SSM - Population Health 19 (2022) 101220

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Data quality of birthweight reporting in India: Evidence from cross-sectional surveys and service statistics

Sayeed Unisa [a], Preeti Dhillon [b,*], Enu Anand [c], Harihar Sahoo [d], Praween K. Agarwal [e]

[a] Dept. of Biostatistics and Epidemiology, International Institute for Population Sciences (IIPS), Mumbai, India
[b] Dept. of Survey Research and Data Analytics, IIPS, Mumbai, India
[c] Doctoral Fellow, IIPS, Mumbai, India
[d] Dept. of Family and Generations, IIPS, Mumbai, India
[e] IPE Global Ltd, New Delhi, India

ARTICLE INFO

Keywords:
Data quality
Birthweight
Low birthweight
Heaping
Underreporting

ABSTRACT

The study aims to assess the quality of birthweight data collected in two surveys, including the National Family Health Survey (NFHS) and the Comprehensive National Nutrition Survey (CNNS), and as reported in the statistics from the Health Management Information System (HMIS). The study also aims to assess the implications of the data on the estimates of low birthweight (LBW). The percentage of newborns whose birthweight is missing continues to be high in the recent surveys (NFHS-4: 22%, CNNS: 30%) despite an improvement from 66% in NFHS-3. The under-coverage of birthweight data in HMIS is around 40%. In the surveys, the percentage of missing data on birthweight is higher among newborns belonging to poor households, Scheduled Tribes, and Scheduled Castes. Irrespective of whether birthweights are reported from the health cards or from mother's recall, there's a high reporting at multiples of 500g and heaping at 2,500g. The prevalence of missing data on birthweight and of heaping is higher among children born at home in comparison to facility-based births. Birthweight data of dead children who were more likely to have had a lower birthweight is highly underreported. The paper demonstrates state-level variations in birthweight reporting and inconsistencies across surveys and HMIS. In 2015–16, the prevalence of LBW as per HMIS data was 12.5%, whereas during the same period, NFHS-4 and CNNS reported a prevalence of 18%. The findings suggest that LBW is likely to be underestimated when missing data as well as heaping at 2,500g are highly prevalent. To generate robust LBW estimates in India, there is an urgent need to devise methods to ensure coverage of all live births (including early neo-natal deaths) as well as the stillbirths, irrespective of the facility where the deliveries take place.

1. Introduction

Birthweight is a strong predictor of weight and height in early childhood, not only for low birthweight children but also for those of normal and high birthweight (Binkin et al. 1988). Low birthweight (LBW) is defined by the World Health Organization (WHO) as weight at birth less than 2500g and continues to be a significant public health problem with short- and long-term consequences. Across the world, an estimated 15% of all babies are born with a low birthweight (LBW), and South Asia accounts for 52% of the global burden of LBW (UNICEF--WHO, 2019). Globally, three of the five countries with a prevalence of LBW of over 20% are from South Asia. These include Pakistan, India, and Bangladesh (Vir, 2016). The situation of LBW in South Asia is so bad possibly because timely and accurate weighing of newborns is a low public health priority and far from a universal practice (Desai et al., 2016). UNICEF-WHO (2019) have estimated that birthweight data is not available for nearly 40 million newborns worldwide, more than half of whom live in sub-Saharan Africa and nearly 40% in South Asia.

Birthweight data is important to study the growth of children. Moreover, it is required to examine the burden of low birthweight on the society and to ascertain the impact of the ongoing programs on health and nutrition. In India, birthweight data is missing for a large proportion of live births according to a study that used the National Family Health Survey data from 2005 to 06 (Subramanyam et al., 2010). Birthweight is often recorded to a round figure of multiples of 500g and particularly at 2,500g to avoid any queries or follow-up management efforts for

* Corresponding author. Dep. of Survey Research and Data Analytics, International Institute for Population Sciences (IIPS), Mumbai, 400088, India.
E-mail addresses: sunisa829@gmail.com (S. Unisa), pdhillon.iips@gmail.com (P. Dhillon), hariharsahoo@gmail.com (H. Sahoo), praween.agrawal.77@gmail.com (P.K. Agarwal).

https://doi.org/10.1016/j.ssmph.2022.101220
Received 27 January 2022; Received in revised form 28 August 2022; Accepted 29 August 2022
Available online 9 September 2022

improved perinatal and neonatal care (Blanc & Wardlaw, 2005). The reporting of the prevalence of LBW does not present the true picture of the implications of women's poor nutritional status on birth outcomes. LBW underestimates the problem of foetal growth restriction or intrauterine growth restriction (Fall, 2013). The Small for Gestational Age measure is considered more appropriate for the assessment of the problem of poor birth outcomes (Lawn et al., 2014; Lee et al., 2013; Qadir & Bhutta, 2009). For instance, nearly 47% of births in India are estimated to be SGA as against 28% that are reported as LBW (Black et al., 2013; IIPS, 2007).

There are several potential sources of bias in birthweight data. Only a little over half of all newborns are weighed at birth, and the gestational age is known for an even smaller proportion (Sreeramareddy et al., 2011). Likewise, there are errors in birthweight measurement and recording, including rounding to multiples of 500g, heaping of recorded birthweights at 2,500g, measurement after the first hour of life when significant weight loss is common, misclassification between live birth and stillbirth, survival bias, missing birthweight data due to home deliveries, and, for administrative data, lack of representation of births at public/private facilities. Studies have also found that there are significant differences in birthweight reporting from health cards and mother's recall (Channon et al., 2011; O'Sullivan et al., 2000; Shenkin et al., 2017). Those most vulnerable to biases of LBW data belong to the socioeconomically disadvantaged populations, who face a greater risk of giving birth to LBW babies. Overall, these biases are likely to result in the underestimation of LBW prevalence (Blencowe et al., 2019).

Low birthweight is a well-recognized indicator of progress towards sustainable development goals. Generating reliable LBW estimates at the national and the state levels is also essential for tracking the progress towards the achievement of the global nutrition target of a 30% reduction in LBW prevalence (Blencowe et al., 2019) and the *Poshan Abhiyaan* (Prime Minister's Overarching Scheme for Holistic Nutrition) target to reduce LBW prevalence by 2% per annum (Press Information Bureau, Government of India, 2020). Population-based nationally-representative surveys – namely, the National Family Health Survey (NFHS), the Rapid Survey on Children (RSOC), the Comprehensive National Nutrition Survey (CNNS), and the National Expanded Programme on Immunization – and the service statistics of the Health Management Information System (HMIS) are important sources of birthweight records in India. The statistics on the prevalence of LBW obtained from the survey data are used as inputs for a global model to calculate LBW, adjusting for possible biases in birthweight discussed earlier (UNICEF-WHO, 2019). But the global modeling estimates for LBW for India have not been accepted by the Ministry of Health and Family Welfare (MoHFW), Government of India, probably because the modeling or weighting techniques cannot address the extent of missing data, especially among the disadvantaged sections of the population (Subramanyam et al., 2010).

