Working paper 1, version 1. Deposited in UCL Discovery 9 April 2020

Technical report

Trial experience and data capture in the Low Birth Weight South Asia Trial, a large cluster-randomised controlled trial in lowland Nepal

Short title: Investigating data capture in a cluster-randomised controlled trial in Nepal.

Authors: Naomi M. Saville[1,*], Dharma S. Manandhar[3], Jonathan C. Wells[2]

1. Institute for Global Health (IGH), University College London (UCL), London, UK
2. Great Ormond Street Institute of Child Health (ICH), University College London (UCL), London, UK
3. Mother and Infant Research Activities (MIRA), Kathmandu, Nepal

* Corresponding author
Email: n.saville@ucl.ac.uk (NS)
Dharma S. Manandhar dsm@mira.org.np
Jonathan Wells jonathan.wells@ucl.ac.uk

Financial disclosure

Primary data collection for the Low Birth Weight South Asia Trial (LBWSAT) was funded by UKaid from Department for International Development (DFID) South Asia Research Hub (PO 5675) (https://www.gov.uk/dfid-research-outputs/protocol-of-the-low-birth-weight-south-asia-trial-lbwsat). Analyses conducted here were funded by The Leverhulme Trust as part of a project entitled "Age at marriage and human capital outcomes in mothers and offspring in Nepal" (Leverhulme project number RPG-2017-264 https://www.leverhulme.ac.uk/listings?field_grant_scheme_target_id=13). This research was partly supported by the National Institute for Health Research (NIHR) Great Ormond Street Hospital Biomedical Research Centre (https://www.gosh.nhs.uk/our-research/nihr-great-ormond-street-hospital-brc). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

JW and NS conceived the paper and the analyses. NS lead data collection on the ground, conducted all the analyses and wrote the first draft of the paper for publication. DSM managed the field data collection team in Nepal. All authors reviewed the manuscript, provided input and approved the paper for publication.

Ethics statement

Research ethics approval for these analyses was obtained from the Nepal Health Research Council (292/2018) and the UCL Ethical Review Committee (0326/015). Ethical approval for the original study was obtained from the Nepal Health Research Council (108/2012) and the UCL Ethical Review Committee (4198/001). The Government of Nepal Ministry of Health and the Social Welfare Council granted permission for the primary data collection to be conducted.

Corresponding Author: Naomi M. Saville, PhD.

Date of upload: 9 April 2020

Abstract

Objectives: i) to describe data capture in the Low Birth Weight South Asia Trial (LBWSAT) and factors affecting it; ii) to analyse to what extent differential data capture created bias in the available data.

Methods: We describe the context, study design, data collection instruments used and their capture rates. Little of the data available were eligible for trial analyses, so use of the data for secondary analyses is important. Data capture was affected by data collector overload, pressure to enrol women in the food and cash transfer arms, delayed receipt of participant ID cards, enrolment of women at any gestational age (including after delivery at the start), in-migration into the food/cash arms to access transfers, political instability, conflict in the field team, logistical issues, establishment of a run-in period, hiatus of data collection during training, and lack of funds to extend the duration of the study.

To assess the extent that differential data capture generated bias we described background characteristics by study arm and in-migration status. Then for each of the main data collection instruments we compared captured and not-captured enrolled women's age, age at marriage, wealth score and height using t-tests and enrolled women's and husband's education using chi squared tests. Using mixed logistic regressions (adjusted for clustering using random effects) we assessed the odd of questionnaire capture in relation to these factors.

Results: Small differences between captured and non-captured women were found. In-migrators were more prevalent in the cash/food transfer arms and compared with permanent residents were more likely to be living in their parental homes, younger, primigravida, adolescent, Muslim, slightly poorer and have some education. Analyses of captured and non-captured women by questionnaire revealed small differences in age, age at marriage, wealth score and education but mostly these differences were very small. The largest differences were between captured and non-captured women in the endline cross-sectional survey, when slightly older, less educated, poorer women were more likely to be captured. These women more likely to report to a measuring station in their community.

Conclusions: Many challenges in implementing large-scale trials in the plains of Nepal affect rates of data capture, especially when several timebound follow-up data collection occasions are needed. Although in-migrated and permanent residents, and captured and non-captured women differed slightly, the differences were not large enough to be of concern.

Introduction

The Low Birth Weight South Asia (LBWSAT) was a cluster-randomised controlled trial (cRCT) implemented with 24682 women and 25,090 pregnancies in the plains of Nepal between 2012 and 2015. The study tested the impact of 3 community-based interventions upon birthweight and child nutritional status within the first 16 months of life.

Aims of this paper

The aim of this working paper is to examine data capture in the Low Birth Weight South Asia Trial (LBWSAT), to help evaluate potential biases arising from use of the data to address secondary research questions on an observational basis. We report data capture rates and any differences between captured and non-captured respondents that might affect further analyses on new research questions.

Low Birth Weight South Asia Trial (LBWSAT) results already reported

Details of the trial are documented in the trial protocol [1], a description of the electronic data collection system [2] and the trial paper which presents the impact of trial interventions on primary and secondary outcomes [3]. A study of intra-household food allocation [4] using a validated photographic food atlas [5] and an electronic data collection form [6] documented determinants of intra-household food allocation [7] and measured the impact of the trial upon intra-household food allocation [8]. Agency and empowerment were measured in a sub-sample of enrolled women using a validated questionnaire [9], and the impact of trial interventions upon markers of agency was tested [10]. A qualitative investigation of cash transfers in the households demonstrated that women were able to utilise the money for improved nutritional intake or other purposes of their choosing but that complex power relations within families affected this [11, 12]. Women's group facilitators and the group members to some extent 'compelled' the women to spend their cash on the intended purpose by 'soft conditioning' [13]. Formative research informed intervention design [14] and process evaluation findings on the interventions have also been reported [15].

Over twenty-five thousand pregnant women enrolled in the study and 19,017 mothers and 19,222 children were measured during an endline cross-sectional survey. The primary trial analysis of birth weight excluded in-migrators and covered a 'full trial period' between 5 June 2014 and 31 March 2015 only. Compared with the *control* arm (n=464), the *PLA + food* arm showed a significant 78·0 g (95% CI 13·9, 142·0; n=626) increase in birthweight measured within 72 hours, while increases of 50·5 g (95% CI -15·0, 116·1; n=509) in the *PLA + cash* arm and 28·9 g (95% CI -37·7, 95·4; n=488) in the *PLA women's group-only* arm were not significant. When children born of the trial between 5 June 2014 and 19 June 2015 were followed up cross-

sectionally at trial endpoint (between 20 June and 19 Oct 2015) there was no impact of trial interventions on any child or maternal anthropometric indicators [3].

Recasting the LBWSAT dataset as an observational resource

Whilst the trial is well documented, the primary results are available in the papers cited above and secondary analyses from the trial are emerging [16-18], these publications did not have the space or scope to describe the study population in sufficient detail for the purposes of secondary investigations. Also, because of constraints upon inclusion in cRCT analyses, data from many women enrolled and surveyed could not be used in the primary analyses because the participants had insufficient exposure to the trial interventions and /or were in-migrators. In addition, largely only primary and secondary outcome variables have been analysed to date, while many variables that were collected to understand the mechanisms by which interventions may have worked or not worked remain unused. These data provide a rich resource for observational analyses to answer novel and important research questions. While undertaking such secondary analyses, it is important to assess all potential sources of bias associated with the different data collection instruments.

The context of Dhanusha and Mahottari

At the time of data collection, Dhanusha and Mahottari in the central plains were two of the 75 districts of Nepal, and our study population covered 80 Village Development Committee (VDC) clusters in the southern parts of the districts bordering Bihar state in India. Following Nepal's federalisation, these 80 VDCs now cover 28 out of 160 rural and urban municipalities of Province 2. The population in these areas mostly lives in rural villages connected by dirt or gravel roads within 2-3 hours' drive from the sub-metropolitan city of Janakpur.

Maithili-speaking society, being relatively conservative and resistant to change, tends to exert more control over married women's movements than other sectors of Nepali society. Young married women in particular are often not permitted to leave the confines of the home compound and if they do, they usually cover their face as they move around. Their participation in decisions is highly restricted [19, 20]. Early marriage is still the norm in this community and many women begin childbearing during adolescence [16, 21]. A traditional patrilocal system dictates that the daughter-in-law goes to live in her marital (husband's) home or "*sasural*" upon marriage and becomes a member of that household. She is permitted to visit her parental home or "*maiti*" for festivals and often around the time of her delivery, especially for the first baby. Family ties between *sasural* and *maiti* are often maintained by the *maiti* sending gifts of food or cash to the sasural. Whilst living in extended families is still the norm, there is an emerging trend towards separation into nuclear family units where families can afford it. Most females either have no education or

leave school before reaching secondary education, while those males that attend school tend to stay in school longer.

The traditional caste system dictates livelihood activities for many who follow occupations typical of their caste, such as farming, pottery, basket-making, leather work, thatching, building, hair-cutting, trading, oil pressing, sweet making, flower garland making and so on [22]. Despite caste discrimination being illegal, Dalits are still treated as untouchable and are highly marginalised [23] and Muslims also suffer from marginalisation [24]. Agricultural livelihoods predominate with landowners practising mixed farming with arable crops and livestock, while the landless work on their fields as labourers paid in food or cash. The better-off provide land to poorer households under share cropping agreements, whilst middle wealth groups farm their own land. The most important crops are one or two cropping cycles of rice during the monsoon/autumn and a winter crop of wheat, supplemented by smaller areas of pulses (e.g. lentils, pigeon pea), oil seeds and vegetables.

The city of Janakpur is the birthplace of the Hindu goddess Sita and the site of her marriage to the god Ram, so is an important pilgrimage location for Hindus from Nepal and India. In association with this, lacto-vegetarianism is relatively common, though the majority consume flesh foods occasionally, especially fish, which is available from the many ponds which are used for fish farming. Cow and buffalo meat are taboo and not consumed by most Hindus, so are not available in the area. Very few households consume eggs, chicken or pork, while goat meat and fish are the preferred flesh foods eaten occasionally.

Labour migration to the Middle East, Malaysia and India is extremely common especially amongst young men in Nepal, accounting for 29% of GDP in 2014 [25]. Indeed, Province 2 has the highest migration rates of anywhere in the country [26]. Most migratory labourers travel on 2 to 3-year contracts, while a poorer minority migrate to India seasonally, mostly for agricultural work. This means that many reproductive age women are living separately from their husbands for 2-3 years at a stretch during each cycle of migration, and that there is a shortage of able-bodied younger men in society.

Why investigate human capital and factors influencing it in the central plains of Nepal?

This part of Nepal makes an interesting location to investigate human capital and factors influencing it because the population offers a broad range of characteristics with variability in terms of women's agency and empowerment, education, wealth, marginalisation and nutritional status. Young married women are particularly disempowered in this context [20], however levels of empowerment also vary [11, 12, 27]. Age at marriage is low, at 15 years on average. Levels of education are very low and a differential exists between men and women as well as between different wealth / status groups [28]. Underweight and short

stature of women is common and undernutrition of children, especially wasting, is prevalent, despite food insecurity levels being relatively low [28]. Inequalities in terms of wealth and status are wide. This means that we are able to ask novel questions about associations between different aspects of human capital and covariates such as household factors and measures of relative empowerment / agency as well as analysing many aspects of nutrition in pregnancy and the first 2 years of life.

Methods

Cluster selection within Dhanusha and Mahottari districts

80 Village Development Committee population clusters (each comprising 9 ward sub-units) were block-randomised (20 clusters each) to *control* or one of three intervention arms. The randomisation was stratified on the basis of cluster size and motor accessibility, using four strata (large accessible, large inaccessible, small accessible and small inaccessible). Clusters were eligible for inclusion if they had a predominantly Maithili-speaking population, were not located near the East-West highway, did not have significant forest cover, had not previously had PLA women's groups running there, and were not being studied by another organisation. These inclusion criteria helped to ensure that enrolled women came from a homogenous population group that faced similar cultural and social constraints and/or enablers to maternal and child nutrition.

