Factors associated with low birth weight in Nepal using multiple imputation

CrossMark

Usha Singh[1,2]* , Attachai Ueranantasun[2] and Metta Kuning[2]

Abstract

Background: Survey data from low income countries on birth weight usually pose a persistent problem. The studies conducted on birth weight have acknowledged missing data on birth weight, but they are not included in the analysis. Furthermore, other missing data presented on determinants of birth weight are not addressed. Thus, this study tries to identify determinants that are associated with low birth weight (LBW) using multiple imputation to handle missing data on birth weight and its determinants.

Methods: The child dataset from Nepal Demographic and Health Survey (NDHS), 2011 was utilized in this study. A total of 5,240 children were born between 2006 and 2011, out of which 87% had at least one measured variable missing and 21% had no recorded birth weight. All the analyses were carried out in R version 3.1.3. Transform-then impute method was applied to check for interaction between explanatory variables and imputed missing data. Survey package was applied to each imputed dataset to account for survey design and sampling method. Survey logistic regression was applied to identify the determinants associated with LBW.

Results: The prevalence of LBW was 15.4% after imputation. Women with the highest autonomy on their own health compared to those with health decisions involving husband or others (adjusted odds ratio (OR) 1.87, 95% confidence interval (95% CI) = 1.31, 2.67), and husband and women together (adjusted OR 1.57, 95% CI = 1.05, 2.35) were less likely to give birth to LBW infants. Mothers using highly polluting cooking fuels (adjusted OR 1.49, 95% CI = 1.03, 2.22) were more likely to give birth to LBW infants than mothers using non-polluting cooking fuels.

Conclusion: The findings of this study suggested that obtaining the prevalence of LBW from only the sample of measured birth weight and ignoring missing data results in underestimation.

Keywords: Multiple imputation, Low birth weight, Survey package and Transform-then impute

Background

Missing data occur almost in all types of studies and cause inefficient and biased estimates of parameters if they are handled improperly. In a survey, missing data occur, when a selected respondent refuses to participate (unit nonresponse) or respondent does not provide answer to entire survey questions (item nonresponse) [1, 2]. For unit nonresponse, the weighting adjustment technique is applied, in which weight of respondents are increased to represent non-respondents [3], whereas for item nonresponse, imputation methods are employed [1, 4].

There are three types of mechanisms under which missing data occur: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) [1, 5]. When missing data are MCAR, the probability of missingness does not depend on the missing and other observed data. An example is when survey papers are lost accidentally. If missing data are MAR, the probability of missingness depends only on observed data, but not on the missing data themselves. For example, people from different demographic backgrounds may decline to answer based on beliefs or traditions. When missing data are MNAR, the probability of missingness depends on both observed and missing data. For example, people with high incomes are less likely to report their incomes than those of people with average or low income. Data under MCAR mechanism can be

* Correspondence: usha.singh36@gmail.com
[1]Nepal Institute of Health Sciences, Gokarneswor Municipality-12, Jorpati, Kathmandu, Nepal
[2]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000, Thailand

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 2 of 10

tested statistically by little's test [6]. However, there is no clear technique to diagnose and distinguish between MAR and MNAR. Thus, MAR and MNAR can only be reasoned or hypothesized [1, 4].

There are several studies about methods used for handling missing data in each type of missing mechanisms [7]. The most common method is case deletion in which subjects with missing values are deleted. The results from this method are inefficient, but unbiased, when the missing data hold MCAR assumption. However, when data are not MCAR, the results from this method are inefficient and biased [4, 5]. Methods like mean substitution, last observation carried forward, hot deck imputation, cold deck imputation and regression imputation come under single imputation in which missing values are replaced by synthetic values [2, 8]. The first two methods of single imputation assume missing data are MCAR, while the remaining methods assume missing data are MAR [7]. The results obtained from mean substitution and hot deck imputation are biased under three missing mechanisms. However, the results obtained from conditional mean imputation are unbiased under MCAR and MAR, but may be biased under MNAR [4]. Furthermore, in single imputation, values are imputed for one time; the uncertainties created by missing values are not accounted for. As a result, there are small standard errors, p-values and narrow confidence intervals [5, 9]. In multiple imputation, unlike single imputation, missing values are imputed for more than one time and the uncertainties created by missing values are incorporated resulting in larger standard errors and wider confidence intervals [1]. In addition, multiple imputation provide unbiased result, when data hold both MAR and MNAR assumption [4].

In southern Asia and sub-Saharan Africa, more than half of women give birth at home [10]. Therefore, analyzing data on infants delivered only at hospital would be biased [11]. As a substitute to hospital based data, household data survey begin to collect information on infants born outside health facilities [12]. However, the data on birth weight from a household survey become limited since mothers are unable to provide numeric birth weight [11, 12]. Nepal Demographic and Health Survey (NDHS), 2011 reported that only 36% of weights of infants were measured at the time of birth [13]. The same survey also reported that the prevalence of low birth weight (LBW) in Nepal was 12%, which was calculated from the available birth weight of infants. Studies conducted in Nepal on LBW by using demographic and health survey (DHS) data either have considered mother's recall for infant's size at birth as an alternative to the birth weight [14] or analyzed the subset of measured birth weight [15] for identifying the prevalence and factors associated with LBW. Estimating prevalence of LBW and identifying determinants associated with it only from the available birth weight may be biased, when missing birth weight are not MCAR. Besides missing values on the birth weight, missing values are also presented on determinants of birth weight, but are not handled in most of previous studies and the results obtained from these studies may be misrepresented. Thus, the main objective of this study is to identify factors associated with LBW using multiple imputation to handle missing data in both outcome and determinants.