Despite its proven importance, accurate information on birthweight continues to lack in India. Moreover, studies on the quality of birthweight data in India are limited. Therefore, the present paper aims to analyse the quality of birthweight data collected in the recent surveys, namely NFHS and CNNS, as also of the service statistics data from the HMIS. It further attempts to analyse and discuss the implications of reporting on LBW estimates.

2. Material and methods

We analysed data from the following sources: a) the third and fourth rounds of the National Family Health Survey (NFHS 3–4) conducted during 2005-06 and 2015-16 respectively b) the Comprehensive National Nutrition Survey (CNNS) conducted during 2016-18 and c) the Health Management Information System (HMIS).

2.1. Survey data

NFHS, the Indian equivalent of the Demographic and Health Surveys (DHS), is an established source of representative data on population and health indicators at the national and state levels, with a special emphasis on maternal and child health outcomes. It utilizes standard model questionnaires widely used in more than 80 developing countries. On the other hand, CNNS is a specially designed survey for anthropometric measures and biochemical indicators for children and adolescents in the Indian population.

In the case of NFHS, birthweight information during 5 years preceding the survey was available for 19,250 out of 56,327 live births in NFHS-3 (IIPS and Macro International, 2007) and for 194,818 out of the 249,967 live births in NFHS-4 (IIPS and ICF, 2017). CNNS, on its part, recorded a total of 38,060 live births during 5 years preceding the survey; however, birthweight information was available for 29,362 children only. The descriptive analysis used survey analytic methods that account for clustering by primary sampling units and the appropriate sampling weights.

Each woman respondent (mother) interviewed was asked to provide a detailed birth history for all the births during the 5 years preceding the survey. Women who reported a live birth were asked whether the child was weighed at birth; those who replied with a 'yes' were asked to report the birthweight of the child. Stillbirths were excluded since NFHS and CNNS record birthweight only for live births. Birthweight was obtained from the health card but in case of those who did not have the health card, self-reported information was recorded. One of the major concerns with birthweight information obtained from survey data is the missing cases, that is, when a respondent provides no information for a particular item. Missing data can reduce the representativeness of a sample.

2.2. Service statistics

The Health Statistics Information Portal facilitates the flow of information on physical and financial performance from the district level to the state headquarters on to the centre using a web-based Health Management Information System (HMIS) interface. The portal provides periodic reports on the status of the health sector. More specifically, it provides information about the reported number of live births and the birthweight of the live births. The present study used HMIS data since 2009-10 to estimate the missing data on birthweight, defined as the percentage of live births whose birthweight was not reported, using the following method:

Missing data on birthweight = 100 - Coverage, where

$$\text{Coverage (of live births in HMIS)} = \frac{\text{Total number of reported live births}}{\text{Estimated number of live births}^{\wedge}} * 100$$

^Estimated number of live births is the number of projected births computed through exponential projection using the total population of states from the population census and the crude birth rate for the respective states from the Sample Registration System (SRS).

2.3. Methods for assessment of data quality

We assessed a survey's quality of reporting birthweight data in two steps. First, we presented 'missing birthweight data,' which is defined as the percentage of newborns who were not weighed at birth or whose birthweight information was not provided by their mothers in the survey. The proportion of missing birthweight data was estimated across categories of covariates (state, region, maternal education, caste, household wealth, place of delivery, assistance during delivery, and infant mortality) from the three surveys.

Second, we assessed heaping, which is a phenomenon inherent in population surveys. Heaping refers to a pattern of misreporting in which the distribution of numbers reported by respondents, such as age or weight, shows implausibly large frequencies of particular values, usually ending in 0 or 5. We then compared the children's birthweight with data from health cards and maternal recall.

2.4. Method for imputing missing birthweight data and for adjusting for heaping

We applied the multiple imputation approach to get the complete data on birthweight and then fitted a normal distribution curve with the mean and the standard deviation of the imputed birthweight data to adjusting for the heaping pattern. By taking five imputations with 1000 random seeds to get reproducible results of multiple imputations, we obtained 66,501 incomplete/missing cases, which were imputed. A linear regression model with such predictors as mother's age, education, caste, perceived size at birth, birth order, multiple births, and place of delivery was fitted to arrive at the imputed birthweight data. A similar approach of multiple imputation has been adopted and suggested in prior studies on LBW estimates (Blencowe et al., 2019; Singh et al., 2017).

3. Results

3.1. Completeness of birthweight data

This section presents our findings on the completeness of birthweight data reporting in CNNS, two rounds of NFHS, and service statistics, that is, HMIS.

3.1.1. Completeness of birthweight data in large-scale surveys

The problem of missing data on birthweight decreased from 66% in 2005–06 (NFHS-3) to 22% in 2015–16 (NFHS-4). But this figure is still on the high side. Even the CNNS survey reported a high figure of 29%. While the improvement in the reporting of birthweight is visible in all the states of India (Fig. 1), Uttar Pradesh and Bihar have consistently reported a high percentage of missing data in all three surveys, whereas Kerala and Goa have reported the highest amount of information on birthweight. In 2005–06 (NFHS-3), the states with the highest birthweight missing information were Uttar Pradesh, Nagaland, Bihar, Jammu and Kashmir, and Jharkhand. By contrast, the states with the least missing information were Kerala (3%), Tamil Nadu (11%), Goa (15%), Mizoram (16%), and Maharashtra (29%). In 2015–16 (NFHS-4), the states and union territories with the highest percentage of missing information were Nagaland (62%), Arunachal Pradesh (49%), Uttar Pradesh (47%), Bihar (41%), and Meghalaya (37%). On the other hand, missing information was the least in Kerala, followed by Puducherry, Andaman and Nicobar Islands, Sikkim, Lakshadweep, Goa, and Telangana. In CNNS, the states with the least missing birthweight information were Kerala, Goa, Telangana, Karnataka, Tamil Nadu, and Odisha, whereas the bottom five states (that is, the states with the highest missing information) were Nagaland, Uttar Pradesh, Bihar, Arunachal Pradesh, and Manipur.