Sample size

At the time of study design, using birthweight within 72 hours as the primary outcome, assuming a standard deviation (SD) of 410 g, an intra-cluster correlation coefficient (ICC) of 0.01, and two-tailed 5% significance level, we estimated that 163 birthweights per cluster from at least 17 clusters per arm would provide 80% power to demonstrate a difference of 50 g between any pair of study arms. After experiencing difficulties obtaining birthweights, but seeing little chance of cluster dropout, we repeated the calculation during trial implementation. Assuming 80 clusters contributing data, 111 birth weights per cluster, 2220 per arm and 8880 in total, provided the same power. For WAZ, the second primary outcome added during trial implementation, we estimated that 150 child measures per cluster would provide 84% power to demonstrate a difference of 0.12 between two arms, assuming an ICC of 0.01 and SD of 1.

LBWSAT interventions

Trial interventions in all three of the intervention arms comprised monthly Participatory Learning and Action (PLA) women's groups that pregnant women and their family members were invited to attend. The groups discussed Low Birth Weight (LBW) and its impact upon survival and morbidity through the life course and its root causes. Each of the series of 20 meetings addressed a different determinant of LBW and /or issues around pregnancy, intra-partum and postpartum care. The groups identified means of increasing

birth weight, barriers to doing so and potential strategies to overcoming the barriers. After prioritising problems that particularly affected their own community and strategies to address them, the groups implemented awareness-raising strategies for a period of approximately 12 months. Strategies included making home visits to women not attending groups, showing picture cards about LBW and pregnancy nutrition in the community, organising separate meetings with mothers-in-law, adolescent girls, or male family members to raise their awareness on the topic, holding rallies on the importance of maternal nutrition, and showing pregnancy-related videos.

Whilst the PLA only arm delivered women's groups only, the other two intervention arms combined PLA with a monthly social transfer to pregnant women. Up to 7 monthly transfers were available to ID card-holding women when they attended the groups. A small minority of women who were physically unable to attend the group could receive the transfer via a home visit. The transfer in the PLA + food arm was a fortified wheat-soya blended food called Super Cereal, 150g of which women were advised to consume daily as an extra snack. The remaining 180g/day provided to the women was allocated to allow for family sharing of the supplement. In the *PLA + cash* arm pregnant women received NPR 750 (c. USD $7.5), which they were advised to spend on nutritious food. This often was spent on a daily ration of milk and sometimes on fruits or but was not sufficient for regular purchase of fish or meat.

LBWSAT data collection

In order to identify pregnancies early and ensure that all pregnant women were involved, a population census (conducted August to Nov 2013) enumerated all women of reproductive age. The names of 64,000 consenting married women who had not had permanent family planning and whose husbands had not had vasectomy were entered into registers for use by ward enumerators. These incentivised community volunteers visited each woman once per month (between December 2013 and February 2015) to record in the register whether the woman had missed any menses or the status of her pregnancy if pregnant. After two missed menses the enumerator called an interviewer to come to conduct a pregnancy test, and to enrol women with verified pregnancies into the study. From this point onwards, during routine trial follow-up, all data were collected on mobile phones using the CommCare data collection platform.

At the time of enrolment, details from the menstrual register were first entered into the 'register woman' form (R). Then the pregnancy was confirmed with a urine pregnancy test and estimates of the Last Menstrual Period (LMP) and Estimated Date of Delivery (EDD) entered into the 'check suspected pregnancy' form (P). After enrolment the design was to capture details of the woman's early and late pregnancy, delivery and postnatal period by conducting follow-up interviews with the 'early pregnancy' (E) form from 8-31 weeks' gestation, the 'late pregnancy' form (L) from 31 weeks' gestation to birth, the

'delivery form' (D) within 72 hours if possible but up to 42 days where necessary and the 'post-neonatal' form (F) from 43 to 84 days after delivery.

Due to challenges with data collection and premature cessation of trial follow-up (described below), capture rates in follow-up questionnaires were low, so additional endpoint outcomes were added in Feb 2015 during trial implementation. These were captured using a cross-sectional survey between 20 June and 17 October 2015 using three forms for recording details of mothers (M), offspring (B), and distribution of trial participation incentives (I), implemented using the ODK data collection platform. Children were aged between 0 and 22 months at the time of the endpoint survey, which was conducted at a central data measurement station in each ward, instead of at home as the follow-up questionnaires had been. A 'clinic GPS' form was used to record the precise location of the endpoint survey measuring stations.

In addition to the enrolment, main follow-up and endpoint questionnaires described above, a 'woman update' form (U) was used to capture pregnancy outcomes and vital status of women from the menstrual monitoring registers as the trial progressed. A 'case close' form (Z) was used to capture pregnancy outcomes for those for whom pregnancy outcome had been missed, and a 'hospital birth weight' form (H) was used to collect birth weights in the Janakpur zonal hospital.

Pregnancy history and socioeconomic status were recorded at enrolment for most women using a 'socioeconomic' form (S), but for those who had not received this module by the time of the endpoint survey, key missing variables were collected in the 'baby endpoint' form (B) during cross-sectional data collection at endpoint.

Numbers and percentage capture rates by study arm for all the electronic data forms described above and for the combined socioeconomic and pregnancy history information from forms S and B are provided in Table 1. Capture rates for the main tools varied widely by data collection tool, with highest capture being for enrolment (P) and socioeconomic (S / B), lowest for late pregnancy (L), followed by post-neonatal (F) and early pregnancy. For the purposes of this paper we will examine capture of the main follow-up and endpoint tools (E, L, D, F, M and B) only as these provide most of the information to be used in secondary analyses apart from enrolment and socioeconomic forms which have very high coverage. We have not analysed capture of the woman update from registers (U), case closure (Z), incentive distribution (I) or hospital birth weight (H) forms as the findings from these mostly only fed into tracking of vital status and pregnancy outcome, not of any other variables.

Table 1. Data collection tools and their capture rates as a proportion of 25090 women enrolled

Data collection tool		**0. *Control***	**1. *PLA Women's group only***	**2. *PLA + Cash***	**3. *PLA + Food***	**Total**
Register woman (R)	n	5,299	5,615	7,245	6,877	25,036
	%	*99.8*	*99.8*	*99.6*	*99.9*	*99.8*
Check suspected pregnancy (P)	n	5,300	5,617	7,257	6,880	25,054
	%	*99.8*	*99.8*	*99.8*	*99.9*	*99.9*
Socioeconomic (S)	n	4,767	4,808	6,052	5,476	21,103
	%	*89.8*	*85.5*	*83.2*	*79.6*	*84.1*
Early Pregnancy (E)	n	673	609	1,483	1,189	3,954
	%	*12.7*	*10.8*	*20.4*	*17.3*	*15.8*
Late Pregnancy (L)	n	646	549	927	781	2,903
	%	*12.2*	*9.8*	*12.8*	*11.4*	*11.6*
Delivery (D)	n	2,036	1,827	1,965	2,104	7,932
	%	*38.4*	*32.5*	*27.0*	*30.6*	*31.6*
Post-neonatal (F)	n	984	760	677	613	3,034
	%	*18.5*	*13.5*	*9.3*	*8.9*	*12.1*
Mother endpoint (M)	n	4,134	4,407	5,383	4,977	18,901
	%	*77.9*	*78.3*	*74.0*	*72.3*	*75.3*
Baby endpoint (B)	n	4,161	4,461	5,436	5,047	19,105
	%	*78.4*	*79.3*	*74.8*	*73.3*	*76.2*
Incentive distribution endpoint (I)	n	4,098	4,396	5,321	4,953	18,768
	%	*77.2*	*78.1*	*73.2*	*72.0*	*74.8*
Case close (Z)	n	1,049	1,211	59	62	2,381
	%	*19.8*	*21.5*	*0.8*	*0.9*	*9.5*
Update woman (U)	n	3,247	3,003	3,721	3,858	13,829
	%	*61.2*	*53.4*	*51.2*	*56.0*	*55.1*
Hospital birth weight ((H)	n	0	0	5	1	6
	%	*0.0*	*0.0*	*0.1*	*0.0*	*0.0*
Socioeconomic data exist from either S or B forms	n	5,199	5,452	6,993	6,519	24,163
	%	97.9	96.9	96.2	94.7	96.3
Total	**n**	**5,309**	**5,626**	**7,271**	**6,884**	**25,090**

[1]note that some data is missing for 36 women who were enrolled using the check suspected pregnancy form which is why the total is 25054 for this tool, however we take the total 25090 pregnancies enrolled as the denominator for assessing capture rates.

Outcomes of pregnancies and numbers of maternal deaths are reported in Table 2. Whilst outcomes of pregnancies were ascertained for 21,549 (86%) and vital status of mothers for 20,503 (82%) of the women enrolled, it is important to note that we do not plan to analyse vital status of the pregnant mothers and the

index children of this study to estimate mortality rates. This is because we were not able to conduct our endline data collection as a house to house census, whereby we could have ascertained the pregnancy outcome and/or vital status of a higher proportion of the mothers and their offspring in a less biased manner. Whilst we were able to capture vital outcomes on more than three quarters of the mothers and offspring at the endpoint survey, and for additional women via other tools such as the woman update, case close and delivery/post-neonatal questionnaires, those not captured may differ systematically from those captured, especially in relation to their vital status and relative poverty. Mothers who died may have been missed. Mothers who had offspring who died would have been less likely to report for measurement at the endpoint data collection stations, whilst poorer women may have been more likely to report for measurement because this was a means of accessing the trial participation incentive they were eligible for.

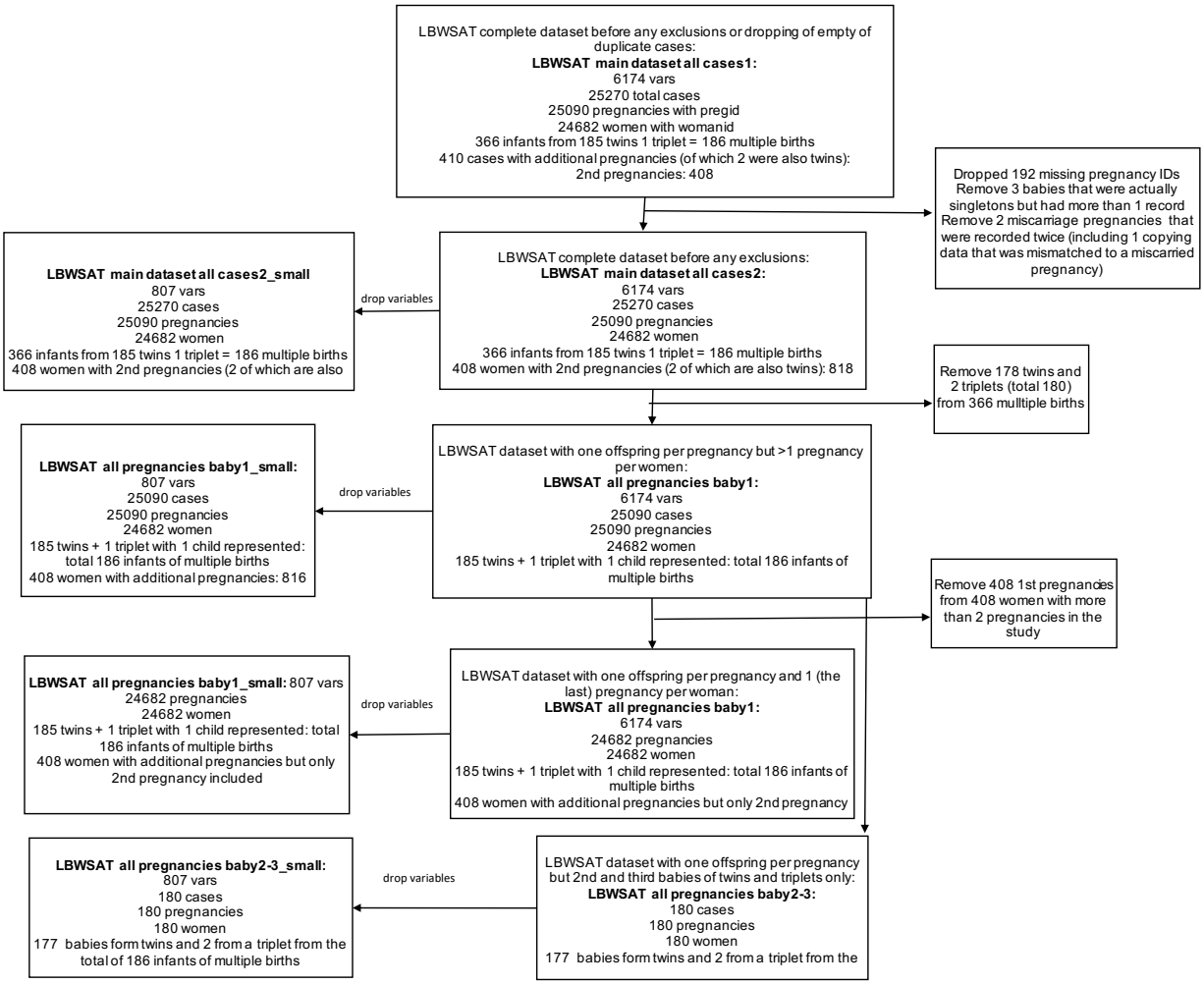
Table 2. Pregnancy outcomes and maternal deaths