Methods

NDHS data

The child dataset from Nepal Demographic and Health Survey (NDHS), 2011 was analyzed in this study. NDHS is a nationally representative household survey conducted every 5 years [13]. Multistage cluster sampling was used in this survey. In the first stage, the probability proportionate to size was used to select wards from rural areas and sub-wards from urban areas. In the second stage, random sampling was done to select households [13, 16]. Details of clustering, listing and sample selection have been mentioned elsewhere [16]. The survey interviewed 12,674 women aged 15 to 49 and 4,121 men aged 15 to 59. Three main questionnaires were administered including household questionnaire, women's questionnaire and men's questionnaire to collect information from different levels. These questionnaires contained different units of analysis and they were eventually converted into seven datasets [13]. In this study, child dataset was used. From this dataset, a total of 5,306 children were born during the period of 2006–2011. Children from multiple births tend to have lower birth weights than singletons[17]. Thus, 66 multiple births were excluded from this study and only 5,240 singleton children were included in this study. However, out of 5,240 children, 766 (13.4%) children had completed the record and 4,474 (86.6%) children had at least one of the measured variables missing.

Study variables

Birth weight of an infant was considered the outcome of this study. Based on World Health Organization (WHO) classification, birth weight was divided into normal birth weight, equal to or greater than 2,500 g, and LBW, lower than 2,500 g [18]. In this study, all the study variables that were included in [15] were employed. The study variables were classified under three major determinants. These are underlying factors, proximate factors and factors related to gestation and fetal growth. The underlying factors were made up of economic status (wealth index), mother's education, women's decision for utilization of health services, ethnicity, residence and development region. Body mass index (BMI), birth interval, antenatal care (ANC) visits,

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 3 of 10

and consumption of iron tablets during pregnancy, smoking and use of polluting cooking fuel constituted the proximate factors. Gestation and fetal growth factors were mother's age at child's birth, parity and gender of child were employed in this study. Besides birth weight, NDHS 2011 asked a specific question to mothers about the size of their babies at the time of birth. Based upon five categories namely very large, large, normal, small and very small, mothers had to recall their babies' size. This variable was used as an auxiliary variable for imputing birth weight in this study. In the dataset, there was no mother's age at child's birth variable. Mother's age at child's birth was calculated from mother's current age minus date of child's birth and was categorized as 15–19 years, 20–24 years, 25–29 years and 30 years and above. Mother's education was categorized into no education, primary education, secondary/higher education. Here, all study variables were in categories and the categorization of the study variables were based on previous studies which used similar DHS datasets conducted in Nepal [14, 15].

Frequency, pattern and reason for missingness of missing data

Before handling missing data, the frequency, pattern and reason for missing data were checked. For the graphical presentation of missing data and its pattern, VIM package in R was used. The percentages of missing values in each variable and the patterns of missing data are displayed in Fig. 1, left (a) and right (b) panel respectively. From Fig. 1a, the highest percentages of missing values were from the variables of birth weight (63.32%), and BMI (52%). The percentages of missing were nearly equal for ANC and consumption of iron tablets during pregnancy, whereas the percentages of missingness were less than 10 for cooking fuel and women's decision for utilization health of services. There were no missing values for the variables such as mother's age at child's birth, gender of child, parity, mother's education, wealth index, ethnicity, residence, ecological region, development region, birth interval and smoking. The pattern of missing data shown in Fig. 1b was arbitrary, because the missing values for the variables of any record were seen in a random fashion. From Fig. 1b, only 13.4% of children had completed the record without missing values, while 21.2% of children data contained missing values only on birth weight and 15.4% children had missing values only on BMI. Furthermore, 21.9% of children had missing values on both birth weight and BMI, and only 8.7% of infants data contained missing values on birth weight, mother's BMI, ANC visit and consumption of iron tablets during pregnancy.