3.1.2. Completeness of birthweight data in HMIS

This study estimated missing LBW information and percentage of low birthweight from the annual reports of HMIS. The missing LBW information and the percentage of LBW, that is, the total number of reported

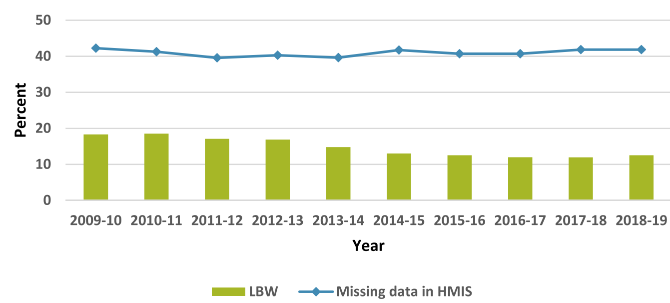


Fig. 2. LBW and missing birthweight data from HMIS, 2009–2019.
Source: Based on authors' compiled data from the HMIS annual reports

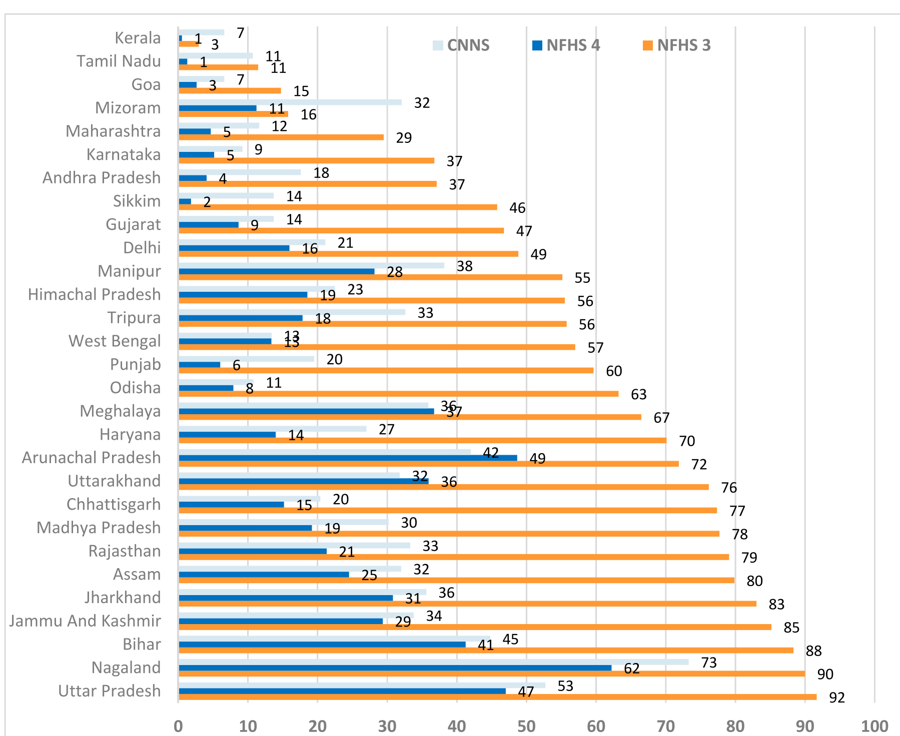


Fig. 1. Percentage of missing data on birthweight by state, NFHS-3 (2005–06), NFHS-4 (2015–16), and CNNS (2017–18).

live births with low weight to the total estimated live births from 2009 to 10 till 2018-19 are provided in Fig. 2. LBW estimates was around 12.5% in 2015–16, which remained same in 2018–19. The missing LBW information in HMIS data remained more or less the same, at around 40%, over the years. The state-wise pattern of missing birthweight data for the last ten years is presented in Table 7. The states of Uttar Pradesh (55%), Nagaland (54%), Madhya Pradesh (50%), Arunachal Pradesh (50%), Bihar (47%), and Odisha (46%) had the most incomplete information on birthweight for the year 2019. Owing to the incompleteness of birthweight information, the prevalence of low birthweight cannot be assessed accurately from the HMIS data. For instance, in 2018–19, the percentage of LBW children was 12.5, with a coverage of only 58% live births.

3.1.3. *Completeness of birthweight data by socioeconomic characteristics*

Fig. 3 illustrates the missing information on birthweight by wealth quintile. The missing information was the highest among the poorest across the three surveys. While there was only 6% of missing information on birthweight in the richest quintile, it was 42% among the poorest in NFHS-4. Similarly, the missing information was the highest among children born to mothers with no formal education or with lower educational levels. For instance, the missing information among children born to mothers without any formal level of education was 42% as opposed to only 7% among those born to mothers with 12 or more years of education. In all the surveys, the missing data on birthweight was the highest among the Scheduled Tribes, followed by the Scheduled Castes, Other Backward Classes, and others (Fig. 4).

3.1.4. *Completeness of birthweight data by place and type of personnel who assisted during delivery*

Missing birthweight information was the highest in the case of home-based births in all three surveys. The missing information for home-based births was 93% in NFHS-3, which declined to 72% in NFHS-4 and 75% in CNNS, indicating improvement in the reporting of birthweight for home births (Table 1). For all facility-based births, missing birthweight information in the case of government facilities declined from 25% in NFHS-3 to 10% in NFHS-4, while that in private health facilities declined from 22% in NFHS-3 to only 8% in NFHS-4. In NFHS-4, missing information was 48% when home-based births were assisted by a doctor/nurse/ANM/midwife. On the other hand, it was 80% when home-based births were assisted by a *dai* [traditional birth attendant (TBA)] or friends or relatives or any other person. Among all deliveries conducted in public health facilities, the percentage of missing data was the highest in the case of CHCs/rural hospitals/block PHCs (18.1% in CNNS; 12.7% in NFHS-4), followed by sub-centers (14.8% in NFHS-4; 11.6% in CNNS).

3.2. *Quality of reported birthweight data*

In the following section, the quality of birthweight data in the surveys is assessed by analyzing digit preference and the heaping pattern of data.

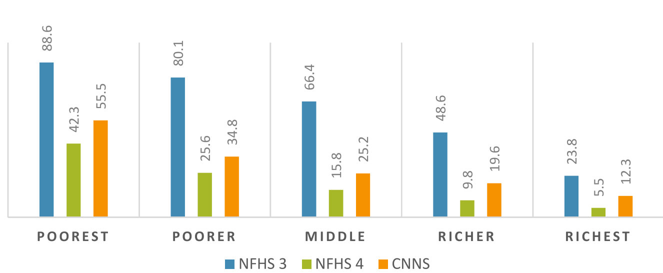


Fig. 3. Percentage of missing data on birthweight in India by wealth quintiles, NFHS-3 (2005–06), NFHS-4 (2015–16), and CNNS (2016–18).

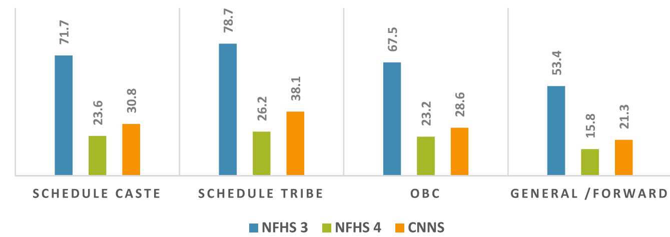


Fig. 4. Percentage of missing data on birthweight in India by caste, NFHS-3 (2005–06), NFHS-4 (2015–16), and CNNS (2016–18).