Pregnancy outcomes	**Freq.**	***Percent***
0. Livebirth singletons and 1st baby of multiple births	19,665	*91.26*
1. Miscarriage / abortion	828	*3.84*
2. Twins / triplets	186	*0.86*
3. Stillbirth	276	*1.28*
4. Neonatal death	446	*2.07*
5. Post neonatal death	148	*0.69*
Total for whom pregnancy outcome available	21,549	*85.9%*
Maternal deaths identified		
Death within 42 days of delivery	15	*0.07*
Death 43-365 days after delivery	9	*0.04*
Mother's death more than 1 year after delivery	1	*0.00*
Total for whom vital status available	20,503	*81.7%*

A flow diagram of generation of LBWSAT datasets available and the exclusions applied is provided in Figure 1. In brief, whilst we began with 25,270 cases in the data, removal of cases with missing pregnancy IDs and duplicate offspring or pregnancies resulted in a dataset with 25,090 pregnancies from 24,682 women. For multiple births we had baby records for more than one child in some cases but not others, so for the purposes of analysing data for one offspring per pregnancy, we retained one child per pregnancy (where available) and removed 180 twins or triplets from 366 multiple births. In this dataset, which we use for offspring outcomes, we retained 408 second pregnancies in the study. For the purposes of analysing outcomes relating to individual women we dropped 408 first pregnancies from the dataset (leaving 408 second pregnancies for these woman) to yield a dataset with 24,682 women, each with one pregnancy. This dataset with one offspring and one pregnancy per women is useful for looking at outcomes where we

want each woman-offspring dyad to be an independent observation (such as those analysing maternal capital outcomes such as height, body mass index (BMI) or education).

Figure 1. Flow chart of LBWSAT data during processing



Challenges with LBWSAT data collection

There were considerable challenges to be overcome during the implementation of LBWSAT, which had an impact of data capture rates using the different data collection instruments. These issues are summarised below.

Data collector overload

A higher number of women enrolled than we had expected, which created a burden upon data collectors to follow-up women at multiple time points. In addition to the enrolment questionnaire, five subsequent data collection forms were required for each mother, which made it difficult to complete follow-up on all woman. This resulted in poor coverage of most of the follow-up questionnaires.

Pressure to enrol women in food and cash transfer arms

Because there was an economic incentive to enrol early for women in the *PLA + food* and *PLA + cash* transfer arms, data collectors came under pressure from community members to enrol women in these arms, whilst in the *control* and *PLA women's group-only* arms, participants were less eager to be enrolled. This led to data collectors focusing their attention on enrolling women at the cost of completing follow-up on women already enrolled. It also meant that women in the *PLA + food* and *PLA + cash* arms enrolled earlier in pregnancy than those in the *control* and *PLA women's group only* arms.

Time lag from enrolment to receipt of participant ID cards

Although data collection upon electronic data collection forms using the CommCare data collection platform made the data available immediately after each data collection form was synched with the server, our system relied upon the printing of photo ID cards for each pregnant mother participating in the trial, using a printer located in the MIRA Dhanusha Janakpur office. This office was up to 3 hours' drive away from the remotest clusters in the furthest areas of Dhanusha and Mahottari districts. Although a robust electronic system of generating the cards was in place, and cards were regularly delivered to field sub-offices, there was inevitably a delay between the registration of the pregnancy during the enrolment questionnaire and the issuing of the card to the mother for subsequent follow-up and/or receipt of food and cash transfers. This may have led to early or late pregnancy questionnaires being missed.

Women delivering before data collection began

At the start of data collection, pregnant women had been waiting for some months between being told about the study and undergoing full enrolment. The delay in the start of the study was partly due to the general election which took place in November 2013. This made the holding of community meetings impossible (by government mandate), and disrupted work as project employees had to return to their natal home for voting. Another reason for the delay in starting enrolment was that we underestimated the time required to programme and finalise the electronic data collection system and individual questionnaires on CommCare. This delay, between community meetings held between July and September and the start of data collection at the end of December 2013, meant that there were women who were pregnant in the final months of 2013 who had expected to receive benefits from the trial and who were disappointed not to have had the chance to enrol. To compensate these women, we allowed 1841 postpartum women to enrol in the trial and receive a food or cash transfer and/or the end of data collection payment even though they had already delivered and therefore could not be included in trial analyses or be interviewed during pregnancy. We felt that these women provided data collectors with an opportunity to practice their interviewing skills and anthropometry, even though they could not provide outcomes in the trial.

Permitting women to enrol at any gestational age

We were aware that, due to the cash and food transfers, pregnant women may have perceived it beneficial to report that they were earlier in their pregnancy than they actually were, in the hope of receiving more transfers during their pregnancy. In actuality, reporting a falsely early LMP did not result in women receiving more transfers, because they were no longer eligible for transfers after delivery. To avoid further

incentivising false reporting of the LMP, and so as not to exclude women from interventions who may have needed them, we allowed all pregnant women to enrol in the trial and receive benefits regardless of their gestational age. This meant that some women enrolled too late in pregnancy to make best use of the interventions, to increase their food intake or to change their behaviour in the way intended by the trial.

In-migration to access transfers

In the context of the plains of Nepal, women routinely move between their marital home and their natal home during pregnancy, often moving to their parents around the time of the delivery. In *PLA + food* and *PLA + cash* arms, some women migrated into the study area in order to access transfers, staying with relatives or their parents rather than in their marital homes. This created some imbalance between arms, as poorer women were more likely to migrate in, so a large number of in-migrating women were excluded from trial analyses. Of the 25,090 pregnancies enrolled, 22% of the sample (n=5,449) were classified in-migrators. We considered women not to be in-migrators if they were either enrolled at the time of the census (August to November 2013) or were newly wedded and living in their marital home with the husband's family at enrolment.

Political instability and conflict in the field team

The study occurred during a period of political instability, when tension between previously ruling hills ethnicities and marginalised plains ethnicities was high. This fuelled conflict within the trial implementation team on the ground, particularly concerning the recruitment of new team members. This in turn led to many of the 84 data collectors and their supervisors not performing data collection as required, especially in the months of April 2014 and from late October to December 2014. It also meant that recruitment of new team members was not possible, so it became difficult to alleviate data collectors' work overload in terms of follow-up.

Logistical issues

There were a number of geographical barriers to visiting women in their homes in this part of Nepal. Many of the roads were either dirt tracks, or dirt tracks with some gravel, while others have been tarmacked in the past but have not received maintenance, such that they had many potholes and were in bad condition. These poor road conditions and the far spread of study clusters from the urban centre of Janakpur made travel to the study sites challenging. In the monsoon, flooding also blocks off road access to communities, and dirt roads become very muddy and slippery. Walking around communities with anthropometric equipment is made difficult by the monsoon rain, floods and muddy conditions, and is also very taxing in the summer (hot season) months when temperatures often exceed 38 degrees Celsius. In the winter, cold temperatures also make work difficult, because participants may not want to strip their babies, and pregnant women may also not want to remove sweaters and shawls for anthropometric measurements. Moreover, in the winter season very dense fogs descend on the Nepal plains such that people do not see the sun for several days at a time. High humidity and low temperatures during these periods lead to illnesses, and driving conditions are very difficult due to extremely poor visibility.

Decision to establish a run-in period

Before starting data collection, in collaboration with the trial steering committee and data monitoring committee, we decided to have a run-in period to allow interventions and surveillance systems to become established before counting any participants' outcomes in the impact evaluation analyses. We included participants in the run-in if they were due or delivered between 13 Feb 2014 to 4 June 2014 inclusive (Table 3). Prior to the run-in of the interventions we also had a period of enrolling and following up women to enable the surveillance system to get up and running. This included women who were due or delivered between 12 Feb 2014 to 4 June 2014 inclusive.

Hiatus in data collection during training

At the end of the run-in period we ran a training session from late May 2014 to early June to refresh existing data collectors, and to train up additional data collectors to increase follow-up rates. At this time, we deleted some non-essential questions and made minor edits to existing questions where necessary. Whilst in principle this approach offered benefits, it resulted in a hiatus in data collection from 22 May to 7 June 2014 inclusive (with just one delivery questionnaire collected on 31 May 2014). This further reduced follow-up rates of enrolled women, especially for the timebound delivery and early/late pregnancy questionnaires.

Funding

Although the original trial was planned to end in February 2015, in practice more time was needed to capture enough birth weights to reach our target sample size. Unfortunately, funds to continue data collection using the existing data collection system of 5 follow-ups per mother were not forthcoming at this time, and the decision was taken to end enrolment of new women on 28 Feb 2015 and to end birth weight surveillance on 31 March 2015, despite a large number of women needing to be enrolled and followed-up to achieve the intended sample size. This left 5,457 women who had been enrolled, but whose birth weights could never be captured due to the cessation of surveillance activities. In order to obtain an outcome from as many trial participants as possible, we added a cross sectional measure of weight-for-age z score as a second primary outcome of the trial. We devised a cross-sectional endpoint survey and, between 20 June and 19 October 2015, sampled as many as possible of the enrolled women and their index children born of the trial with anthropometry and questionnaires. This provided a measure for most of the participants but could not provide the timebound measures at birth, early and late pregnancy and post neonatal timepoints, which meant that these follow-ups had very low capture rates.

Table 3 provides a summary of the cases available for analysis from LBWSAT on the basis of the period in which the offspring was born. Interventions began on 13 February 2014, and 9% (n=2,317 enrolled mothers) delivered before this. A further 14% of pregnant mothers (n=3,501) gave birth during the period when interventions we running-in (13 Feb to 4 June). These women had some exposure to interventions if they resided in intervention arms but it had been agreed with the data monitoring committee (DMC) and

trial steering committee (TSC) that the interventions, especially the participatory learning and action women's groups (PLA), should be allowed time to get established before birth weights could be included in analyses. The majority (55%) of mothers (n=13,815) gave birth between 5 June 2014 and 31 March 2015 during the 'full trial' phase of the study. These had the opportunity to receive the interventions and to have their birth weights measured, and so could be included in the primary analyses. A further 8% of women (n=2,079) gave birth after the end of the birth weight surveillance, when the interventions were continuing but the main surveillance tools (forms R, P, S, E, L, D, F) were no longer being used. These women were permitted to be included in endpoint outcome analyses but had missed the opportunity for birth weight measurement. Another 13.5% (n=3,378) delivered after the endpoint measurement had started. These were included in the endpoint study but not included in any trial analyses. This illustrates the vast quantity of data collected which could not be used for cluster RCT purposes, and also provides a rationale for using the dataset for secondary analyses going forwards.