The missing mechanism was diagnosed by implementing Little's test [6] to identify whether missing data were MCAR. The test revealed that data were not MCAR
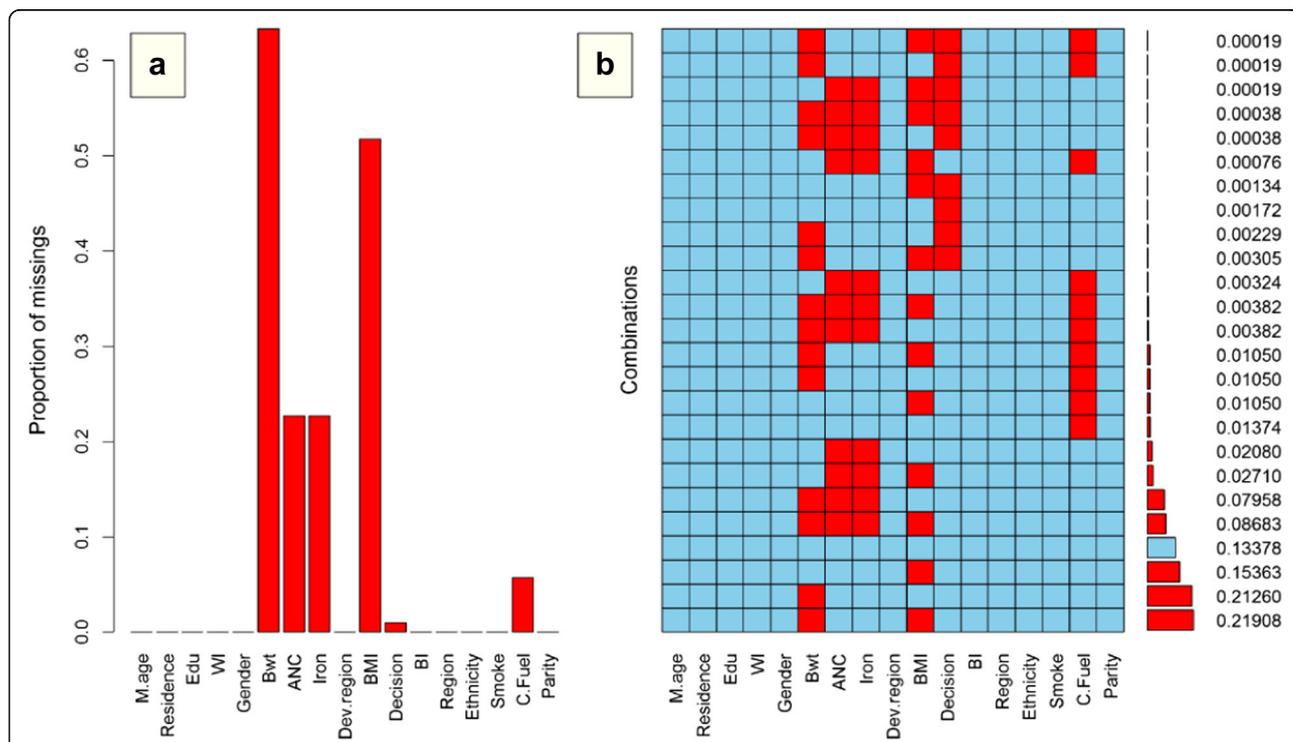
Fig. 1 Percentage and pattern of missing data. Note: *M.age: Mother's age at child's birth, Edu: education, WI: wealth index, Bwt: birth weight, Iron: consumption of iron tablets during pregnancy, Decision: women's decision for utilization of health services, BI: birth interval and C. Fuel: cooking fuel*

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 4 of 10

since *p*-value was found to be around 0.000. From the data, missing birth weight was due to home delivery. The reason for missing values on ANC visit was probably because mothers living in rural areas felt shy to report a number of ANC visits during the time of interview. Missing values on consumption of iron tablets during pregnancy might be due to missing values on ANC visit, because mothers who did not report their ANC visit were less likely to report any consumption of iron tablets during pregnancy. Missing values on mother's BMI were because of refusal to measure height and weight either by a respondent or a respondent's mother. Furthermore, missing values for cooking fuel were for those mothers who did not belong to the household (non de jure residents), but presented at the place for the time of an interview. It was evident that missing values in the variables were not missing due to themselves, but were missing due to other characteristics. Thus, in this study, missing data were under MAR assumption.

Background of Multiple imputation

Multiple imputation yields unbiased estimates of parameters, when missing data hold MAR assumption [4], and as aforementioned in the last section, missing data in this study held MAR assumption. Therefore, multiple imputation was applied to handle missing data. In multiple imputation technique, each missing value is imputed by $m > 1$ times resulting into m datasets. Each dataset is analyzed by using complete data method. The estimates of parameters of m datasets are pooled to calculate overall estimates of parameters and confidence intervals that identify missing data uncertainty [1].

For combining estimates of parameters of m datasets, formulas derived by [1] is used. Suppose the regression coefficient for an imputed dataset i is Q_i and U_i be the variance where $i = 1, 2,...,m$. Therefore, the overall regression coefficient is the average of all Q_i and shown in Eq. (1).

$$\overline{Q} = \frac{1}{m}\sum_{i=1}^{m} Q_i \tag{1}$$

The variance within imputation is average of all U_i and is shown in Eq. (2).

$$\overline{U} = \frac{1}{m}\sum_{i=1}^{m} U_i \tag{2}$$

The variance between imputations is displayed in Eq. (3).

$$B = \frac{1}{m-1}\sum_{i=1}^{m} \left(Q_i - \overline{Q}\right)^2 \tag{3}$$

The total variance is a combination of variance within and in between imputations which is displayed in Eq. (4).