3.2.1. *Heaping in birthweight data*

The frequency distribution of birthweight data from NFHS-3, NFHS-4, and CNNS is presented graphically in Fig. 5a, b, and 5c, respectively. The analysis suggests a considerable and consistent heaping at certain numeric weights, particularly at multiples of 500, i.e., 2,000g, 2,500g, and 3,000g. Although heaping is an overall indicator of data quality, heaping at 2500 gm – the cut-off point for low birthweight – is most crucial to estimate the percentage of infants with a low birthweight (Blanc & Wardlaw, 2005). The amount of heaping at 2,500g could have a substantial effect on the estimation of LBW. Nearly 17–20% of newborns were reported to have weighed exactly 2,500g at birth (Table 2). If we assume that a certain proportion of the newborns reported as weighing 2,500g actually weighed less, some of the LBW babies would have been misclassified as having a normal birthweight.

3.2.2. *Heaping in birthweight data by place of delivery*

Heaping of birthweight data at multiples of 500g is observed irrespective of the place of delivery (Fig. 6). In the recent surveys (NFHS-4 and CNNS), heaping at exactly 2,500g was slightly higher for home-based births than for facility-based births, which was the opposite in NFHS-3. On the other hand, heaping at 3000 g was considerably more among home-based births than facility-based births. Heaping at 3500 g was observed to be more among facility-based births than home-based births in all three surveys. It is noteworthy that birthweight data reporting increased at 2,500g irrespective of place of delivery.

3.2.3. *Heaping differentials from the health card records and mother's recall*

In the recent surveys, including NFHS-4 and CNNS, only 54–56% of birthweight information was available from the health cards, whereas the rest was self-reported based on the memory of the mother (Fig. 7). In NFHS-3, reporting of birthweight from health cards was a lot less at 14%. Although there has been an increase in birthweight reporting via the health cards between the two rounds of NFHS, almost half of the birthweights are still self-reported. It is observed that mothers who reported the birthweight through recall were not always able to report the exact birthweight of their children.

Heaping is observed from both the sources of reporting in the data (Fig. 8). One would expect the birthweights reported from health cards to show less heaping than those recalled from memory; nevertheless, our analysis shows that this is not necessarily the case. Although health cards displayed less clustering, the birthweights were still highly heaped at 2,500g and 3,000g. This indicates that birthweight figures are often rounded by medical personnel who weigh a newborn and record its weight in a health card and report it to the mother as well as by mothers themselves when recalling the figure.

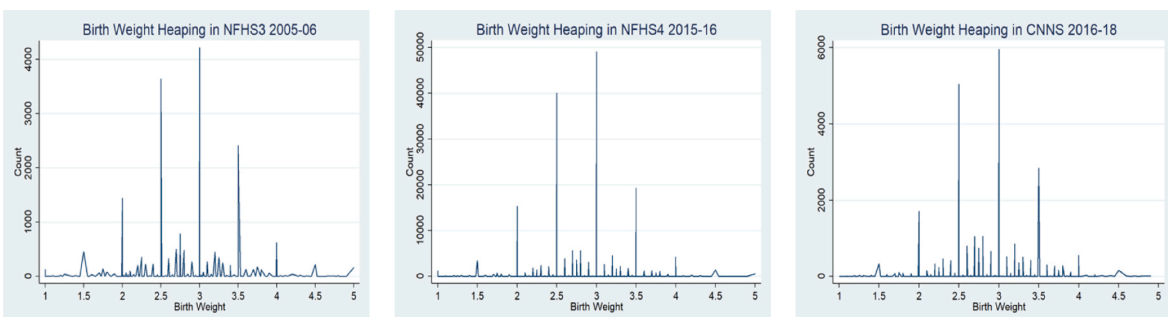
3.3. *Reporting of birthweight by survival status of infants*

Table 3 presents the survival status of infants by their birthweight in NFHS-3 and NFHS-4. Infant mortality was higher among newborns with a low birthweight. The percentage of missing information on birthweight was also considerably higher for children who died within a year of birth. Thus, the probability of dying within one year was more among

Table 1
Percentage of live births with missing birthweight (MBW) data by place of delivery and type of personnel who assisted during delivery at home.

Type of facility	NFHS 3		NFHS 4		CNNS	
	% MBW	Number of births	% MBW	Number of births	% MBW	Number of births
Public	25.0	10,166	9.5	1,30,199	14.6	18,436
Private	22.3	11,810	7.4	67, 599	12.1	9941
Home	92.7	34,461	72.4	52, 010	75.3	7091
Total	65.8	56,437	22.1	2,49, 809	26.0	35,468 *
Delivery at home assisted by						
Doctor	82.3	1711	47.5	5455	51.2	156
Nurse/ANM/Midwife	82.4	2467	42.5	5113	68.0	315
Other health personnel	84.7	591	62.7	1310	71.6	73
Dai/TBA	93.1	22,282	75.3	29,042	79.3	1430
Friends/Relatives	94.5	21,962	78.9	24,288	79.3	1421
Other	–	–	81.7	2742	75.3	186
No one	95.3	266	75.3	623	81.4	152
Delivery at public facility						
Govt./municipality	20.2	7650	6.9	589,623	14.2	12,355
Govt. dispensary	27.8	117	8.2	4143	9.5	1319
UHC/UHP/UFWC	21.0	102	7.8	4172	13.8	492
CHC/rural hospital/Block PHC	40.2	2100	12.7	42, 869	18.1	3264
PHC/additional PHC			10.2	17, 269	10.1	1286
Sub-centre	63.4	126	14.8	2490	11.6	231
Other public health facility	23.3	71	11.6	295	10.3	136

*1774 cases missing for births in health facilities.



(a) (b) (c)

Fig. 5. (a–c): Heaping in birthweight data in NFHS-3, NFHS-4, and CNNS.

Table 2
Percentage of birthweight data reported in multiples of 500g in India, NFHS-3, NFHS-4, and CNNS.

Birthweight	NFHS 3	NFHS 4	CNNS
500 g	0.0	0.0	0.1
1000 g	0.6	0.6	0.4
1500 g	2.2	1.8	1.1
2000 g	6.9	7.9	5.8
2500 g	17.4	20.7	17.2
3,000g	20.2	25.4	20.3
3500 g	11.5	10.0	9.7
4000 g	3.0	2.2	1.9

the LBW children. Thus, there is a concern with regard to survival bias in the survey data. Birthweight reporting from the card was also lower for dead children in both the rounds of the survey.

3.4. Implication of birthweight data quality on LBW estimation

In 2015–16, the LBW prevalence as per HMIS data was 12.5%, whereas NFHS-4 and CNNS, which were conducted around the same time, reported a much higher prevalence (18%). Table 4 gives a summary of the potential sources of bias in birthweight data and the implication of the bias on the estimation of LBW. Most of the factors have a potential influence on the underreporting of LBW estimates. However, very few factors like recall bias and instrument measurement errors may affect both sides of estimates.

