 - K \geq 1 \tag{7}$$

Error Rate Estimation

The holdout or cross-validation approach was used to assess the performance of the classification function. This method usually holds one observation and classifies the hold-out observation. The process is repeated until all observations are classified, producing unbiased estimates of the misclassification probabilities (Lachenbruch and Mickey 1968).

Organization of Data

Three categories, below EDC, within EDC, and above EDC, were used to categorize the gestational differences of mothers. Conceptually, the various gestational categories were viewed as follows: Within EDC is defined as delivery between 38 to 40 weeks of gestation, above EDC is defined as delivery after 41 weeks of gestation or above, while below EDC is defined as delivery in exactly 37 weeks of gestation or less. Some characteristics of the mothers and their neonates were examined quantitatively as the study's independent variables. These factors were as follows: the weight of the infant, the mother's height, her parity, and her age.

Results and Discussion

Table 1 displays the general characteristics of mothers and their newborns. According to the findings, women who give birth below the EDC have a mean age of 25, with a standard deviation of 5.91; those who give birth within the EDC have a mean age of 26, with a standard deviation of 6.64 and those who give birth above the EDC have a mean age of 32, with a standard deviation of 6.63. The result did not show much variation in mean maternal height between the various categories of time differences of birth. As compared to the parity of 1 for the two categories of women who give birth below and within the EDC of gestational differences, the results demonstrated a higher parity of 2 for women who deliver above EDC. Additionally, the results showed that babies born below EDC had a mean weight of 2.76 kg, babies born within EDC had a mean weight of 2.98 kg, and babies born above EDC had a mean weight of 2.95 kg. This demonstrates that babies born within the EDC are generally heavier than those born above or below the EDC.

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

Table 1: The Descriptive Statistics for Selected Variables

Variables	Below EDC		Within EDC		Above EDC	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Maternal Age	25.370	5.9149	26.791	6.6394	31.634	6.6345
Maternal Height	160.120	3.0584	160.670	3.4716	160.656	3.3281
Parity	1.295	1.1457	1.597	1.2031	2.527	1.2648
Baby's Weight	2.763	0.4251	2.980	0.3943	2.951	0.4252

Quadratic Discriminant Analysis

The Box M test of equality of population covariance matrices was initially ran in order to test the three groups under consideration for equal covariance matrices. The log determinant of the groups was as illustrated in Table 2. The Box M test was found to be significant at 1% level under the null hypothesis of equal covariance matrices, showing a violation of the assumption of equal covariance matrices.

Table 2: Test for Equality of Population Covariance Matrices.

Gestation	Rank	Log Determinant	Chi Square	df	P value
Below EDC	4	3.224	50.365523	20	0.0002*
Within EDC	4	3.508			
Above EDC	4	3.200			
Pooled	4	3.431			

*Significant at 1%

A diagnostic test for multicollinearity also revealed that there was no multicollinearity among the variables because the variance inflation (VIF) values of the independent variables ranged from 1 to 10, as shown in Table 3. Violations of the normality assumption are typically not "fatal" as long as it is caused by skewness and not outliers (Tabachnick and Fidell, 1996). The linearity assumption in discriminant analysis is frequently ignored, unless transformed variables are used as new predictor variables (Clelok, 2017).

Table 3: Test for Multicollinearity

Statistic	**Baby's Weight**	**Maternal Height**	**Parity**	**Age**
Tolerance	0.9255	0.9227	0.2942	0.2939
VIF	1.0806	1.0838	3.3985	3.4030

The data was then fitted with a quadratic classification function. The quadratic classifier's results demonstrated a significant performance at 1% significance level (Table 4).