$$T = \overline{U} + \left(1 + \frac{1}{m}\right) B \tag{4}$$

The overall standard error is the square root of total variance T and is displayed in Eq. 5.

$$SE = \sqrt{T} \tag{5}$$

In multiple imputation, methods like Joint Modeling (JM) and Multiple Imputation by Chained Equations (MICE) also called as Fully Conditional Specification (FCS) have been proposed to impute missing data [19]. In MICE approach, a series of regression models are performed in which each variable with missing data is modeled conditionally upon other variables in the dataset. This signifies that each variable has its own imputation model. For example, logistic regression model is used for binary variables and linear regression model used for continuous variables [20]. As described in [19], multiple imputation involves three main steps: imputation, analysis and pooling. Firstly, an imputation model is used to generate the missing values using possible values. In the imputation model, auxiliary variables and variables that can explain a missing mechanism are kept for a better prediction of missing values and making MAR hypothesis more possible [21, 22]. Initially, three to five imputations are suggested for obtaining outstanding results [23]; however, [24] recommended the number of imputation should be over than or equal to the percentage of missing data. Secondly, an analysis model is applied to estimate parameters for each imputed dataset. Basically, in theory, the analysis model and the imputation model need to be the same, but they can be different in practice [22]. Finally, the estimated coefficients, standard errors and confidence intervals from each model are pooled together using Rubin's rule.

In this study, missing values were in both independent and dependent variables. As stated by [6], if the missing values are presented in both determinants (X) and outcome (Y), then cases with the missing outcome (Y) can confer a little information for the regression of interest, by improving prediction of missing determinants (X) for cases with the outcome (Y) present. Therefore, under a particular condition Multiple Imputation then Deletion (MID) performs better than standard multiple imputation in which all missing values on determinants (X) and outcome (Y) are imputed, and then deleting cases with imputed values on outcome (Y) before analysis [25]. However, standard multiple imputation performs better than MID when auxiliary variables are included in an imputation model as stated by [26]. Hence, mother's opinion on infant's size at birth was employed as an auxiliary variable in this study for the better result.

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 5 of 10

Implementation and statistical analysis

The pattern of missing data in this study was arbitrary and missing variables were categorical; hence, FCS method was considered appropriate [19]. Therefore, mice package in R was used in this study, because in mice package, multiple imputation using FCS is implemented by MICE algorithm [19]. For combining each imputed dataset, mitools package by [27] was applied. In this study, multiple imputation was carried out for 65 times, because the highest percentage of missing was 63.32. As suggested by [28], mother's opinion on infant's birth size can be used as an alternative to the birth weight. Therefore mother's opinion on infant's size at birth was considered an auxiliary variable in this study. Before the imputation model, possible interaction between the variables like ANC and iron tablets consumption during pregnancy, wealth index and mother's education, education and women's decision for utilization of health services, ecological region and developmental region, and development region and women's decision for utilization of health services was checked using transform-then-impute method as described by [29]. It was found that no interaction among them presented, because the *p*-value was greater than 0.05. Consequently, all the study variables along with auxiliary variable were included into the imputation model. Survey package by [30] in R was applied to each imputed dataset to account for sampling method and sample weights. Survey logistic regression model as an analysis model was applied to identify the factors associated with LBW. Under complex survey data, the parameters are estimated by pseudo likelihood method instead of maximum likelihood [31]. Therefore, the adjusted Wald test statistic was applied for selecting significant variables.

Results

Prevalence of LBW

The overall and crude subgroup estimations of LBW prevalence and their 95% confidence intervals (95% CI) were calculated and shown in Table 1. The overall prevalence of LBW was 15.4% (95% CI = 12–18%) after imputation. The prevalences of LBW for the determinants like residence, ethnicity, mother's age at child's birth, parity and gender of child were nearly equal in each subgroup. However, the prevalence of LBW was different in each subgroup for the rest of variables. For the variables such as wealth index, BMI and ANC visit, the percentages of LBW were showing a decreasing trend starting from poor, underweight and no ANC visit respectively. For ecological region, the lowest prevalence of having LBW babies was for mothers living in Terai (13.7%), while mothers from other two subgroups had nearly similar prevalences of having LBW babies. Likewise, the percentage of giving

birth to LBW infants was the highest for mothers who gave birth to infants within a gap of less than 24 months from the previous birth (19.1%), while mothers for other two subgroups had almost equal prevalences of giving birth to LBW infants. For mothers who were not consuming iron tablets during pregnancy (18.9%), being smoker (21.0%) and using highly polluting cooking fuel (16.1%) showed the highest prevalences of having LBW babies compared to their respective subgroups. The percentages of giving birth to LBW infants among mothers who attended primary education (18.3%) and no education (16.1%) were close and higher than uneducated mothers (16.1%). In case of development region, the higher prevalences were evident in mothers residing in Far-western (19.2%), Eastern (18.3) and Mid-western (17.4%) than in mothers residing in other development regions.