3.4.1. LBW estimates after imputing missing data and adjusting for heaping

The linear regression model, as shown in Table 5, was considered for multiple imputations of LBW data using NFHS-4 data. All the predictors taken in the model, including mother's age, education, caste, perceived size at birth, birth order, multiple births, and place of delivery, were statistically significant and the model was fitted well ($p < 0.001$). Using the imputed data at the 5th imputation, the estimated LBW was 21.8% (95%CI: 21.52, 21.84), higher than the LBW of 18.2%, estimated based on reported data (95% CI: 18.04, 18.38) (Table 6). The effect of smoothing – that is, adjusting for heaping – was much more on the LBW estimates. Using the normal distribution of the imputed birthweight data (mean = 2.7817 kg, standard deviation = 0.5914 kg), the estimated LBW was 38.1% (95% CI: 37.89, 38.27).

4. Discussion

The problem of missing birthweight information is highly common in Indian health data, with birthweight unknown for at least one in five births as evident from NFHS-4 and CNNS, two of the recent surveys. Surveys done in some of the other countries have reported a similar proportion of births with no birthweight records (Singh et al., 2017). The challenges involved in utilizing birthweight information gathered from surveys cannot be ignored. Given that birthweight was reported for

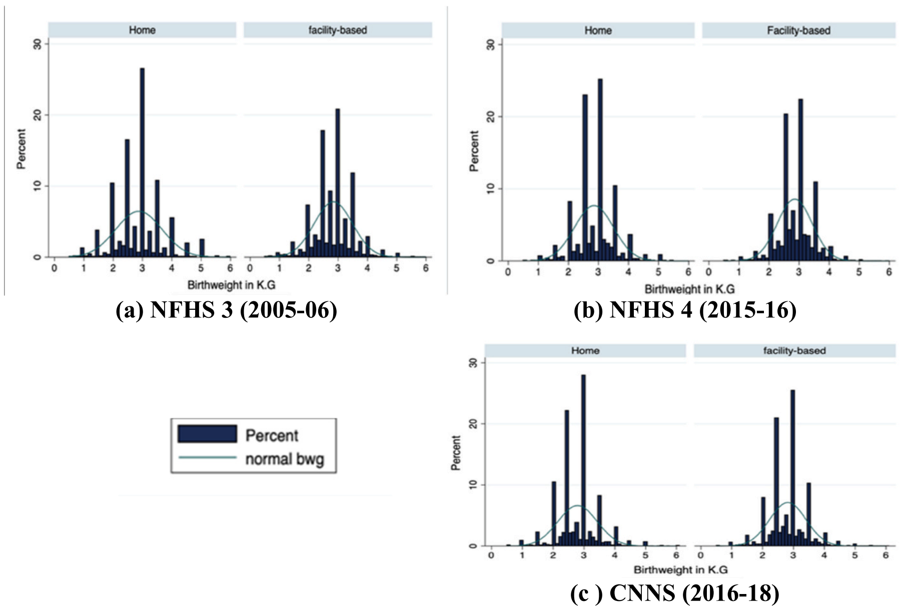


Fig. 6. (a–c): Heaping in birthweight data by type of facility.
Note: Bwg = Birthweight in grams.

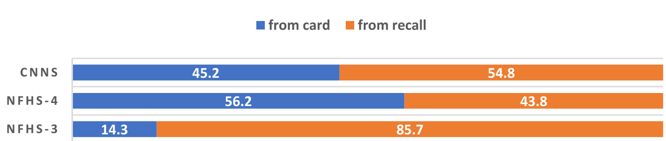


Fig. 7. Percentage of birthweight reporting by health cards and mother's recall, NFHS-3, NFHS-4, and CNNS.

only one-third of all births in NFHS-3 and around two-thirds of all births in NFHS-4 and CNNS, the results of birthweight should be interpreted with caution. For example, in the case of Uttar Pradesh, which represents 16% of the country's population but has birth records for only half of its children, the prevalence of LBW may be an underestimation and may be misleading. However, estimates from the survey data suggest that missing data on birthweight reduced between 2005 and 2018, indicating some improvement in the quality of data over time.

The estimates of birthweight data missing in HMIS have remained unchanged at around 40% in a decade, and the incompleteness of data makes these estimates questionable. A state-wise analysis of HMIS data showed an inconsistent pattern in the prevalence of LBW over 10 years, reaffirming that one has to be especially careful in estimating LBW using HMIS as the source of information (Appendix A2). One of the major limitations of the HMIS data is that it only provides data pertaining to the estimated number of births, the number of births reported, and the number of LBW babies born alive. Not all livebirths are reported and so it lacks representativeness. Other researchers in the past have also raised concerns over the HMIS data quality on account of completeness, timeliness, and reliability/accuracy (Husain et al., 2012; Pandey et al., 2010). The completeness of the data cannot be assured since the number of data elements reported against the total data elements is often unmatched. In most cases, the reported data elements are less than the actual data elements that should be presented. Also, the reporting from private facilities is poor. Timeliness is another important component of data quality. Studies show that many health facilities fail to submit the reports in time (Husain et al., 2012). Poor internet connectivity, lack of

Table 3
Percentage of LBW babies and percentage of newborns with MBW by survival status of infants, India, NFHS-3 and NFHS-4.

	Infant died	LBW (%)	% MBW	Birthweight reported from card
NFHS-4	Yes	37.4	48.4	39.4
	No	17.6	21.8	56.0
NFHS-3	Yes	41.2	79.0	6.1
	No	20.9	59.0	15.1

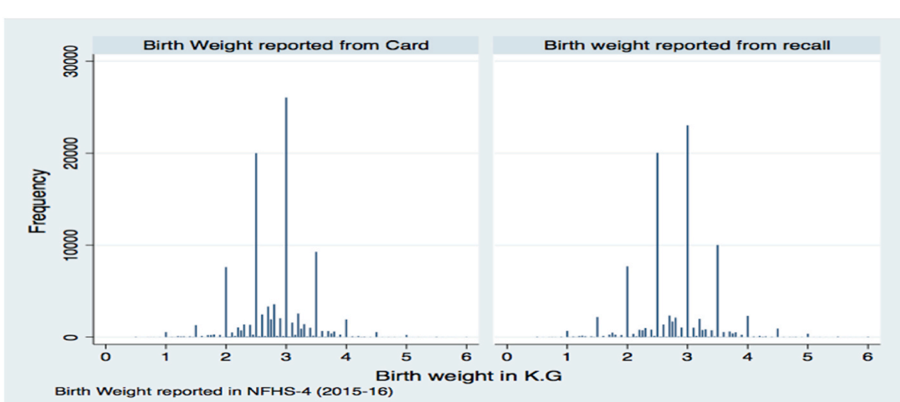


Fig. 8. Birthweight heaping by reporting from health cards and mother's recall, NFHS-4.