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

Table 4: Test of Model Adequacy

Test Statistic	Value	F Value	Num DF	Den DF	P Value
Wilks' Lambda	0.87260139	19.02	8	2158	<.0001*
Pillai's Trace	0.13040666	18.83	8	2160	<.0001*
Hotelling-Lawley Trace	0.14255142	19.22	8	1539.1	<.0001*
Roy's Greatest Root	0.11168620	30.16	4	1080	<.0001*

* Significant at 1%

Table 5 presents the result of classification and misclassification rates. 32.98 % of the women were correctly classified as below EDC of gestation with a misclassification rate of 67.02%. However, 14.94% of women Within EDC of gestation were misclassified and 85.06% correct classification was achieved. The results further indicated that for women above EDC of gestation, 7.53% were correctly classified whiles 16.13% and 76.34% were misclassified into below EDC and within EDC, respectively. Consequently, an overall error rate of 0.3963 was achieved under the classification model. Further, the cross-validation option provides a better assessment of classification accuracy. For this data, 84.42% of women who gave birth Within EDC were classified correctly with a misclassification rate of 15.58% into the Below EDC category. From the result, it can be observed that approximately 60.37% (1–0.3963) correct classification of gestation was achieved under classification with QDF, as well as 60.00% (1–0.4000) correct classification rate under the cross-validated results.

Table 5: Quadratic Function Classification Results

	Classified			
	Below EDC	**Within EDC**	**Above EDC**	**Total**
True/Original				
Below EDC	124	252	0	376
Percent	32.98	67.02	0.00	100.00
Within EDC	92	524	0	616
Percent	14.94	85.06	0.00	100.00
Above EDC	15	71	7	93
Percent	16.13	76.34	7.53	100.00
Total	231	847	7	1085
Percent	21.29	78.06	0.65	100.00
Error Rate	0.6702	0.1494	0.9247	0.3963
Priors	0.3465	0.5677	0.0857	
Cross Validation				
Below EDC	124	252	0	376
Percent	32.98	67.02	0.00	100.00
Within EDC	96	520	0	616
Percent	15.58	84.42	0.00	100.00
Above EDC	15	71	7	93
Percent	16.13	76.34	7.53	100.00
Total	231	847	7	1085
Percent	21.29	78.06	0.65	100.00
Error Rate	0.6702	0.1558	0.9247	0.4000

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

The eigenvalue and canonical correlation coefficient were also used to examine the performance of the discriminant function. The canonical correlation's strength indicates how well the discriminant function can distinguish between different groups. According to Johnson and Wichern (2007), the total structure coefficient is deemed beneficial if it is equal to or higher than 0.30. The eigenvalue and canonical correlation coefficient in Table 6 demonstrate a well-defined model. The hypothesis that the canonical correlation in the current row and all that follows are zero indicated significance at 5 % level of significance, which showed that QDF was correctly specified.

Table 6: Test of Canonical Correlation

	Can. Corr.	Adj. Can. Corr.	Approx. SE	Square Can. Corr.	Eigenvalue
Function 1	0.316963	0.310859	0.027321	0.100466	0.1117
Function 2	0.173035	0.169916	0.029463	0.029941	0.0309
Test	**Likelihood Ratio**	**F Value**	**Df 1**	**Df 2**	**P – Value**
Function 1	0.87260139	19.02	8	2158	<.0001*
Function 2	0.97005892	11.11	3	1080	<.0001*

* Significant at 5%

The univariate test of class means (Table 7) reveals the minimum number of variables necessary for discrimination as well as the relevance of each variable in discrimination. The findings show that parity, age, and baby's weight were all significant at 1% ($P < 0.01$). While maternal height was significant at 5% ($P < 0.05$). The R-square and the adjusted R-square values show the amount of variation explained by each discriminating variable. Parity, Age and baby's weight explained large proportions of the variability (7.45%, 6.63% and 6.28%) among the classes and hence indicated their level of contribution to the group separation (Table 7). In contrast to a previous study on maternal height by Derraik et al. (2016), which found that globally, idiopathic preterm births are likely influenced by maternal small stature, partly because of anatomical limitations, the results of this study showed that maternal height was not a contributing variable to the group separation. However, the findings showed that a woman's age had an impact on the gestational differences, which supported a previous publication by UNICEF, WHO, UNESCO, UNFPA, UNDP, UNAIDS, WFP, World Bank (2010), on facts of life, which stated that women between the ages of 15 and 18 are more likely to give birth prematurely, while those over 35 are more likely to have post-term birth.