Factors associated with LBW

For the univariate analysis, all study variables were analyzed by using simple survey logistic regression and results are displayed in Table 2. Women's decision for utilization of health services and cooking fuel were found statistically significant. Mothers were more likely to give birth to LBW infants, when decision on utilization of health services relied on husband and others (OR 1.91, 95% CI = 1.34, 2.72) and mother and her husband together (OR 1.54, 95% CI = 1.03, 2.30). Mothers using highly polluting cooking fuels (OR 1.56, 95% CI = 1.07, 2.28) were more likely to give birth to LBW infants than mothers using non-polluting cooking fuels. However, the variables like wealth index, mother's education, ethnicity, residence, ecological region, developmental region, mother's BMI, birth interval, ANC visit, consumption of iron tablets during pregnancy, smoking, mother's age at child's birth, parity and gender of child remain insignificant with LBW.

The significant variables in univariate analysis were further analyzed by using a multiple survey logistic regression model. Adjusted odds ratio (OR) and its 95% CI are shown in Table 3. The inference statistical tests were nearly unchanged for the final model. Women with the lowest autonomy on their own health compared to those with involvement of husband or others (adjusted OR 1.87, 95% CI = 1.31, 2.67) and with husband and women together (adjusted OR 1.57, 95% CI = 1.05, 2.35) had a greater chance to give birth to LBW infants. For the other significant variable, mothers using highly polluting cooking fuels (adjusted OR 1.56, 95% CI = 1.03, 2.22) were more likely to give birth to LBW infants than mothers using non-polluting cooking fuels.

Discussion

The overall prevalence of LBW from this study is 15.4% which is different from the study including only infants

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 6 of 10

Table 1 Overall and subgroup prevalences of LBW after imputation

Variables	Estimate	SE	95% CI
Overall	0.154	0.015	0.12, 0.18
Underlying factors			
Wealth index			
Poor	0.174	0.024	0.13, 0.22
Middle	0.152	0.026	0.10, 0.20
Rich	0.123	0.014	0.10, 0.15
Mother's education			
No education	0.161	0.025	0.11, 0.21
Primary education	0.183	0.022	0.14, 0.23
Secondary/higher education	0.124	0.014	0.10, 0.15
Women's decision for health service utilization			
Women	0.105	0.016	0.07, 0.14
Women and husband together	0.152	0.019	0.11, 0.19
Husband or others	0.181	0.020	0.14, 0.22
Ethnicity			
Relatively advantaged	0.159	0.018	0.12, 0.19
Relatively disadvantaged (Janjati)	0.142	0.020	0.10 0.18
Relatively disadvantaged (Dalit)	0.160	0.024	0.11, 0.21
Residence			
Rural	0.155	0.016	0.13, 0.19
Urban	0.140	0.014	0.11, 0.17
Ecological region			
Mountain	0.178	0.033	0.11, 0.24
Hill	0.170	0.020	0.13, 0.21
Terai	0.137	0.018	0.10, 0.17
Development region			
Eastern	0.183	0.024	0.14, 0.23
Central	0.121	0.021	0.08, 0.16
Western	0.131	0.025	0.08, 0.18
Mid-western	0.174	0.025	0.12, 0.22
Far-western	0.192	0.029	0.13, 0.25
Proximate factors			
Body mass index (BMI)			
< 18.5 (Underweight)	0.176	0.028	0.12, 0.23
18.5–23.0 (Normal)	0.160	0.018	0.12, 0.19
> 23.0 (Overweight)	0.115	0.021	0.07, 0.16
Birth interval			
No interval	0.150	0.014	0.12, 0.18
< 24 months	0.191	0.037	0.12, 0.26
≥ 24 months	0.146	0.019	0.11, 0.18

Table 1 Overall and subgroup prevalences of LBW after imputation *(Continued)*

ANC visit during pregnancy			
No visit	0.207	0.048	0.11, 0.30
One-three visits	0.167	0.023	0.12, 0.21
Four or more visits	0.125	0.012	0.10, 0.15
Consumption of iron tablets during pregnancy			
No	0.189	0.036	0.12, 0.26
Yes	0.143	0.013	0.12, 0.17
Smoking			
No	0.149	0.014	0.12, 0.18
Yes	0.210	0.058	0.10, 0.32
Fuel			
Low polluting fuel	0.110	0.017	0.08, 0.14
Highly polluting fuel	0.161	0.016	0.13, 0.19
Gestation and fetal growth factors			
Mother's age at child's birth (Years)			
15–19	0.160	0.020	0.12, 0.20
20–24	0.149	0.016	0.12, 0.18
25–29	0.149	0.022	0.11, 0.19
≥ 30	0.162	0.034	0.09, 0.23
Parity			
One	0.140	0.015	0.11, 0.17
Two-three	0.154	0.014	0.13, 0.18
Four and above	0.164	0.034	0.10, 0.23
Gender of baby			
Male	0.146	0.017	0.11, 0.18
Female	0.161	0.017	0.13, 0.20

with measured birth weight conducted by [15] in which the prevalence of LBW was found to be 11.5%. The difference is expected, because in this study there is an inclusion of additional 3,318 missing birth weight in the analysis. A study conducted by [14] found the prevalence of small size at birth as 16% which is close to the prevalence of this study. This may be because mother's recall of infant's size at birth and other variables are used for imputing missing values in this study. As shown in Table 1, the prevalences of LBW for the determinants like mother's age at child's birth, gender of child, residence, ethnicity and parity are almost equal in each subgroup. It can be concluded that each subgroup has equal chance of having LBW infants. In this study, the prevalences of having LBW infants are higher among mothers living in low standard such as being poor, using highly polluting cooking fuels, not attending ANC visit and not consuming iron tablets during pregnancy than those of