Table 3. Periods of classifying cases on the basis of date of birth in the Low Birth Weight South Asia Trial

Period in which infant born or due	**Freq.**	**%**
0. Born/due before 13Feb14 when integrated interventions started	2,317	*9.2*
1. Born/due 13Feb-04Jun14 during run-in of interventions	3,501	*14.0*
2. Born/due during full trial period between 05Jun14 & 31Mar15 when birth weights stopped	13,815	*55.1*
3. Born/due between 01Apr15 & 20Jun15 when nutrition clinics started	2,079	*8.3*
4. Born/due between 20Jun15 & 15Oct15 nutrition clinics started	3,378	*13.5*
Total	**25,090**	*100.0*

Statistical analyses to assess bias in the women sampled in different tools

We undertook analyses to estimate the extent that differential data collection may have resulted in bias. First, we descriptively compared background characteristics by study arm and by in-migration status. Second, we analysed the differences between captured and non-captured women for each of the main questionnaires in terms of age of marriage, age at enrolment, wealth score, maternal height (using t-tests) and by maternal and husband's education using Chi-squared tests. Thirdly we fitted mixed logistic regression models, with study cluster as a random intercept, to estimate the odds of data capture for each of the tools. For each questionnaire, capture was the binary dependent variable and individual sociodemographic characteristics (named above) were the independent variables. This enabled us to see which sociodemographic characteristics were significantly associated with data capture and which were not. Finally, we mutually adjusted to for the sociodemographic variables in one model, so as to provide hints as to what factors should be adjusted for in analyses.

Results

Factors affecting capture of each of the main data collection tools

The combined factors (described above) of data collector overload, pressure in the *PLA + food* and *PLA + cash* arms to preferentially enrol more women earlier in their pregnancies, delays in ID cards reaching women, women enrolling late in gestation or even after the end of pregnancy, in-migration of women to access food or cash transfers, political conflict, logistical problems, hiatus in data collection during training, and cessation of funding, explain the poor capture rates of birth weight and follow-up questionnaires. Further details of these issues specific to each of the main data collection tools are also described in Tables 4 and 5 below.

Table 4 describes the process of being enrolled into and followed through the study by ward enumerators, while Table 5 describes the data collection tools implemented in CommCare and factors affecting their capture. This may be referred to by any researcher seeking to understand specific factors that may have affected data capture of each of the tools. In summary, factors which may have affected who was captured included: the motivation and skill of the ward enumerators and interviewers; enrolment of *PLA + food* and *PLA + cash* arm mothers earlier in their pregnancies than those in *control* or *PLA women's group-only* arms, which meant that gestational age at sampling differed by arm; women were enrolled in the *PLA + food* and cash transfer areas; women who migrated to their parental home around the time of delivery were poorer or more likely to be Muslim than those who stayed with the husband's family; mothers and their offspring who had negative birth outcomes may have got missed from endpoint, while poorer women and women in *control* and women's group areas reported more to measuring stations due to participation cash incentive distribution.

Table 4. Process of being enrolled into and followed through the study by ward enumerators

Information source	Type of information collected by ward enumerator	Factors affecting differential data capture
Census of married women of reproductive age	All consenting married women aged 10 to 49 years were recorded in ward-level registers (4 sectors per each ward) Aug to Nov 2013 Register had woman's name, address, age, caste grouping. A unique woman ID identifier was pre-printed on each page and allocated to the woman	Some census data collectors may have been more thorough than others.
Monthly	Ward enumerators (WE) visited each sector once per	Some ward enumerators may

checking of menstrual or pregnancy status	month (over 1 week) to record menstrual status. After two missed menses were recorded in the register, the WE sent an automated SMS message to alert the associated interviewer to come to enrol the woman.;	have been more diligent or able than others and some received better support from interviewers than others.
WE monitoring of pregnancies	Each pregnant mother was monitored by the WE for any change in status and this was recorded in the register. Each miscarriage, abortion, permanent migration, maternal death or live / still birth and the date it occurred was recorded. When a birth was detected, WE sent an automated SMS message to alert the interviewer to come to weigh the baby and fill the delivery questionnaire *immediately.*	Some ward enumerators were more diligent or able than others and some received better support from interviewers than others.

513

514 Table 5. Data collection tools implemented in CommCare and factors affecting their capture

Information source	Type of information collected during follow-up (i.e. tracking of pregnancies/deliveries/ children at the same stage for each woman)	Factors affecting differential data capture
R: register woman	Recorded data from the menstrual register (at time of pregnancy test): age, caste, woman ID	Some women who resided in a ward may not have been included in menstrual monitoring and so were missed (low risk).
P: check suspected pregnancy	Consent, Last menstrual period, allocates pregnancy ID to pregnant woman (PW), first pregnancy or not, recently married or not, address of sasural (marital) and maiti (parental) homes, location of enrolment, London measure of unwanted pregnancy (LMUP), GPS location of the household.	As above. As this is the enrolment questionnaire, we have this on all pregnancies (N= 25,090). However, in *PLA + food* and cash intervention arms, mothers enrolled on average 7.5 weeks earlier than mothers in women's group and *control* arms. Possibility of inaccurate reporting of last menstrual period (LMP) or gestational age in *PLA + food* and *PLA + cash* thinking they might be entitled to more transfers. Mean gestation in weeks at enrolment: *control* 26.7; women's group 26.5; cash 19.2; *PLA + food* 19.1.
S: socio-economic	Collected at time of P in most cases, but if missed was collected at endline in baby questionnaire or during the empowerment module (collected by Lu Gram and team for his	Most women (N= 21,103, 84.1%)(prefix S socioeconomic data) had SE and previous pregnancy data collected before delivery of the child (in early pregnancy) but those with

Information source	Type of information collected during follow-up (i.e. tracking of pregnancies/deliveries/ children at the same stage for each woman)	Factors affecting differential data capture
	PhD [10, 11]). Domains covered in this tool included caste/language/ethnicity, assets, toilet, water access, cooking fuel, education of mother and her husband, migratory labour or husband and household, age of woman, pregnancy history and status of previous children (including school enrolment), and breastfeeding status, amongst others.	prefix B (n= 2,889, 11.5%) or Empowerment module (N=1,050, 4.2%) had data collected after delivery so interventions or time might have changed their status and index child could be included in pregnancy history. Age of marriage and age of mother both subject to poor recall as people (especially the illiterate) do not know their age or DOB.
E: early pregnancy: 8-31 weeks' gestation	Domains included: Vital status of PW, plans for moving during pregnancy, food security, nausea/vomiting, eating behaviour (dietary diversity, fasting, food taboos, eating down), smoking/pan/alcohol consumption, decision making, agency/empowerment questions. GPS. Anthropometry: MUAC, Height, weight of PW, Other (youngest) child's weight and height/length.	Capture rate very low (N= 3,954, 15.8 %) Potential for bias in who got captured or not. Women's group and *control* arms sampled about 1- 2 weeks later than *PLA + food* and *PLA + cash* arms.
L: late pregnancy - 3rd. trimester, 31 weeks' gestation to birth	Domains largely repeated from early pregnancy with some additional measures of morbidity. Tool included: Vital status of PW, plans for moving during pregnancy, ultrasound information, consumption of supplements and iodised salt, food security, nausea/vomiting, eating behaviour (dietary diversity, fasting, food taboos, eating down), receipt of food or cash transfers, exposure to women's groups and their strategies and to home visits, consumption of super cereal (food supplement) decision making, agency/empowerment questions, pregnancy morbidity, hand washing, GPS. Anthropometry: MUAC, weight of PW, Height (if missed earlier or if PW adolescent).	Capture rate very low (N= 2,903, 11.6%). Potential for bias in who got captured, especially if some mothers started to move to their maiti (parental) home in late pregnancy. Women's group and *control* arms sampled <1 weeks later than *PLA + food* and *PLA + cash* arms so gestational age bias probably not so important.

Information source	Type of information collected during follow-up (i.e. tracking of pregnancies/deliveries/ children at the same stage for each woman)	Factors affecting differential data capture
D: delivery - within 72 hrs up to 42 days	This was the tool to capture the primary birth weight outcome. Domains included: vital status of mother and index offspring, multiple birth, sex, date of birth (DOB), birth location and institution/home, type of delivery, delivery assistant, immediate care after birth (cord care, bathing, breastfeeding, colostrum, pre-lacteal feeds), birth defects, delivery complications. GPS. Anthropometry of index child born of the study: birth weight, length and head circumference. Sibling (next youngest) child's weight and height/length- repeated on the child measured in E where possible.	Capture rate low (N=7,932, 31.6%). Bias in who got captured generated most probably by women moving to their parental home in late pregnancy for delivery and early postpartum. Compared with those whose infants were not measured, mothers of infants whose weights were captured were significantly older, had more children, were more likely to be Hindu than Muslim, and were less likely to have primary or secondary education.
F: post-neonatal 43 to 84 days	Domains included: vital status of mother and index offspring; details of antenatal and postnatal check-ups (tetanus, iron-folic acid, vitamin A, type of checks, FCHV visits); maternal morbidity in pregnancy and postpartum; neonatal morbidity, vaccination, care seeking, breastfeeding, food security, how health care costs were covered; GPS. Anthropometry of mother (weight, MUAC, height if missing or adolescent) and index baby (length, weight, head circumference).	Very low capture rate (N=3,034, 12.1%). Potential for women who went for longer periods to their maiti (parental) home to be differentially missed out (as with delivery questionnaire). As the time after delivery increases, the potential for recall bias of pregnancy and delivery / neonatal morbidity increases. Only 66% of cases were 43-84 days of delivery, remaining 1/3rd of cases ranged from 85 to 323 days old. Mean age 84 days (SD 41).
H: health facility delivery	Used in the Janakpur zonal hospital (and selected nursing homes) to collect vital status of mother and baby, multiple birth, sex, DOB, birth institution. Anthropometry of index child born of the study: birth weight, length and head circumference.	N=6 (0.02%). This was an attempt to capture birth weights of babies born in health institutions, but this was not effective as but was started too late and ran for only a short time. Only 6 cases could be merged into the data.
Z: Case close	This was used to close cases from the CommCare system and provide women with the participation incentive so that they	N= 2,381, 9.5%. Some cases may have been missed depending on how active the interviewer was, especially in *PLA + food* and

Information source	Type of information collected during follow-up (i.e. tracking of pregnancies/deliveries/ children at the same stage for each woman)	Factors affecting differential data capture
	could enrol with a second pregnancy if needed. Recorded the dob, multiple birth, vital status of the baby and mother, timing of deaths if mother or baby died, stillbirth questions (to differentiate still born and birth asphyxia).	*PLA + cash arm*s. Most cases collected in *control* and women's group only arms where participation incentives were payable after the end of the pregnancy only (as opposed to receiving food or cash transfers during interventions as occurred in the other arms).
U: woman update	This tool was to update the 'case' information in CommCare when an enrolled woman changed status. The data were not collected from interviews but entered into CommCare forms from records in enumerators registers. Information on the following were recorded: still pregnant, miscarriage, abortion, died in pregnancy without delivering a child, delivered and is alive, delivered and died in childbirth or later. If mother died – timing of death. If baby born still or live birth, dob, date of death (dod) if died.	N= 13,829 (55.1%). Some interviewers seemed to use this much more than others. Potential for maternal deaths or births/stillbirths and infant deaths to be missed if the enumerator was not maintaining the register or if the enumerator and their associated interviewer were not checking very carefully.
Information source	Type of information collected cross sectionally at endpoint	Factors affecting data capture N= 18,733 (74.5%) of all enrolled women captured with *both* maternal and baby modules; N= 19,275 (77.0%) had one or both
B: endpoint baby	Cross-sectional endpoint data collection on the (youngest if more than 1) index child born of the trial: checked status of mother and baby and dob, child gender etc as per records from surveillance questionnaires. Also collected data on: where the baby was born, type of delivery (c-section, normal), breastfeeding at time of initiation and in last day, infant feeding (fed solids or not, dietary	N=19,105 (76.2%). Those who had a stillbirth, neonatal or infant death or maternal death were less likely to have reported for interview at the central ward-level interview / anthropometry measuring station, so mortality data not useful for calculating child mortality or still birth *rates.* Problem with jump sequence in questionnaire meant that breastfeeding