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

Table 7: Univariate Test of Class Means

Variable	Total SD	R-Square	Adjusted R-Square	F value	P value
Parity*	1.2313	0.0694	0.0745	40.32	<.0001
Age*	6.5996	0.0622	0.0663	35.86	<.0001
Maternal Height**	3.3289	0.0062	0.0062	3.35	0.0353
Baby's Weight*	0.4200	0.0591	0.0628	33.99	<.0001

* Significant at 1%; ** Significant at 5%

The Receiver Operating Characteristic (ROC) curve is shown in Figure 1. ROC curve is a useful way to interpret sensitivity and specificity levels and to determine related cut scores. The area under the curve (AUC) of a ROC curve represents the overall diagnostic accuracy. The findings of this investigation supported the model's correct specification with an Area Under the Curve (AUC) of 65.4% which was fairly high and a significant P value at 5% level.

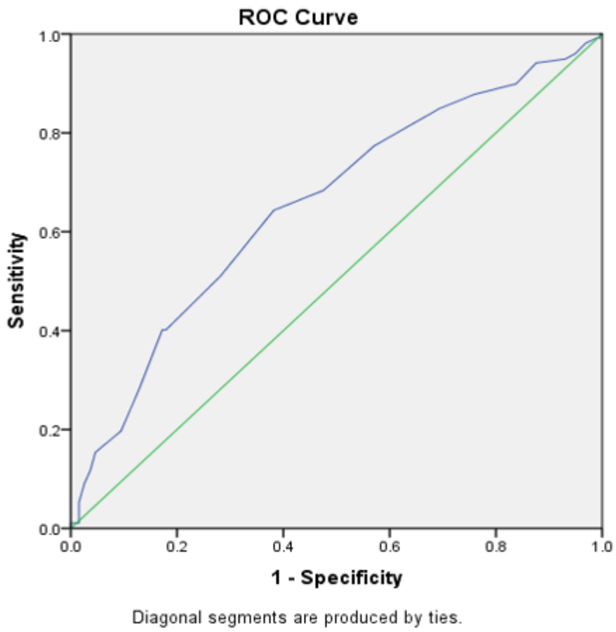

Figure 1: The Receiver Operating Characteristic (ROC)

The structural matrix in Table 8 was used to analyze the significance of each variable in the discriminant function. According to the findings, the first function's main discriminating variables were parity, age and baby's weight, while the second function's only most important discriminating variable is baby's weight. Therefore, these elements are what contribute to birth time discrepancies. However, because parity had the highest structural coefficient of the three factors, it was found to be the most important among the three variables.

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

Table 8: Structure Matrix

Variables	Function 1	Function 2
Parity	0.762873	-0.603169
Age	0.710869	-0.616739
Maternal Height	0.216592	0.220065
Baby's Weight	0.645211	0.760002

Table 9 displays the standardized and unstandardized canonical discriminant coefficients of the QDF for below EDC, within EDC, and above EDC of gestational differences in women, with first canonical class means of -0.40, 0.14, and 0.72 and second canonical class means of -0.11, 0.13, and -0.43, respectively. The score functions for the quadratic discriminant analysis are computed using the standardized canonical discriminant function coefficients in the table. From the results it was observed that for both functions (1 and 2), baby's weight had the greatest magnitude amongst the other variables. To classify future observations of pregnant women, the unstandardized canonical discriminant function coefficients of Table 9 can be used. For each function, women are classified as belonging to the class whose canonical coefficient is closest to the class mean.

Table 9: Unstandardized and Standardized Canonical Discriminant Coefficients

Variables	Unstandardized		Standardized	
	Canonical 1	**Canonical 2**	**Canonical 1**	**Canonical 2**
Parity	0.464871473	-0.252724515	0.5724168710	-0.3111909080
Age	0.039490654	-0.056991583	0.2606240095	-0.3761238007
Maternal Height	0.018835953	-0.001096003	0.0627038206	-0.0036485304
Baby's Weight	1.541879715	1.865502707	0.6476329653	0.7835637492

The results of this study supported earlier research on post-term and preterm births that found parity, age, and baby's weight to be significant variables influencing the time variations in birth (Marie, et al., 2018: UNICEF, WHO, UNESCO, UNFPA, UNDP, UNAIDS, WFP, World Bank, 2010). The findings also support the conclusions made by Marie D. *et al.,* (2018) that common risk factors underpin changes in the gestational age distribution among women.