Table 4
Potential sources of bias in birthweight data and the implication of the bias on the estimation of LBW.

	Sources of bias and measurement error of BW data	Implication on LBW estimate
1.	Loss of birthweight data: bias in missing birthweight data • In the surveys, there is a bias in card retention. (Birthweight not available for babies who died and were more likely to have been LBW) • Babies who are extremely sick or who die soon are most likely classified as stillbirth. Their weight is often not measured. Moreover, these babies are more likely to be LBW. Missing service statistics for the severely sick babies (most often transferred immediately to ICU and weighed in the newborn ward).	Possible underreporting
2.	Missing data bias: Non-response pattern • Birthweight data for lower socioeconomic groups is more likely to be missing	Possible underreporting
3.	Coverage of weighing: bias in newborns weighed at birth • Many newborns are not weighed at birth, especially if born at home. • Newborns born at home are likely to be inadequately weighed given who measures their weight and which machines are used. Furthermore, there is a delay in taking them to a health facility for weight measurement.	Possible underreporting
4.	Measurement errors: individual/recording/weighing/heaping • Errors in birthweight measurement (poorly calibrated machines; outdated weighing machines; inaccurate measurements taken by the concerned personnel) • Inappropriate training of health staff responsible for measuring weight, leading to errors in birthweight measurement • Sub-optimal weighing practices (e.g. delay in weighing the newborn after birth; baby weighed while clothed) • Heaping of recorded birthweight exactly at 2,500g • Cross-sectional estimates of birthweight data are subject to recall bias	Both side possible error
5.	LBW from HMIS: Denominator calculation errors in the computation of LBW prevalence LBW is calculated as the number of births with a weight less than 2,500g for all live births (whether weighed or not)	Possible underreporting

Table 5
Linear regression model used for multiple imputations of LBW data.

Independent variables	Coefficient [95% CI]
Age	0.005***[0.004,0.005]
Mother's education	
No education®	
Primary	0.029***[0.021,0.037]
Secondary	0.076***[0.069,0.082]
Higher	0.127***[0.117,0.137]
Caste	
Scheduled Caste®	
Scheduled Tribe	0.103***[0.095,0.111]
OBC	0.028***[0.021,0.035]
General	0.038***[0.03,0.046]
Other/missing	0.029***[0.015,0.042]
Place of residence	
Urban®	
Rural	<0.001[-0.006,0.006]
Size at birth	
Very large/Larger than average®	
Average	−0.161***[-0.167,-0.154]
Smaller than average	−0.651***[-0.661,-0.642]
Very small	−1.114***[-1.13,-1.098]
Don't know/missing	−0.329***[-0.39,-0.268]
Birth order	
1®	
2	0.022***[0.016,0.028]
3	0.042***[0.034,0.05]
4+	0.061***[0.051,0.071]
Multiple births	
No®	
Yes	−0.555***[-0.574,-0.536]
Place of delivery	
Home®	
Public	0.019***[0.01,0.028]
Private	0.042***[0.032,0.052]
Other/missing	0.074***[0.02,0.128]
Constant	2.756***[2.736,2.775]

Note: ®reference category; ***$p < 0.01$.

Table 6
Estimates of LBW in India, 2015-16.

Multiple imputations based on linear regression model ##				
Variable	Observations per m			
	Complete	Incomplete	Imputed	Total
LBW	193,126	66,501	66,501	259,627

LBW estimates based on	**LBW (%)#**	Weighted N	Unweighted N
Reported birthweight data	18.2 [18.04, 18.38]	1,94,818	1,93,345
Imputed birthweight data##	21.8 [21.52, 21.84]	2,49,967	2,59,627
Imputed and smoothed birthweight data###	38.1 [37.89, 38.27]	2,49,967	2,59,627

Note: 219 cases were above 5.5 kg, which were considered missing and imputed; #95% CI in parentheses; ## in Multiple imputations: Number of imputations = 5, random seeds = 1000; ###by fitting normal distribution curve with mean 2.7817 kg, standard deviation 0.5914 kg and LBW is $P(Z_X < Z_{2.5})$.

essential hardware, lack of staff, lack of supervision, and poor training may explain the incompleteness and untimeliness of the data to some extent. Accuracy of the HMIS data, defined as the correctness of data collected in terms of the actual number of services provided or health events organized, has also drawn considerable attention. Accuracy errors may occur due to inadequate reporting, systematic errors, or data entry errors.

The present findings, which reveal a greater extent of missing birthweight data from lower socioeconomic groups in the surveys, have implications on the estimation of LBW. Similar to our results, Subramanyam et al. (2010) found that children from households in the lowest wealth quantile were underrepresented in birthweight data in 2005–06. However, recent research, using information on sites in Bangladesh, Ethiopia, Ghana, Guinea-Bissau, and Uganda, shows no variations in missing birthweight data by social status (Biks et al. 2021), though, this study suggested a better reporting of birthweight data from educated mothers. Birthweight information is often missing for the socioeconomically vulnerable groups in facility-based data as well. Pregnancies of women belonging to the disadvantaged sections of the society are most likely to result in LBW babies in both high-income (Martinson & Reichman, 2016) and low-/middle-income countries, including India (Mishra et al. 2021; Subramanyam et al. 2010). Also, home-based births are generally more prevalent among these women. Similar to Singh et al. (2017), the present study also indicates that missing birthweight information is the highest for live births at home. Considering that about 20% of childbirths in India occur at home, collecting information on LBW becomes especially complex. There is a strong linkage between disadvantaged populations, home-based births, and their birthweight reporting.

Our study also found that deliveries conducted at CHCs/rural hospitals/Block PHCs and sub-centers are more likely to have missing birthweight data. This suggests a need to check the availability of weighing machines at rural facilities and to train the grassroots-level

Table 7
Percentage of missing birth weight data in India and States HMIS (2009–2019).