Information source	Type of information collected during follow-up (i.e. tracking of pregnancies/deliveries/ children at the same stage for each woman)	Factors affecting differential data capture
	diversity) child morbidity in last 2 weeks, receipt of food and cash transfers, exposure to PLA women's groups and women's group strategies. SES and previous pregnancy questions for those who had missing surveillance SES module: including education of mother and husband, assets, cooking fuel, toilet, landholding, water. Anthropometry of index child: weight, length, head circumference.	details were only collected on a subset of exclusively breastfed children. SES and past pregnancy questions collected after delivery so do not come from early pregnancy like the rest of the data. Age of child sampled: 0-90 days (0-3m) n= 2,223 (12.2%); 91-180 days (4-6m) n= 2,287 (12.5%); 181-270 days (7-9m) n= 3,240 (17.7%); 271-365 days (10-12m) n= 3,855 (21.1%); 12-18 months n= 4,342 (23.7%); 18-24 months n= 2,340 (12.8%)
M: endpoint mother	Cross-sectional endpoint data collected on the mother after delivery including the following: vital status of mother, currently pregnant or not, checked names and which VDC she was enrolled at. Anthropometry of mother: weight, MUAC, height if missing previously or if adolescent.	N= 18,901 (75.3%). Those who had stillbirth, neonatal or infant death or maternal death less likely to have reported for interview at central interview / anthropometry location so mortality data not useful for calculating maternal mortality *rates.* There were one or two VDCs sampled at the beginning of the endpoint study where some maternal heights were missed due to a wrong jump sequence but every VDC does have height data (n= 12476 measured at endpoint and n=6280 measured earlier in pregnancy/ postpartum): Total maternal height n= 16,295 (65%).
I: endpoint incentive	Cross-sectional endpoint data collection on delivery of incentives for participating in the trial	N=18,768. Only really useful for checking for receipt of participation incentives which were payments of NPR1000 and provision of a towel and two soaps. The distribution of NPR 1000 to women's group and *control* arm women who had not received it before meant that high capture rates were found in these arms at endpoint.

Information source	Type of information collected during follow-up (i.e. tracking of pregnancies/deliveries/ children at the same stage for each woman)	Factors affecting differential data capture
Clinic GPS	GPS waypoint of endpoint (anthropometry station) location. Identity of staff. Serial numbers of anthropometric equipment used.	Useful for mapping the location of the cluster to which mother was allocated but not for her home. Useful to check if any scales or length boards had problems. However, we assume not as calibration of equipment was tested.

Variables available for analyses

Types of maternal and child outcomes and descriptors of the women, children and their households and the number of cases available are provided in Table 6.

Table 6. Information available upon women, children and their households in the LBWSAT dataset, the source of the information and the number of cases available

	Cases available (n)	Enrolment[1]	Early pregnancy[1]	Late pregnancy[1]	Delivery[1]	After 42 days[1]	Endpoint[2]	% of enrolled
Household-level descriptors								
Wealth score or quintile	23861	1					1	95.1%
Multi-dimensional Poverty Index (MPI)	22685	1					1	90.4%
Characteristics of housing - roof, walls and floor	24159	1					1	96.3%
Ownership of key assets (including radio/ TV, mobile, computer, means of transport, agricultural tools)	24159	1					1	96.3%
Drinking water source	24159	1					1	96.3%
Toilet	24159	1					1	96.3%
Access to electricity	24159	1					1	96.3%
Fuel used for cooking	23966	1					1	95.5%
Use of non-biomass fuel	24075	1					1	96.0%
Anyone in the household with at least 5 years of education	22864	1						91.1%
Husband's education level	24161	1						96.3%
Travel distance to nearest local shop, local market (haat bazaar) and big market	21098	1						84.1%
Landownership	24159	1						96.3%
Landholding size	22143	1						88.3%
Main source of staple food	21098	1						84.1%
Food security status (household food insecurity access scale HFIAS)	2617		1	1				10.4%
Months of adequate household provisioning	3954		1					15.8%
Household dietary diversity score	3954		1					15.8%
Migratory labour of husband	21420	1						85.4%
Migratory labour of other household members	21420	1						85.4%
Descriptors of mother								
Age at enrolment	25054	1						99.9%
Gravida	24274	1					1	96.7%
Parity	24094	1					1	96.0%
Caste	25054	1					1	99.9%

	Cases available (n)	Enrolment[1]	Early pregnancy[1]	Late pregnancy[1]	Delivery[1]	After 42 days[1]	Endpoint[2]	% of enrolled
Religion	21423	1						85.4%
Ethnicity (hills/plains)	25054	1						99.9%
Language	25054	1						99.9%
Maternal outcomes								
Anthropometric measures at endpoint								
Weight	14064		1	1		1	1	56.1%
Height	16295		1				1	64.9%
BMI	14011						1	55.8%
Mid-Upper Arm Circumference (MUAC)	14095						1	56.2%
Low BMI <18.5 kg/m^2	14011						1	55.8%
Low stature <145cm	16295						1	64.9%
Thinness by MUAC	14095						1	56.2%
Overweight / obese BMI >25 kg/m^2	14011						1	55.8%
Education								
Ability to read	24012							95.7%
Combined literacy	24162							96.3%
Years of education	24162							96.3%
Previous pregnancies / deliveries								
Gravida	24144	1						96.2%
Parity	23964	1						95.5%
Indicators of agency or empowerment								
Age at marriage	21422	1						85.4%
Age at first pregnancy (calculated from DOB of first child and age at enrolment)	23614	1						94.1%
Who decides upon food purchase	3954		1					15.8%
Who decides about food (cooking, who eats what)	3954		1	1				15.8%
Who decides upon use of cash generally	2617			1				10.4%
Who decides about use of cash transfer to pregnant women	2617			1				10.4%
Who does the shopping	2617		1	1				10.4%
Ladder' question of woman's perceived agency status (now, in 5 years' time and in comparison with her neighbours)	3954		1	1				15.8%
Amount of rest during pregnancy	2617			1				10.4%
Extent of help from family members during pregnancy	2890			1				11.5%
Order of eating (number of family members that eat before and after the pregnant woman)	2889		1	1				11.5%
Newly married in last 9 months	4882	1						19.5%
London Measure of Unwanted Pregnancy (LMUP)	15631	1						62.3%
Maternal care uptake								
Antenatal care	3292					1		13.1%
Institutional delivery	19668						1	78.4%
Skilled health worker at delivery	6454				1			25.7%
Postnatal care	1152					1		4.6%
Maternal morbidity in pregnancy / delivery / postpartum								
Late pregnancy anaemia signs	2890			1				11.5%
Morbidity in labour and delivery (bleeding, retained placenta, pre-eclampsia signs)	3034					1		12.1%
Sought care for maternal health once ill	1926					1		7.7%
Birth anthropometry (0-42 days) of index child born of the trial								
Anthropometry								
Weight	6583				1	1	1	26.2%
Length	6515				1	1	1	26.0%
Head circumference	6539				1	1	1	26.1%
Height/length for age z-score (HAZ)	6478				1	1	1	25.8%
Weight for age z-score (WAZ)	6565				1	1	1	26.2%
Weight for length/height z-score (WHZ)	6044				1	1	1	24.1%
Head circumference for age z-score (HCAZ)								
Stunting	6478				1	1	1	25.8%

	Cases available (n)	Enrolment[1]	Early pregnancy[1]	Late pregnancy[1]	Delivery[1]	After 42 days[1]	Endpoint[2]	% of enrolled
Underweight	6565				1	1	1	26.2%
Wasting	6044				1	1	1	24.1%
Child outcomes of index child born of the trial								
Anthropometry								
Weight	16667				1	1	1	66.4%
Length or height	17255				1	1	1	68.8%
Head circumference	17203				1	1	1	68.6%
Height/length for age z-score (HAZ)	17107				1	1	1	68.2%
Weight for age z-score (WAZ)	16609				1	1	1	66.2%
Weight for length/height z-score (WHZ)	16562				1	1	1	66.0%
Head circumference for age z-score (HCAZ)								
Stunting	17106				1	1	1	68.2%
Underweight	16660				1	1	1	66.4%
Overweight	16660					1	1	66.4%
Wasting	16608				1	1	1	66.2%
Neonatal / Child Morbidity								
7 neonatal danger signs in 7YICSS	2982					1		11.9%
Other individual neonatal danger signs	2982					1		11.9%
any illness in the last 2 weeks at endline	13057						1	52.0%
Sought care for neonatal health	2630					1		10.5%
Mother's Mortality information available								
Maternal or pregnancy-related death within 42 days of delivery	20503					1		81.7%
Late maternal or pregnancy-related death 43-365 days of delivery	20502					1		81.7%
Index child mortality information available								
Birth outcome	21549							
Neonatal death outcome	20359						1	81.1%
Post-neonatal death outcome	20362					1		81.2%
Child Feeding								
Exclusive breastfeeding	12931						1	51.5%
Colostrum discarding	14677						1	58.5%
Time of initiation of breast feeding	5319						1	21.2%
Prolonged breastfeeding	9695						1	38.6%
Child dietary diversity and key food groups	12627						1	50.3%
Child outcome on sibling (during pregnancy of index child)								
Anthropometry								
Weight	532		1		1			2.1%
Length or height	401		1		1			1.6%
Height/length for age z-score (HAZ)	374		1		1			1.5%
Weight for age z-score (WAZ)	506		1		1			2.0%
Weight for length/height z-score (WHZ)	373		1		1			1.5%
Stunting	374		1		1			1.5%
Underweight	506		1		1			2.0%
Overweight	506		1		1			2.0%
Wasting	373		1		1			1.5%
Indicators from previous pregnancy / children histories								
No of previous pregnancies	24144	1					1	96.2%
No of previous livebirths	23636	1					1	94.2%
No of previous stillbirths	23636	1					1	94.2%
No of previous miscarriages	23976	1					1	95.6%
No of previous abortions	23646	1					1	94.2%
No of children still alive or dead	23523	1					1	93.8%
No of previous girl children	23523	1					1	93.8%
No of previous boy children	23523	1					1	93.8%
Education of children from previous pregnancy / children histories								
Enrolled in school - for at least 1 child	5601	1					1	22.3%
Enrolled in school - summed across all siblings	6309	1					1	25.1%

	Cases available (n)	Enrolment[1]	Early pregnancy[1]	Late pregnancy[1]	Delivery[1]	After 42 days[1]	Endpoint[2]	% of enrolled
Education of children (class at school) - of at least 1 child	3932	1					1	15.7%
Education of children (class at school) - summed across all siblings	24117	1					1	96.1%
School absenteeism - of at least 1 child	5601	1					1	22.3%
School absenteeism - summed across all siblings	9877	1					1	39.4%
Mortality of previous children								
Child death types - for at least 1 child	612	1					1	2.4%
Child death types -summed across all siblings incl. previous twins	1203	1					1	4.8%
Vital status of previous children - at least 1 child	14,396	1						57.4%
Vital status of previous children - summed across siblings	29,648	1						118.2%

[1]main surveillance questionnaires delivered to participants between 8 weeks gestation and 3 months postpartum; [2] cross-sectional endpoint data collection.