Conclusions

Parity, maternal age, and infant birth weight are critical factors influencing gestational outcomes in Navrongo, Ghana. Parity emerged as the most influential predictor, underscoring its potential role in maternal health interventions. Further research should investigate the biological mechanisms underlying this relationship, with emphasis on parity-related physiological and biochemical processes. Monitoring these factors may enhance preterm birth prevention strategies and improve maternal and neonatal health outcomes.

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

Acknowledgments

We are grateful to the administrator at War Memorial Hospital for authorizing the release of the data used for the study.

Conflict of Interest: The authors reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The authors did not obtain any funding for this research.

References:

1. Bakhteyar, K., Lorzadeh, N., Pouria, Y., Birjardi, M., Ebrahimzadeh, F., and Karman, A., (2012). Factors associated with preterm delivery in women admitted to hospitals in Khorramabad: A case control study. *International Journal of Health and Allied Sciences*, 1(3): 147-152.
2. Célok, T. (2017). *Testing the assumptions for discriminant analysis.* Retrieved from: http://www.tankonyvtar.hu/hu/tartalom/tamop425/0049_08_quantitative_information_forming_meth Accessed 5 April 2017.
3. Derraik, J., Lundgren, M., Cutfield, W., & Ahlsson, F. (2016). Maternal height and preterm birth: A Study on 192,432 Swedish Women. *PLos ONE*.
4. Ghana Statistical Service (2021). 2010 Population and Housing Census (PHC) Final Results. [Online] Available: www.statsghana.gov.gh/docfiles/2010phc/ (October 28, 2012)
5. Johnson, R.A. and Wichern, D.W. (2007) Applied Multivariate Statistical Analysis. 6th Edition, Pearson Prentice Hall, Upper Saddle River
6. Lachenbruch PA, Mickey MA (1968). Estimation of error rates in discriminant analysis. Technometrics 10(15):1–11
7. Marie D., Laust M., Ashna D. H., Beatrice B., Mika G., Michael R. K., Jennifer L. R., Paromita D., Jocelyn R., Naho M., Natasha N., Francisco B., Sylvie B., Anne-Marie N. A., and Michael S. K., Jennifer Z. (2018). International variations in the gestational age distribution of births: an ecological study in 34 high-income countries. *The European Journal of Public Health*, Vol. 28, No. 2, 303–309. *doi:10.1093/eurpub/ckx131*.
8. Ohuma E, Moller A-B, Bradley E (2023). National, regional, and worldwide estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. Lancet. (In Press).
9. Sam, U. (2021). Post Term Pregnancy Causes and Risk. *Birth Injury Guide*. https://www.birthinjuryguide.org/causes/post-term-pregnancy/

 ISSN: 1857-7881 (Print) e - ISSN 1857-7431

10. Tabachnick, Barbara, G., & Fidell, L. S. (1996). *Using Multivariate Statistics.* New York,NY: HarperCollins College Publishers.
11. UNICEF, WHO, UNESCO, UNFPA, UNDP, UNAIDS, WFP, World Bank (2010) - The Mother and Child Health and Education Trust. https://web.archive.org/web/20210909235356/https://factsforlife.org/00/foreword.html
12. UNICEF-WHO. (2015). Low Birth Estimates: Levels and Trends. https://www.who.int/publications/i/item/WHO-NMH-NHD-19.21
13. WHO (2021). Born Too Soon: The Global Action Report on Preterm Birth. Eds CP Howson, MV Kinney, JE Lawn.
14. Yamoah, N. A. (2014). *Using Binary Logistic and Quantile Regressions for Determinants of Preterm.*