	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17	2017–18	2018–19
All India	42	41	40	40	40	42	41	41	42	42
Andaman and Nicobar Islands	43	50	34	36	37	33	40	40	35	35
Andhra Pradesh	24	27	22	27	30	42	29	29	35	35
Arunachal Pradesh	73	69	69	59	56	56	54	54	49	49
Assam	51	46	43	41	39	37	38	38	40	40
Bihar	67	65	56	50	48	49	46	46	46	46
Chandigarh	25	26	OR	OR	OR	OR	OR	OR	OR	OR
Chhattisgarh	32	33	37	44	43	43	43	43	42	42
Dadra and Nagar Haveli	28	27	43	40	42	34	30	30	22	22
Daman and Diu	100	77	71	57	42	44	42	42	45	45
Delhi	61	60	47	40	40	38	37	37	38	38
Goa	35	36	19	32	40	35	31	31	32	32
Gujarat	39	32	31	32	33	32	30	30	25	25
Haryana	31	33	29	34	32	31	34	34	32	32
Himachal Pradesh	34	35	35	35	34	37	40	40	44	44
Jammu and Kashmir	41	48	49	40	37	38	37	37	36	36
Jharkhand	47	44	46	43	42	43	40	40	31	31
Karnataka	40	46	42	41	43	42	41	41	41	41
Kerala	25	25	26	26	24	24	28	28	25	25
Lakshadweep	50	60	43	51	47	49	36	36	39	39
Madhya Pradesh	39	40	43	48	48	48	47	47	50	50
Maharashtra	40	30	29	28	29	31	35	35	32	32
Manipur	37	35	29	28	28	29	32	32	33	33
Meghalaya	26	23	22	21	17	16	17	17	19	19
Mizoram	0	7	22	13	11	9	22	22	24	24
Nagaland	71	68	57	49	50	48	48	48	53	53
Odisha	44	44	36	38	37	37	41	41	45	45
Puducherry	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR
Punjab	39	38	29	29	29	31	34	34	37	37
Rajasthan	44	44	42	40	39	42	42	42	43	43
Sikkim	55	45	44	47	44	45	47	47	50	50
Tamil Nadu	24	25	31	33	32	35	38	38	38	38
Telangana					23	31	31	31	26	26
Tripura	40	35	30	25	25	26	36	36	34	34
Uttar Pradesh	46	46	47	50	50	54	50	50	55	55
Uttarakhand	49	51	47	43	38	39	42	42	46	46
West Bengal	32	28	26	28	25	29	31	31	34	34

Note: OR stands for over reporting of reported birthweight data over estimated number of life births.

staff on the importance of weighing newborns. Adding to the several challenges of gathering robust birthweight information is the non-availability of the timing of birthweight measurement of the newborns. Studies have documented the importance of weighing a newborn within 24 h of birth (Channon et al., 2011). The delay in the time of birthweight measurement may impact the exact prevalence of LBW. Newborns born with a LBW are more likely to be at the risk of infant mortality. There is a strong likelihood that the birthweight is missing for many live births that ended in early neonatal deaths.

The cross-sectional estimates of birthweight data are subject to recall bias and measurement errors. With an increase in the rate of institutional delivery, the survey data based on recall from mothers would reduce; however, in some states, it may persist and is a matter of concern for estimating LBW. The accuracy and quality of birthweight data reported in the health cards by health systems also raise concerns since a significant heaping at multiples of 500g, especially at 2,500g (Blanc & Wardlaw, 2005), the standard cut-off to identify low-birthweight, is observed in the health card records too. Such heaping points to the loopholes in the measurement of birthweight by health personnel, the precision of the measurement, and the quality and condition of scales used to measure birthweight. It is likely that there are no formal standards of recording birthweight within the health systems and, hence, the tendency to round birthweights in health cards (Channon et al., 2011). In a multi-country hospital-based study, weight heaping was found to reduce with a greater use of digital scales compared to analog scales (Kong et al., 2021).

Quality birthweight records, with minimal missing information, are vital for estimating the actual prevalence of LBW. Using sample measured birthweights without accounting for the missing values and the heaping of the observed values results in the underestimation of the prevalence. Prior studies have shown the relevance of using multiple imputations of missing birthweight in the estimation of LBW (Blencowe et al., 2019; Singh et al., 2017). Our study showed that the effect of heaping on LBW estimates is much higher. Using partial data for India, the LBW estimate for the South Asian region was 26.4% (18.6–35.2) (Blencowe et al., 2019). Our study applied a similar method, except that we assumed only one normal curve for smoothing (or adjusting for heaping) of Indian data, whereas they fitted two normal distributions on global birthweight data. In alignment with prior studies, the findings of the present study suggest a need for further studies to improve birthweight data quality and provide robust LBW estimates using MI and adjusting for heaping.

The present study found the prevalence of LBW estimated from HMIS to be considerably lower than that estimated from the three surveys. Therefore, there is a need to strengthen facility-based data reporting in service statistics. The increase in card-based birthweight reporting in the surveys will result in a better birthweight data over time. However, extra efforts are needed from health programmes to record good quality (accurate and reliable) data at the facility level.

5. Conclusion

The present study evaluated the quality of birthweight information by estimating the percentage of missing birthweight data and heaping at multiples of 500g in data collected through three large-scale national surveys, including NFHS-3, NFHS-4, and CNNS, as also data obtained from the service statistics of HMIS over the last one and a half decade. The findings of this study suggest that the currently available sources of birthweight information in India are inadequate to capture the actual prevalence of low birthweight as quite a few live births go unrecorded. Large amounts of missing birthweight information result in an underestimation of low birthweight, particularly in lower socioeconomic settings, and are likely to portray an overly optimistic picture of health of children. There is an urgent need to devise methods to ensure coverage of all births, whether live births (including early neo-natal deaths) or stillbirths, irrespective of the facility where the births take place, to generate robust birthweight data. The increasing trend of reduction in reporting birthweight data from recall in the surveys, along with the rise in institutional births, will enhance the completeness of birthweight data. However, programmatic efforts, such as providing a sufficient number of trained staff, increasing the resources in the facilities, and monitoring the reporting by health personnel, are needed to capture quality data in health cards at the facility level. The study concludes that missing and heaping of birthweight data tend to underestimate the LBW estimates. Therefore, programmatic efforts are required to get robust estimates of LBW in India.

Funding

This paper was prepared as part of a CNNS Knowledge Network Project of International Institute for Population Sciences (IIPS). The contents of this paper are the sole responsibility of the authors and do not necessarily reflect the views of IIPS, Government of India or IPE Global Ltd..

Authors contribution

Conceptualization SU, PA; Data curation EA; Formal analysis EA, SU, PD; Funding acquisition SU, PA; Methodology SU, RJ, PD; Supervision SU; Validation PD; Visualization EA, PD; Writing –EA, PD; Writing - review & editing SU, HS.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgements:

Authors are thankful to **Dr. Anjula Srivastava** for her comments on the prior version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2022.101220.