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 7 of 10

Table 2 Unadjusted odds ratio and 95% CI of study variables

Variables	Unadjusted OR	95% CI	p-value
Underlying factors			
Wealth index			
Rich	1.00		0.107
Middle	1.27	0.83, 1.96	
Poor	1.50	1.03, 2.20	
Mother's education			
Secondary/higher education	1.00		0.107
Primary education	1.58	1.09, 2.27	
No education	1.34	0.90, 2.00	
Women's decision for health service utilization			
Women	1.00		0.002*
Women and husband together	1.54	1.03, 2.30	
Husband or others	1.91	1.34, 2.72	
Ethnicity			
Relatively advantaged	1.00		0.758
Relatively disadvantaged (Janjati)	0.88	0.61, 1.26	
Relatively disadvantaged (Dalit)	1.00	0.67, 1.49	
Residence			
Urban	1.00		0.466
Rural	1.12	0.82, 1.53	
Ecological region			
Terai	1.00		0.289
Hill	1.29	0.89, 1.87	
Mountain	1.37	0.87, 2.17	
Development region			
Central	1.00		0.072
Eastern	1.63	1.03, 2.57	
Western	1.09	0.64, 1.84	
Mid-western	1.53	0.92, 2.57	
Far-western	1.72	1.07, 2.78	
Proximate factors			
Body mass index (BMI)			
> 23.0 (Overweight)	1.00		0.138
18.5–23.0 (Normal)	1.50	0.96, 2.33	
< 18.5 (Underweight)	1.67	1.03, 2.71	
Birth interval			
No interval	1.00		0.338
< 24 months	1.32	0.82, 2.15	
≥ 24 months	0.96	0.69, 1.33	
ANC visit during pregnancy			
Four or more visits	1.00		0.131
One-three visits	1.38	0.96, 1.98	
No visit	1.83	0.96, 3.51	
Consumption of iron tablets during pregnancy			
Yes	1.00		0.199
No	1.39	0.84, 2.30	
Smoke			
No	1.00		0.247
Yes	1.47	0.75, 2.88	
Fuel			
Low polluting fuel	1.00		0.023*
Highly polluting fuel	1.56	1.07, 2.28	
Gestation and fetal growth factors			
Mother's age at child's birth (Years)			
≥ 30	1.00		0.970
25–29	0.91	0.55, 1.51	
20–24	0.92	0.56, 1.50	
15–19	0.99	0.60, 1.66	
Parity			
Four and above	1.00		0.748
Two-three	0.94	0.60, 1.48	
One	0.84	0.51, 1.38	
Gender of baby			
Male	1.00		0.379
Female	1.13	0.86, 1.48	

p-value was calculated from Wald test, *statistically significant at 5% level

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 8 of 10

Table 3 Adjusted odds ratio and 95% CI of study variables

Variables	Adjusted OR	95% CI	*p*-value
Women's decision for health service utilization			
Women	1.00		0.006*
Women and husband together	1.57	1.05, 2.35	
Husband or others	1.87	1.31, 2.67	
Fuel			
Low polluting fuel	1.00		0.045*
Highly polluting fuel	1.49	1.03, 2.22	

p-value was calculated from Wald test, *statistically significant at 5% level

their respective subgroups and this finding is consistent with the previous study conducted by [14].

The prevalences of LBW for BMI and ethnicity in each subgroup are surprisingly different from normal perception. For BMI, women with overweight have the lower prevalence and the lower odds of LBW compared to women with normal and underweight. The possible explanation for this is that overweight mothers are likely to give birth to bigger babies and underweight mothers are likely to give birth to smaller babies. This finding is consistent with the studies conducted by [32, 33]. Furthermore, the results of this study reveal that the prevalence of LBW among relatively advantaged mother is higher than relatively disadvantaged mother (janajati). Even though, there have been studies on ethnicity affecting on LBW, these studies were performed in the high income countries [34, 35]. From those studies, it seems that mothers from the advantaged group are less likely to give birth to LBW infants. However, in this study, the different effects on LBW from mothers with different ethnic backgrounds are insignificant because *p*-value is higher than 0.05 from unadjusted odds ratio. Therefore, it is inconclusive to state that the odds of having LBW infants from differently ethnic mothers can be distinguished.