The subject domains available in the data fall in the following categories: i) **household descriptors** such as assets, wealth scores, multi-dimensional poverty index, housing characteristics (materials the home is made from, toilet, electricity, cooking fuel), landholdings, market access, migratory labour; ii) **maternal characteristics** – age, gravida/parity, caste/ethnicity, religion; iii) **maternal capital outcomes** – anthropometry, education; iv) various indicators of **agency and empowerment** in the mother including decision-making, perception of own agency (ladder question), whether newly-wed or not, access to family planning (London Measure of Unwanted Pregnancy), age at marriage and at first pregnancy, eating order in the household, education of mother relative to husband; vi) **maternal care seeking and morbidity**; vii) **index child outcomes** including anthropometry and morbidity between birth and 22 months (3 time points); viii) **child feeding** practices – breast feeding and dietary diversity of the index child; ix) **anthropometry of sibling** of index child; x) **histories of previous pregnancies and births** including miscarriages, live/still-births, child deaths; xi) **access to education** and class in which each of the surviving children is studying or has studied; xii) **vital status** of index child and mother and of previous children. Specific variables (or groups of variables) can be found directly in Table 6 (above).

Characteristics of the study population

Characteristics of the women enrolled in the Low Birth Weight South Asia Trial and their households summarised below and a selection (subset) of descriptors of women by study arm of the trial are provided in Table 7.

Nearly all participants were married women of Maithili-speaking plains 'Madhesi' ethnicity. Most (84%) were Hindu while the remainder were Muslims. The most common caste group (35%) was Middle Madhesi including Sah, Sudi, Teli, Mandal and others while 20% were Yadavs and 19% were Dalits (traditionally treated as untouchable) with few (7%) Janajati or Brahmin/Chettris (3%). Over one third (36%) of women were primigravida, 26% had 1 child already and 38% had 2 or more children. 41% were adolescent and most of the remainder were between 20 and 29 years (mean age 21.5 years SD 4.6). More than half of the

women had been married between 13 and 15 years of age and one third between 16 and 18 years. Only 6% of participants were married at 19 years or above, while 5% were married at 12 years or below (mean age at marriage 15.3 years, SD 2.1). Nearly two thirds (64%) of women had never been to school, 20% had primary education and only 16% had secondary education or above. In contrast, 48% of women's husbands lacked schooling, 27% had primary and 25% had secondary education. More than a third (36%) of households had no family member with more than 5 years education.

Over one third (34%) of households were landless, 30% had marginal landholdings of less than 0.5 hectares, 25% between 0.5 and 2 hectares and 6% two hectares or more. Average landholding size was small at 0.48 hectares (SD 0.93). Two thirds of households produce their own food while 31% ranked labour exchange (being paid in food) and 24% ranked share cropping as a main source of food. Four fifths of households relied on purchase of food to meet their needs and 36% said this was their primary source. Using the household food insecurity access scale 79% of households were food secure, 14% had mild, 6% had moderate and 1% had severe food insecurity.

Three-quarters of households lacked a toilet and nearly all burned biomass fuels for cooking. Houses were mainly small: 16% had only 1 room, 32% 2, 25% 3 and 28% 4 or more rooms. Most households had a television (69%), bicycle (76%) and mobile phone (88%), but only 20% owned motorcycles, 3% a computer, 2% a tractor and less than 1% any other kind of vehicle. A small proportion (8%) had no access to electricity for lighting. All had easy access to water from a hand pump or bore hole either inside their home or interior courtyard (49%), in their yard (25%) or in a nearby public location (26%) less than 5 minutes away. Most participants live in traditional houses comprising a local timber frame and walls made from plant stems plastered with mud (66%), with roofs made of traditional tiles (44%) or modern tiles/metal (26%) or thatch (10%), and mud floors (86%). Only 29% had the more modern cement and brick walls, 20% had cement roofing and 14% cement plastered flooring.

In terms of proxy measures of agency and empowerment, 9% of women covered their face for the whole enrolment interview and 9% for part of the time while 82% showed their face. In late pregnancy, half of the women ate neither first nor last, 12% ate first and 35% ate last in the household. In early pregnancy, only 30% of respondents said they had some role in decision making over food purchases. Comparing women's education level with that of their husbands (using a 5-category scale) 18% of husbands were much more educated than their wives, 17% slightly more educated, 54% had same level and 11% of women were more educated than their husbands.

Analysis of nutritional status of women sampled at endpoint after pregnancy showed that 15.5% had low stature <145cm height. We used BMI cut-offs as follows: Using these cut-offs, 10.6% of women were moderately to severely underweight (BMI <17 kg/m^2), 21.5% were mildly underweight (BMI 17 - 18.49 kg/m^2), 63.7% normal (BMI 18.50 - 24.99 kg/m^2), and 4.2% overweight or obese (BMI >=25 kg/m^2). Of the index children sampled at endpoint during the monsoon season, 23.0% were moderately and 9.9% severely

585 stunted (mean HAZ -1.48 SD 1.27), 26.5% moderately and 11.0% severely underweight (mean WAZ -1.69 SD
586 1.11) and 15.9% moderately and 3.5% severely wasted (mean WHZ -1.14 SD 1.03).
587
588 Table 7. Characteristics of women enrolled in LBWSAT by study arm[1]

	n	**0. *Control***		**1. *PLA Women's group***		**2. *PLA + Cash***		**3. *PLA + Food***		**Total**	
Study arm * *(row %)*	25090	5309	*21%*	5626	*22%*	7271	*29%*	6884	*27%*	25090	
Randomisation stratum	25054										
1. small, inaccessible		1069	20%	1179	21%	1408	19%	1475	21%	5131	20%
2. small, accessible		1125	21%	1276	23%	1571	22%	1534	22%	5506	22%
3. large, inaccessible		1605	30%	1510	27%	2269	31%	1862	27%	7246	29%
4. large, accessible		1501	28%	1652	29%	2009	28%	2009	29%	7171	29%
Home of enrolment	25054										
1. Husbands home		4633	87%	4864	87%	5968	82%	5435	79%	20900	83%
2. Parental home		662	12%	732	13%	1273	18%	1435	21%	4102	16%
3. Other		5	0%	21	0%	16	0%	10	0%	52	0%
Age category of Mother	25054										
1. 11-19 years		2066	39%	2256	40%	3119	43%	2896	42%	10337	41%
2. 20-24 years		2026	38%	2109	38%	2641	36%	2552	37%	9328	37%
3. 25-29 years		875	17%	950	17%	1137	16%	1092	16%	4054	16%
4. 30-55 years		333	6%	302	5%	360	5%	340	5%	1335	5%
Age of Mother (mean sd)	25054	21.8	4.7	21.6	4.6	21.4	4.5	21.4	4.5	21.5	4.6
Age of Mother at 1st pregnancy (mean sd)	23614	17.7	2.8	17.6	2.7	17.5	2.5	17.4	2.6	17.5	2.6
Age of Mother at marriage (mean sd)	21422	15.3	2.0	15.2	2.0	15.4	2.1	15.3	2.1	15.3	2.1
Weeks of gestation at enrolment (mean sd)	25054	26.7	11.5	26.4	11.6	19.2	9.7	19.1	9.0	22.4	11.0
No of previous pregnancies	24144										
0		1897	37%	1955	36%	2474	35%	2315	36%	8641	36%
1		1325	26%	1422	26%	1858	27%	1776	27%	6381	26%
2		960	18%	1103	20%	1372	20%	1214	19%	4649	19%
3		568	11%	583	11%	722	10%	697	11%	2570	11%
4 or more		440	8%	387	7%	560	8%	516	8%	1903	8%
Female children already (mean sd)	14882	1.13	1.07	1.12	1.05	1.15	1.08	1.12	1.08	1.13	1.07
Male children already (mean sd)	14882	0.78	0.77	0.76	0.73	0.76	0.76	0.77	0.75	0.77	0.75
Adolescent <=19 years	21101	1842	39%	1941	40%	2585	43%	2298	42%	8666	41%
Mother education	24162										
1. never went to school		3479	67%	3353	62%	4404	63%	4252	65%	15488	64%

	n	0. *Control*		1. *PLA Women's group*		2. *PLA + Cash*		3. *PLA + Food*		Total	
2. primary to lower sec'		1003	19%	1194	22%	1456	21%	1263	19%	4916	20%
3. secondary and above		717	14%	903	17%	1133	16%	1005	15%	3758	16%
Husband education	24161										
1. never went to school		2639	51%	2520	46%	3451	49%	3028	46%	11638	48%
2. primary to lower sec'		1328	26%	1515	28%	1790	26%	1913	29%	6546	27%
3. secondary and above		1232	24%	1415	26%	1752	25%	1578	24%	5977	25%
Asset quintile	23861										
1 (poorest)		1079	21%	1088	20%	1319	19%	1284	20%	4770	20%
2		996	19%	1091	20%	1326	19%	1359	21%	4772	20%
3		1087	21%	1038	19%	1342	20%	1337	21%	4804	20%
4		1041	20%	991	18%	1444	21%	1337	21%	4813	20%
5 (least poor)		973	19%	1166	22%	1442	21%	1121	17%	4702	20%
1 SD change in Wealth score (mean sd)	23861	-0.034	1.003	0.030	1.055	0.041	1.026	-0.040	0.983	0.001	1.017
Family member migrated for labour	21099	2273	48%	1956	41%	2549	42%	2580	47%	9358	44%
Husband migrated overseas	21420	1481	31%	1284	26%	1431	23%	1503	27%	5699	27%
Caste 3 broad groups	25054										
1. Dalit/Muslim-disadvantaged		2109	40%	1934	34%	2441	34%	2461	36%	8945	36%
2. Janjati/Other Terai castes - middle		2236	42%	2534	45%	2824	39%	2904	42%	10498	42%
3. Yadav/Brahmin least disadvantaged		955	18%	1149	20%	1992	27%	1515	22%	5611	22%
Deprived of toilet (none/shared)	24159	3968	76%	3901	72%	5060	72%	4941	76%	17870	74%
Owns land	24159	3218	62%	3555	65%	4780	68%	4290	66%	15843	66%
Uses mainly gas or other non-biomass fuels	21097	213	4%	266	6%	328	5%	191	3%	998	5%
Mother is Hindu (vs. Muslim/Other)	25053	4299	81%	4782	85%	6056	83%	5854	85%	20991	84%
Months of adequate HH food provisioning in 12m before early pregnancy (mean sd)	3954	11.8	0.7	11.8	0.9	11.8	1.3	11.9	0.4	11.8	0.9
Household dietary diversity score (out of 12) in early pregnancy (mean sd)	3954	6.9	1.5	7.1	1.6	7.2	1.6	7.1	1.5	7.1	1.5
Mother's height (mean sd)	16295	150.24	5.20	150.38	5.50	150.59	5.30	150.52	5.60	150.45	5.40

	n	0. *Control*		1. *PLA Women's group*		2. *PLA + Cash*		3. *PLA + Food*		Total	
Mother's short stature	16295	542	15.60%	570	16.0%	708	14.6%	707	16.0%	2527	15.5%
Total		5309	100%	5626	100%	7271	100%	6884	100%	25090	100%

[1] column percentages presented to show characteristics within each arm

Potential sources of bias that may be adjusted for in secondary analyses

Potential sources of bias associated with the different data collection instruments have been itemised in Table 3 column entitled "factors affecting capture". The following section provides a closer look at characteristics by in-migration and provides analysis to measure the extent that respondents differed by whether or not they were captured in each of the main follow-up tools.

1) In-migration

Table 8 provides a comparison of the characteristics of in-migrators and permanent residents, as per the trial inclusion definition of permanent resident of having been captured in the population census or being newly wedded and living in their husband's home. Using capture during the baseline population census or being newly married and living with the husband's family as a definition of permanent residency, we found 18% in-migrators in *control* and PLA only arms, whereas they comprised 23% in PLA + cash and 26% in *PLA + food* arms. In-migrators tended to be women who were staying in their parental homes (*maiti*) rather than in the household of their husbands (*sasural*). In *control* and PLA only arms, 13% of participants enrolled from their maiti compared with 17% in *PLA + cash* and 20% in *PLA + food* arms. Compared with permanent residents, in-migrators tended to be younger (56% adolescent as compared with 37% amongst permanent residents), married younger, and more likely to be primigravida, Muslim, slightly poorer and have primary or secondary education.