Abbreviations

ANM	*Auxiliary Nursing Midwifery*
BCG	*Bacille Calmette-Guerin*
CHC	*Community Health Centre*
CNNS	*Comprehensive National Nutrition Survey*
HMIS	*Health Management Information System*
LBW	*Low birthweight*
MBW	*Missing Birthweight*
MoHFW	*Ministry of Health and Family Welfare*
NFHS	*National Family Health Survey*
PHC	*Public Health Centre*
UFWC	*Urban Family and Welfare Centre*
UHC	*Urban Health Centre*
UHP	*Urban Health Post*
UNICEF	United Nations Children's Fund
WHO	*World Health Organization*

References

Biks, G. A., Blencowe, H., Hardy, V. P., et al. (2021). Birthweight data completeness and quality in population-based surveys: EN-INDEPTH study. *Population Health Metrics, 19*, 17. https://doi.org/10.1186/s12963-020-00229-w

Binkin, N. J., Yip, R., Fleshood, L., & Trowbridge, F. L. (1988). Birthweight and childhood growth. *Pediatrics. 1988 Dec, 82*(6), 828–834.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., De Onis, M., Ezzati, M., Grantham-Mcgregor, S., Katz, J., Martorell, R., & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet, 382*(9890), 427–451. https://doi.org/10.1016/S0140-6736(13)60937-X

Blanc, A. K., & Wardlaw, T. (2005). Monitoring low birthweight: An evaluation of international estimates and an updated estimation procedure. In *Bulletin of the world health organization* (Vol. 83, pp. 178–185), 3.

Blencowe, H., Krasevec, J., de Onis, M., Black, R. E., An, X., Stevens, G. A., Borghi, E., Hayashi, C., Estevez, D., Cegolon, L., Shiekh, S., Ponce Hardy, V., Lawn, J. E., & Cousens, S. (2019). National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: A systematic analysis. *Lancet Global Health, 7*(7), e849–e860. https://doi.org/10.1016/S2214-109X(18)30565-5

Channon, A. A. R., Padmadas, S. S., & McDonald, J. W. (2011). Measuring birthweight in developing countries: Does the method of reporting in retrospective surveys matter? *Maternal and Child Health Journal, 15*(1), 12–18. https://doi.org/10.1007/s10995-009-0553-3

Desai, S., Haddad, L., Chopra, D., & Thorat, A. (2016). *Undernutrition and public Policy in India: Investing in the future. The state of undernutrition in India (Part1)*. Routledge India.

Fall, C. H. D. (2013). Fetal malnutrition and long-term outcomes. *Nestle Nutrition Institute Workshop Series, 74*, 11–25. https://doi.org/10.1159/000348384

Husain, Z., Saikia, N., & Bora, R. S. (2012). *Opportunities and challenges of health management information system in India: A case study of uttarakhand*. https://mpra.ub.uni-muenchen.de/id/eprint/40014.

International Institute for Population Sciences (IIPS) and ICF. (2017). *National family health survey (NFHS-4) 2015-16 India*. Mumbai: IIPS.

International Institute for Population Sciences (IIPS) and Macro International. (2007). *National family health survey (NFHS-3), 2005–06: India, ume I*. Mumbai: IIPS.

Kong, S., Day, L. T., Zaman, S. B., et al. (2021). Birthweight: EN-BIRTH multi-country validation study. *BMC Pregnancy and Childbirth, 21*, 240. https://doi.org/10.1186/s12884-020-03355-3

Lawn, J. E., Blencowe, H., Oza, S., You, D., Lee, A. C. C., Waiswa, P., Lalli, M., Bhutta, Z., Barros, A. J. D., Christian, P., Mathers, C., & Cousens, S. N. (2014). Every newborn: Progress, priorities, and potential beyond survival. *The Lancet, 384*(9938), 189–205. https://doi.org/10.1016/S0140-6736(14)60496-7

Lee, A. C. C., Katz, J., Blencowe, H., Cousens, S., Kozuki, N., Vogel, J. P., Adair, L., Baqui, A. H., Bhutta, Z. A., Caulfield, L. E., Christian, P., Clarke, S. E., Ezzati, M., Fawzi, W., Gonzalez, R., Huybregts, L., Kariuki, S., Kolsteren, P., Lusingu, J., … Black, R. E. (2013). National and regional estimates of term and preterm babies born small for gestational age in 138 low-income and middle-income countries in 2010. *Lancet Global Health, 1*(1). https://doi.org/10.1016/S2214-109X(13)70006-8

Martinson, M. L., & Reichman, N. E. (2016). Socioeconomic inequalities in low birth weight in the United States, the United Kingdom, Canada, and Australia. *American Journal of Public Health, 106*(4), 748–754. https://doi.org/10.2105/AJPH.2015.303007

Mishra, P. S., Sinha, D., Kumar, P., Srivastava, S., & Bawankule, R. (2021). Newborn low birth weight: Do socio-economic inequality still persist in India? *BMC Pediatrics, 21*(1), 518. https://doi.org/10.1186/s12887-021-02988-3

O'Sullivan, J. J., Pearce, M. S., & Parker, L. (2000). Parental recall of birthweight: How accurate is it? *Archives of Disease in Childhood, 82*(3), 202–203. https://doi.org/10.1136/adc.82.3.202

Pandey, A., Roy, N., Bhawsar, R., & Mishra, R. M. (2010). Health information system in India: Issues of data availability and quality. *Demography India, 39*(1), 111–128. https://www.researchgate.net/publication/232084914.

Press Information Bureau, Government of India. Ministry of women and child development. Poshan abhiyaan. 6 March, March 2020. Retrieved from https://pib.gov.in/newsite/PrintRelease.aspx?relid=199916. (Accessed 23 January 2022).

Qadir, M., & Bhutta, Z. A. (2009). Low birthweight in developing countries. In *Small for gestational age. Karger* (Vol. 13, pp. 148–162). https://doi.org/10.1159/000165998

Shenkin, S. D., Zhang, M. G., Der, G., Mathur, S., Mina, T. H., & Reynolds, R. M. (2017). Validity of recalled v. recorded birthweight: A systematic review and meta-analysis. *Journal of Developmental Origins of Health and Disease, 8*(2), 137–148. https://doi.org/10.1017/S2040174416000581

Singh, U., Ueranantasun, A., & Kuning, M. (2017). Factors associated with low birth weight in Nepal using multiple imputation. *BMC Pregnancy and Childbirth, 17*, 67. https://doi.org/10.1186/s12884-017-1252-5

Sreeramareddy, C. T., Shidhaye, R. R., & Sathiakumar, N. (2011). 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, 11*(1), 403. https://doi.org/10.1186/1471-2458-11-403

Subramanyam, M. A., Ackerson, L. K., & Subramanian, S. V. (2010). Patterning in birthweight in India: Analysis of maternal recall and health card data. *PLoS One, 5* (7), Article e11424. https://doi.org/10.1371/journal.pone.0011424

UNICEF-WHO. (2019). Low birthweight estimates. In *UNICEF data: Monitoring the situation of children and women*. Retrieved from https://data.unicef.org/topic/nutrition/low-birthweight/. (Accessed 23 January 2022).

Vir, S. C. (2016). Improving women's nutrition imperative for rapid reduction of childhood stunting in South Asia: Coupling of nutrition specific interventions with nutrition sensitive measures essential. *Maternal and Child Nutrition, 12*, 72–90. https://doi.org/10.1111/mcn.12255