The current study finds that a mother has higher odds to give birth to LBW babies, when her decision on utilization of health services is relied only on others instead of herself and this finding is supported by [36] in which women with the lowest decision making autonomy were more likely to have LBW. This is probably because women with the lowest decision making autonomy on their health care are less likely to receive regular health checkups together with ANC visit during pregnancy including safe deliveries and health information regarding pregnancy and childbirth. Apart from that, women with the lowest decision making autonomy on their own health may have poor nutrition uptake during pregnancy and that may consequently impair fetal growth [36]. The variables such as ANC visit during pregnancy and consumption of iron tablets during pregnancy are not significant with LBW in the current study. However, studies performed by [14, 15] found that mothers

who did not attend ANC visit during pregnancy and mothers who did not consume iron tablets during pregnancy were more likely to give birth to LBW infants. This difference may be because [14, 15] assumed the missing values presented on ANC visit and iron tablets consumption during pregnancy as no ANC visit and no consumption of iron tablets during pregnancy respectively. The result from this study also finds that mothers who use highly polluting fuel are more likely to give birth to LBW infants and this finding is supported by a study conducted in India [37]. However, cooking fuel was found insignificant in the previous studies conducted in Nepal by [14, 15]. This is probably because [14, 15] supposed that mothers who did not belong to households (non dejure residents) used highly polluting cooking fuel.

The current study consists of missing data on the variables like birth weight, BMI, ANC visit, consumption of iron tablets during pregnancy, cooking fuel and women's decision for utilization of health services. For birth weight, even though there has been a considerable rise in the percentage of measurement of infants birth weight at birth in the past 5 years from 17% in 2006 to 36% in 2011 [13, 38], but home delivery is still a preferred choice for most mothers in Nepal as stated in [39, 40]. Eventually, the problem of missing data on birth weight may continue for a long period. This suggests promoting and strengthening institutional delivery, provision of weighing scale and training to community health workers for measurement of birth weight of those infants who are born at home. However, missing data in other variables can be minimized with other measures. For instance, in DHS survey, the questions related to cooking fuel, collected in household level, should be assigned to individuals in the individual data file. Thus, a mother who is not member of household lack the data on cooking fuel and the problem of missing data on cooking fuel can be avoided, if questions related to cooking fuel are included in women's questionnaire too.

Multiple imputation is employed in this study to handle missing data, because the analysis based on only complete cases of measured birth weight cannot be used since missing data are presented in more than one variable and the missing data are MAR. Moreover, using multiple imputation reduces bias downwards compared to analysis of complete cases, but it does not mean that using imputation methods for replacing missing values removes the bias completely.

The limitation of this study is that the efficiency of multiple imputation cannot be determined, because the data lack the completed record. Secondly, this efficiency might be lower because of high numbers of missing data. The study conducted by [7] mentioned that the results from statistical analysis are more prone to be biased, when the amount of missing is greater than 10%. However, as stated by [41], missing the data pattern and

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 9 of 10

missing mechanism are more important than the percentage of missing data. Furthermore, the current study utilized the secondary data; thus, the exact reason for missing data is not clear for many variables.

Conclusions

The findings of this study suggest that obtaining the prevalence of LBW from only the sample of measured birth weight results in underestimation of the prevalence. In addition, assuming missing values as non missing provides different results from the results with imputed data. Therefore, it is suggested for future researchers conducting studies on LBW with DHS data from low income countries that missing data on birth weight and its determinants should be imputed.

Abbreviations
ANC: Antenatal care; BMI: Body mass index; CI: Confidence interval; DHS: Demographic and health survey; FCS: Fully conditional specification; LBW: Low birth weight; MAR: Missing at random; MCAR: Missing completely at random; MICE: Multiple imputation by chained equations; MID: Multiple imputation then deletion; MNAR: Missing not at random; NDHS: Nepal demographic and health survey; OR: Odds ratio; WHO: World health organization

Availability of data and materials
This study used the data from Nepal Demographic and Health Survey, 2011 and the data is available in the DHS website http://www.dhsprogram.com/.

Authors' contributions
US involved in extracting data from DHS dataset, data analysis and interpretation, and drafting of the manuscript. AU involved in drafting the manuscript and revising it critically for intellectual content. MK involved in critically revising the manuscript. All the authors read and approved the final manuscript.

Authors' information
US MSc student in Research methodology, Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani, Thailand. AU: Lecturer in the Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani, Thailand. MK: Assistant professor in the Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani, Thailand.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Consent for publication was not required for this study because this study used secondary data.

Ethics approval and consent to participate
This study utilized secondary data from Nepal Demographic and Health Survey, 2011; therefore, no ethical approval was required. However, for the data access, permission from DHS measure was obtained.