Table 8. Comparison of in-migrators and permanent residents

	N=	In-migrators		Permanent residents		Total	
Study arm mother enrolled in	25090						
0. *Control*		975	18%	4325	22%	5300	21%
1. *PLA Women's groups only*		1024	19%	4593	23%	5617	22%
2. *PLA + Cash*		1678	31%	5579	28%	7257	29%
3. *PLA + Food*		1772	33%	5108	26%	6880	27%
Randomisation stratum	25054						
1. small, inaccessible		1029	19%	4102	21%	5131	20%
2. small, accessible		1296	24%	4210	21%	5506	22%
3. large, inaccessible		1601	29%	5645	29%	7246	29%
4. large, accessible		1523	28%	5648	29%	7171	29%

	N=	In-migrators		Permanent residents		Total	
Home where mother enrolled	25054						
1. Husbands home (sasural)		2149	39%	18751	96%	20900	83%
2. Parental home (Maiti)		3261	60%	841	4%	4102	16%
3. Other		39	1%	13	0%	52	0%
Age category of Mother	25054						
1. 11-19 years		3044	56%	7293	37%	10337	41%
2. 20-24 years		1712	31%	7616	39%	9328	37%
3. 25-29 years		532	10%	3522	18%	4054	16%
4. 30-55 years		161	3%	1174	6%	1335	5%
Age of Mother (mean sd)	25054	20.2	4.1	21.9	4.6	21.5	4.6
Age of Mother at 1st pregnancy (mean sd)	23614	17.6	2.4	17.5	2.7	17.5	2.6
Age of Mother at marriage (mean sd)	21422	15.6	2.0	15.2	2.1	15.3	2.1
Weeks of gestation at enrolment (mean sd)	25054	22.4	10.0	22.0	11.3	22.4	11.0
No of previous pregnancies (up to 4 or more) before index pregnancy	24144						
0		2676	52%	5965	31%	8641	36%
1		1233	24%	5148	27%	6381	26%
2		732	14%	3917	21%	4649	19%
3		281	6%	2289	12%	2570	11%
4		184	4%	1719	9%	1903	8%
Total girls alive during index pregnancy (mean sd)	14882	0.93	0.98	1.16	1.09	1.128	1.073
Total boys alive during index pregnancy (mean sd)	14882	0.75	0.73	0.77	0.76	0.766	0.752
Pregnant mother adolescent <=19 years	21101	2622	56%	6044	37%	8666	41%
Mother education 3groups	24162						
1. never went to school		3121	61%	12367	65%	15488	64%
2. primary to lower secondary		1096	21%	3820	20%	4916	20%
3. secondary and above		886	17%	2872	15%	3758	16%
Husband education 5groups	24161						
1. never went to school		2531	50%	9107	48%	11638	48%
2. primary to lower secondary		1266	25%	5280	28%	6546	27%
3. secondary and above		1306	26%	4671	25%	5977	25%
Asset quintile based on 12 items-without education	23861						
1		1107	22%	3663	19%	4770	20%
2		1033	20%	3739	20%	4772	20%
3		1005	20%	3799	20%	4804	20%
4		995	20%	3818	20%	4813	20%
5		914	18%	3788	20%	4702	20%
1 SD change in Wealth score of 12 items without education (mean sd)	23861	-0.056	1.010	0.016	1.018	0.001	1.017
Family member migrated for labour	21099	2263	48%	7095	43%	9358	44%
Husband migrated overseas	21420	1347	29%	4352	26%	5699	27%
Caste 3 broad groups	25054						

	N=	In-migrators		Permanent residents		Total	
1. Dalit/Muslim- disadvantaged		2114	39%	6831	35%	8945	36%
2. Janjati/Other Terai castes - middle		2115	39%	8383	43%	10498	42%
3. Yadav/Brahmin least disadvantaged		1220	22%	4391	22%	5611	22%
Deprived of toilet (no improved toilet or toilet is shared)	24159	3810	75%	14060	74%	17870	74%
Owns land	24159	3212	63%	12631	66%	15843	66%
Uses mainly gas or other non-biomass fuels	21097	215	5%	783	5%	998	5%
Mother is Hindu (vs. Muslim/Other)	25053	4402	81%	16589	85%	20991	84%
Months of adequate HH food provisioning in 12m before Early Pregnancy (mean sd)	3954	11.9	0.7	11.8	1.0	11.8	0.9
Household dietary diversity score (out of 12) in early pregnancy (mean sd)	3954	7.2	1.6	7.1	1.5	7.1	1.5
Mother's height (mean sd)	16295	150.2	5.3	150.5	5.4	150.5	5.4
Mother's short stature	16295	230	16%	2297	15%	2527	16%
Total		5449	100%	19605	100%	25090	100%

2) Analyses comparing those captured with those not captured

Comparisons of captured versus non-captured pregnant woman's age at marriage, wealth score, age at enrolment, height, and educational status and the education of her husband are presented for each of the follow-up questionnaires after enrolment (questionnaire codes E, L, D, F, M, B) in Table 9. Differences in maternal age, age at marriage, age at first pregnancy and maternal height and the significance of differences analysed using linear regressions (adjusted for clustering) are presented, including differences in age of respondent groups in months. Also, results of chi-squared tests comparing the educational status of captured and not captured mothers and their husbands are provided.

Women captured in early pregnancy were 2 months younger, 0.06 wealth score units less poor than those not captured. Other differences were not significant. Women captured in late pregnancy were 2 months older, were married 1.5 months younger, were 0.07 wealth score units poorer than those not captured. Other differences in husband's education were very small.

Women captured in at birth were 4 months older, were married 2.4 months younger, were 0.28 cm shorter and were 0.13 wealth score units poorer than those not captured. More than expected of those captured had no schooling and fewer had primary or secondary education. Slightly more husbands were uneducated in the captured group, but differences were small. These differences were probably associated with migration patterns to the parental home for delivery / postpartum. Women captured in at the >42 days postpartum timepoint were 4 months older, were married 1 months younger, were 0.32 cm shorter than those not captured. The woman's education and poverty level didn't differ but more husbands of those captured had secondary education than expected. The differences were probably associated with migration patterns to the parental home for delivery / postpartum.

Women captured with the mother and baby questionnaires at endpoint differed from those who were not captured more than at other time points and in different ways. They were 18 months older, married 4 months younger, were 0.15 wealth score points poorer. Compared with expected frequencies, captured women and their husbands were more likely to have never been to school. It could be that the trial participation incentives for reporting for measurements were more appealing to the older women from less educated, poorer families.

Results of univariate logistic regression analyses analysing the odds of capturing a particular questionnaire in association with socioeconomic covariates are presented in Table 10. For each standard deviation increment in wealth score the odds of capture was slightly and significantly lower for L, D, M and B questionnaires and was slightly higher at early pregnancy (E). In most cases poorer women were more likely to be captured than less poor but odds ratios are close to zero in all cases. For all tools except early pregnancy, the odds of capture increased very slightly but significantly with every year of age at enrolment and increased very slightly but significantly with every year of age of marriage. Meanwhile mothers with secondary education and those whose husband's had primary secondary education had lower odds of capture. The effect was especially strong at endpoint when the ORs relative to uneducated women were 0.75 for primary educated mothers and 0.61 for secondary educated mothers. When all the variables were co-adjusted for included in the models secondary education of mother and primary education of husband associated with capture at endpoint, age was still associated with capture for a E, D, and F, assets were still associated with capture for E, L, and D but not at endpoint.

Whilst some small differences were found between captured and non-captured groups for selected variables, we conclude that none of these differences are of importance for secondary analyses of the data. Relative to the span of the variable in the whole sample all of the differences are extremely small. These overall differences are between group, while within each group there is still a lot of variability.

656 Table 9. Comparisons of age at marriage, wealth score, age of pregnant mother at enrolment, maternal height, maternal education and husband's education in
657 captured and non-captured households, with tests for differences using regression, t tests and chi-squared tests.

			n=21422	*n*=23861	*n*=25054	*n*=16295		*n*=24162		*n=24161*	
Questionnaire			Age married	Wealth score	Age of pregnant woman	Maternal height	Level of education	Maternal education n	%	Husband's education n	%
Early Pregnancy (anthro, diet, empowerment, exposures) n=3975, 15.8%	Not captured	Mean	15.3	-0.01	22.6	150.5	1. never went to school	13078	*64.4*	9812	*48.3*
		SD	2.0	1.91	4.6	5.4	2. primary to lower secondary	4109	*20.2*	5478	*27.0*
		n	17566	20023	21210	12413	3. secondary and above	3133	*15.4*	5029	*24.8*
	Captured	Mean	15.3	0.05	22.4	150.3	1. never went to school	2499	*62.9*	1897	*47.8*
		SD	2.1	1.89	4.6	5.5	2. primary to lower secondary	828	*20.9*	1100	*27.7*
		n	3969	3967	3975	3975	3. secondary and above	645	*16.2*	975	*24.6*
	Significance test	p	0.990	0.040	0.029	0.215	Chi-squared	3.16		0.91	
		difference	-0.04	-0.06	0.17	0.1	p	0.206		0.634	
		difference in (unit)	*-0.51 months*		*2.03 months*						
Questionnaire			Age married	Wealth score	Age of pregnant woman	Maternal height	Level of education	Maternal education n	%	Husband's education n	%
Late Pregnancy (anthro, diet,	Not captured	Mean	15.3	0.01	22.5	150.5	1. never went to school	13710	*64.1*	10293	*48.1*
		SD	2.0	1.91	4.6	5.4	2. primary to lower secondary	4351	*20.3*	5802	*27.1*
		n	18679	21089	22272	13537	3. secondary and above	3328	*15.6*	5293	*24.8*
		Mean	15.2	-0.06	22.7	150.4	1. never went to school	1867	*64.3*	1416	*48.8*

		difference in (unit)	*1.45 months*		*-2.27 months*						
Questionnaire			Age married	Wealth score	Age of pregnant woman	Maternal height	Level of education	Maternal education		Husband's education	
								n	%	n	%
Delivery (birth anthro of baby, delivery details) n=8001, 31.8%	Not captured	Mean	15.4	0.04	22.4	150.6	1. never went to school	10373	*63.5*	7810	*47.8*
		SD	2.0	1.93	4.5	5.4	2. primary to lower secondary	3371	*20.6*	4473	*27.4*
		n	13806	16058	17184	10294	3. secondary and above	2600	*15.9*	4060	*24.8*
	Captured	Mean	15.17	-0.09	22.76	150.3	1. never went to school	5204	*65.5*	3899	*49.1*
		SD	2.10	1.85	4.71	5.4	2. primary to lower secondary	1566	*19.7*	2105	*26.5*
		n	7729	7932	8001	6094	3. secondary and above	1178	*14.8*	1944	*24.5*
	Significance test	p	<0.001	<0.001	<0.001	<0.001	Chi-squared	9.66		3.65	
		difference	0.20	0.13	-0.35	0.28	p	0.008		0.161	
		difference in (unit)	*2.40 months*		*-4.15 months*						
Questionnaire			Age married	Wealth score	Age of pregnant woman	Maternal height	Level of education	Maternal education		Husband's education	
								n	%	n	%
Post-neonatal (pregnancy, postpartum, neonatal morbidity, baby and mother anthro) n=3056, 12.1%	Not captured	Mean	15.31	0.00	22.48	150.5	1. never went to school	13608	*64.0*	10231	*48.1*
		SD	2.05	1.91	4.57	5.4	2. primary to lower secondary	4348	*20.5*	5813	*27.3*
		n	18572	20961	22129	13722	3. secondary and above	3303	*15.5*	5214	*24.5*
	Captured	Mean	15.23	0.00	22.81	150.2	1. never went to school	1969	*64.9*	1478	*48.7*
		SD	2.06	1.85	4.68	5.3	2. primary to lower secondary	589	*19.4*	765	*25.2*
		n	2963	3029	3056	2666	3. secondary and above	475	*15.7*	790	*26.1*
	Significance test	p	0.002	0.904	0.008	0.047	Chi-squared	1.76		7.10	
		difference	0.08	0.00	-0.32	0.32	p	0.415		0.029	
		difference	*0.93*		*-3.87*						

		in (unit)	*months*			*months*						

Questionnaire			Age married	Wealth score	Age of pregnant woman	Maternal height	Level of education	Maternal education		Husband's education	
								n	%	n	%
Mother Questionnaire at endpoint (anthro, delivery details) n=19017, 75.5%	Not captured	Mean	15.56	0.12	21.40	150.2	1. never went to school	3063	*57.7*	2390	*45.0*
		SD	2.03	1.95	3.88	5.4	2. primary to lower secondary	1208	*22.8*	1379	*26.0*
		n	5283	5231	6168	1218	3. secondary and above	1036	*19.5*	1537	*29.0*
	Captured	Mean	15.21	-0.03	22.89	150.5	1. never went to school	12514	*65.9*	9319	*49.1*
		SD	2.05	1.89	4.74	5.4	2. primary to lower secondary	3729	*19.6*	5199	*27.4*
		n	16252	18759	19017	15170	3. secondary and above	2742	*14.4*	4467	*23.5*
		p	<0.001	<0.001	<0.001	0.300	Chi-squared	132.13		66.66	
	Significance test	difference	0.34	0.15	-1.49	-0.24	p	<0.001		<0.001	
		difference in (unit)	*4.12 months*		*-17.86 months*						