References
1. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
2. Bethlehem J. Applied Survey Methods: A Statistical Perspective. Hoboken: John Wiley and Sons, Ltd; 2009.
3. Lohr S. Sampling: Design and Analysis. 2nd ed. Boston: Cengage Learning; 2010.
4. Schafer JL, Graham JW. Missing data: Our View of the State of the Art. Psychol Methods. 2002;7:147–77.
5. Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2nd ed. New York: Wiley Series in Probability and Statistics; 2002.
6. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc. 1988;83:1198–202.
7. Bennett DA. How can I deal with missing data in my study?Aust N Z J. Public Health. 2001;25:464–9.
8. De Waal T, Pannekoek J, Scholtus S. Handbook of Statistical Data Editing and Imputation. Hoboken: John Wiley and Sons, Ltd; 2011.
9. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
10. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. Lancet. 2014; 384:189–205.
11. Robles A, Goldman N. Can accurate data on birthweight be obtained from health interview surveys? Int J Epidemiol. 1999;28:925–31.
12. Blanc AK, Wardlaw T. Monitoring low birth weight: an evaluation of international estimates and an updated estimation procedure. Bull World Health Organ. 2005;83:178–85.
13. Ministry of Health and Population (MOHP). New ERA, ICF International Inc. Nepal Demographic and Health Survey. Kathmandu: Ministry of Health and Population, New Era, and ICF International; 2012.
14. Khanal V, Sauer K, Karkee R, Zhao Y. Factors associated with small size at birth in Nepal: further analysis of Nepal Demographic and Health Survey 2011. BMC Pregnancy Childbirth. 2014;14:32.
15. Khanal V, Zhao Y, Sauer K. Role of antenatal care and iron supplementation during pregnancy in preventing low birth weight in Nepal: comparison of national surveys 2006 and 2011. Arch Public Health. 2014;72:4.
16. Ministry of Health and Population. Annual Report 2009/2010. Kathmandu: Department of Health Services, Ministry of Health and Population, Nepal; 2010.
17. Gomella TL, Cunningham MD, Eyal FG, Zenk KE. Neonatology: Management, Procedures, On-Call Problems, Diseases and Drugs. 5th ed. New York: McGraw-Hill Medical; 2004.
18. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 10th revision. Geneva: World Health Organization; 2004.
19. van Buuren S. Groothuis-Oudshoorn K.mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011;45:1–67.
20. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20:40–9.
21. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods. 2001;6:330.
22. Schafer JL. Analysis of incomplete multivariate data. New York: Chapman and Hall; 1997.
23. Schafer JL, Olsen MK. Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. Multivar Behav Res. 1998;33:545–71.
24. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30:377–99.
25. von Hippel PT. Regression With Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data. Sociol Methodol. 2007;37:83–117.
26. Sullivan TR, Salter AB, Ryan P, Lee KJ. Bias and Precision of the "Multiple Imputation, Then Deletion" Method for Dealing With Missing Outcome Data. Am J Epidemiol. 2015;182:528–34.

Singh *et al. BMC Pregnancy and Childbirth* (2017) 17:67

Page 10 of 10

27. Lumley T. mitools: Tools for multiple imputation of missing data, R package version 2.0. 2010.
28. Boerma JT, Weinstein KI, Rutstein SO, Sommerfelt AE. Data on birth weight in developing countries: can surveys help?Bull World Health. Organ. 1996;74:209–16.
29. von Hippel PT. How To Impute Squares, Interactions, and Other Transformed Variables. Sociol Methodol. 2009;39:265–91.
30. Lumley T. Complex Surveys: A Guide to Analysis Using R. Washington: John Wiley and Sons Inc.; 2010.
31. Lee ES, Forthofer RN, editors. Analyzing Complex Survey Data. Sage Publications Inc. 2005.
32. Chu SY, Kim SY, Lau J, Schmid CH, Dietz PM, Callaghan WM, et al. Maternal obesity and risk of stillbirth: a metaanalysis. Am J Obstet Gynecol. 2007;197:223–8.
33. Han Z, Mulla S, Beyene J, Liao G, McDonald SD. Maternal underweight and the risk of preterm birth and low birth weight: A systematic review and meta-analyses. Int J Epidemiol. 2011;40:65–101.
34. Fulda KG, Kurian AK, Balyakina E, Moerbe MM. Paternal race/ethnicity and very low birth weight. BMC Pregnancy Childbirth. 2014;14:385.
35. Kelly Y, Panico L, Bartley M, Marmot M, Nazroo J, Sacker A. Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. J Public Health (Oxf). 2009;31:131–7.
36. Sharma A, Kader M. Effect of Women's Decision-Making Autonomy on Infant's Birth Weight in Rural Bangladesh. ISRN Pediatr. 2013;2013:159542.
37. Sreeramareddy CT, Shidhaye RR, Sathiakumar N. Association between biomass fuel use and maternal report of child size at birth - an analysis of 2005–06 India Demographic Health Survey data. BMC Public Health. 2011;11:403.
38. Ministry of Health and Population (MOHP). New ERA, ICF International Inc. Nepal Demographic and Health Survey 2006. Calverton: Ministry of Health and Population, New Era, and ICF International; 2007.
39. Karkee R, Lee AH, Khanal V. Need factors for utilisation of institutional delivery services in Nepal: an analysis from Nepal Demographic and Health Survey, 2011. BMJ Open. 2014;4:e004372.
40. Sreeramareddy CT, Joshi HS, Sreekumaran BV, Giri S, Chuni N. Home delivery and newborn care practices among urban women in western Nepal: a questionnaire survey. BMC Pregnancy Childbirth. 2006;6:27.
41. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th ed. Allyn and Bacon. 2012.