Questionnaire			Age married	Wealth score	Age of pregnant woman	Maternal height	Level of education	Maternal education		Husband's education	
								n	%	n	%
Baby Questionnaire at endpoint (anthro, IYCF, morbidity) n=19222, 76.3%	Not captured	Mean	15.57	0.12	21.39	150.1	1. never went to school	2925	*57.4*	2291	*45.0*
		SD	2.03	1.95	3.88	5.4	2. primary to lower secondary	1158	*22.7*	1320	*25.9*
		n	5095	5026	5963	1326	3. secondary and above	1012	*19.9*	1483	*29.1*
	Captured	Mean	15.21	-0.03	22.87	150.5	1. never went to school	12652	*65.9*	9418	*49.1*
		SD	2.05	1.89	4.73	5.4	2. primary to lower secondary	3779	*19.7*	5258	*27.4*
		n	16440	18964	19222	15062	3. secondary and above	2766	*14.4*	4521	*23.6*
	Significance	p	<0.001	<0.001	<0.001	0.050	Chi-squared	140.69		67.57	

	test	difference	0.36	0.15	-1.48	-0.39	p	<0.001	<0.001
		difference in (unit)	*4.33 months*		*-17.77 months*				

Table 10. Odds of capturing different follow-up questionnaires in relation to different indicators of socioeconomic status or human capital, applying logistic regression.

	Early Pregnancy (anthro, diet, empowerment, exposures) n=3975, 15.8%			Late Pregnancy (anthro, diet, empowerment, exposures) n=2913, 11.6%			Delivery (birth anthro of baby, delivery details) n=8001, 31.8%			Post-neonatal (pregnancy, postpartum, neonatal morbidity, baby and mother anthro) n=3056, 12.1%			Mother Questionnaire at endpoint (anthro, delivery details) n=19017, 75.5%			Baby Questionnaire at endpoint (anthro, IYCF, morbidity) n=19222, 76.3%		
Characteristic	OR	95% CI	p	OR	95% CI	p	OR	95% CI	p	OR	95% CI	p	OR	95% CI	p	OR	95% CI	p
1 SD change in Wealth score[1]	1.04	(1.00,1.08)	0.0380	0.93	(0.89,0.97)	0.0008	0.93	(0.90,0.96)	0.0008	1.00	(0.96,1.04)	0.9178	0.92	(0.89,0.94)	<0.0001	0.92	(0.89,0.95)	<0.0001
Age of pregnant mother [2]	0.99	(0.98,1.00)	0.0290	1.01	(1.00,1.02)	0.0176	1.02	(1.01,1.02)	0.0176	1.01	(1.00,1.02)	0.0105	1.09	(1.08,1.09)	<0.0001	1.09	(1.08,1.09)	<0.0001
Age of marriage[3]	1.00	(0.98,1.02)	0.9530	0.97	(0.95,0.99)	0.0009	0.96	(0.94,0.97)	0.0009	0.97	(0.95,0.99)	0.0019	0.92	(0.90,0.93)	<0.0001	0.91	(0.90,0.93)	<0.0001
Mother's education[4]																		
1. never went to school	ref			ref			ref			ref			ref			ref		
2. primary to lower secondary	1.07	(0.98,1.17)	0.3392	0.96	(0.86,1.06)	0.1863	0.93	(0.87,1.00)	0.1863	0.96	(0.87,1.06)	0.6629	0.75	(0.69,0.81)	<0.0001	0.75	(0.69,0.81)	<0.0001
3. secondary and above	0.00	(0.94,1.15)	<0.0001	0.90	(0.80,1.01)	<0.0001	0.88	(0.81,0.95)	<0.0001	0.97	(0.87,1.09)	<0.0001	0.63	(0.58,0.68)	<0.0001	0.61	(0.56,0.66)	<0.0001
Husband's education[5]																		
1. never went to school	ref			ref			ref			ref			ref			ref		
2. primary to lower secondary	1.03	(0.94,1.12)	0.2912	0.92	(0.84,1.02)	0.0430	0.90	(0.84,0.97)	0.0430	0.90	(0.82,1.00)	0.1186	0.96	(0.89,1.03)	<0.0001	0.96	(0.89,1.03)	<0.0001
3. secondary and above	0.95	(0.87,1.04)	<0.0001	0.89	(0.80,0.98)	<0.0001	0.92	(0.86,0.99)	<0.0001	0.99	(0.90,1.10)	<0.0001	0.73	(0.68,0.79)	<0.0001	0.73	(0.67,0.78)	<0.0001
Mother's height in cm[6]	1.00	(0.99,1.00)	0.2280	1.00	(0.99,1.01)	0.6570	0.99	(0.99,1.00)	0.6570	0.99	(0.98,1.00)	0.1007	1.00	(0.99,1.02)	0.3939	1.01	(1.00,1.02)	0.8610

Adjusted analysis co-adjusting for wealth score, age of PW, education of mother and husband:	Age and wealth score and husband secondary education highly associated with q capture. Overall p< 0.0001	Education and age no longer significant but wealth score highly associated with q capture. Overall p= 0.0159	Education no longer significant but age and wealth score highly associated with q capture. Overall p= 0.0159	Only age of mother significant in adjusted analysis. Overall p< 0.0001	Assets no longer associated but age, secondary education of mother and primary education of husband associated with q capture. Overall p< 0.0001	Assets no longer associated but age, secondary education of mother and primary education of husband associated with q capture. Overall p< 0.0001

661 1 SD change in Wealth score[1] *n*= 23,861; Age of pregnant mother [2] *n*= 25054; Age of marriage[3] *n*=21422; Mother's education[4] *n*= 24162; Husband's education[5] *n*= *24161;* Mother's height[6] *n*=16295

Conclusion

There are many challenges in implementing large-scale cluster-randomised controlled trial in the plains of Nepal and these affected rates of data capture, especially when several timebound follow-up data collection occasions are required for each participant as was the case in LBWSAT. Although some differences between in-migrated and permanent residents were found and captured and non-captured women differed slightly, on the whole differences were not large enough to be of concern for the conduct of secondary analyses on this study cohort. We surmise that any biases are not problematic because the magnitude of the differences between captured and non-captured women are trivial relative to the range in the entire sample. The differences are only statistically significant because we have a large sample size.

Acknowledgments

We thank all the researchers involved in the primary collection and analysis of data in the Low Birth Weight South Asia Trial including Bhim P. Shrestha, Helen Harris-Fry, Lu Gram, Sarah Style, B. James Beard, Aman Sengupta, Sonali Jha, Anjana Rai, Vikas Paudel, Raghbendra Sah, Puskar Paudel, Andrew Copas, Bishnu Bhandari, Rishi Neupane, Joanna Morrison, Anni-Maria Pulkki-Brännström, Jolene Skordis-Worrall, Machhindra Basnet, Saskia de Pee, Andrew Hall, Jayne Harthan, Meelan Thondoo, Sonja Klingberg and Janice Messick.

Thank you to the District Public Health Offices of Dhanusha and Mahottari and their health workers at all levels who supported the trial on the ground. We thank: the Trial Steering Committee: Prof Keith West (chair), Prof Simon Cousens (statistical adviser), Dr Elaine Ferguson, Prof Sonia Bhalotra, and Dr Shobha Rao; the Data Monitoring Committee: Dr Rajiv Bahl, Dr Andrea Rehman, Dr Lynette Neufeld, Dr Amod Poudyal and Dr Rita Thapa; and advisors from the Donor and Government Advisory Group in Nepal, including the Family Health Division, Child Health Division and Nutrition Section of the Nepal Ministry of Health.

We thank the following for their efforts in implementation of the study: Shiva Shanker Chaube, Neha Sharma, Ritesh Shrestha and all the field team of MIRA. Lalbabu Kushuwa and Dhruba Devkota of Save the Children helped us to implement the food and cash transfers. Pramila Ghimire, Srijana Nakermi, Marco Cavalcante, and Allison Prather of World Food Programme Nepal provided Super Cereal for the study and advice on managing its distribution. We are grateful to administrators in UCL Institute for Global Health, Save the Children UK, and MIRA for administrative and managerial support.

We thank all the committed staff of MIRA who made the study possible, including more than 66 data collectors and their 16 supervisors, 720 volunteer enumerators, 40 intervention implementers, 539

volunteer mobilisers and 540 Female Community Health Volunteers responsible for facilitating women's group and social transfer interventions, and the thousands of women who agreed to participate.

Author contributions

NS analysed the data and drafted the manuscript. JW and NS conceived the paper topic and analyses to be undertaken. NS and DSM designed the original LBWSAT trial and directed implementation and data collection. NS and DSM provided day-to-day management and necessary logistical support to trial field activities. All authors reviewed and provided critical comments on the manuscript.

Literature cited

1. Saville, N.M., et al., *Protocol of the Low Birth Weight South Asia Trial (LBWSAT), a cluster-randomised controlled trial testing impact on birth weight and infant nutrition of Participatory Learning and Action through women's groups, with and without unconditional transfers of fortified food or cash during pregnancy in Nepal.* BMC Pregnancy and Childbirth, 2016. **16**(1): p. 320.
2. Style, S., et al., *Experiences in running a complex electronic data capture system using mobile phones in a large-scale population trial in southern Nepal.* Glob Health Action, 2017. **10**(1): p. 1330858.
3. Saville, N.M., et al., *Impact on birth weight and child growth of Participatory Learning and Action women's groups with and without transfers of food or cash during pregnancy: Findings of the low birth weight South Asia cluster-randomised controlled trial (LBWSAT) in Nepal.* PLoS One, 2018. **13**(5): p. e0194064.
4. Harris-Fry, H.A., *Intra-household food allocation in rural Nepal*, in *Institute for Global Health*. 2016, University College London: PhD Thesis.
5. Harris-Fry, H., et al., *Development and validation of a photographic food atlas for portion size assessment in the southern plains of Nepal.* Public Health Nutrition, 2016. **FirstView**: p. 1-13.
6. Harris-Fry, H., et al., *Smartphone tool to collect repeated 24 h dietary recall data in Nepal.* Public Health Nutrition, 2017: p. 1-13.
7. Harris-Fry, H.A., et al., *Status and determinants of intra-household food allocation in rural Nepal.* Eur J Clin Nutr, 2018. **72**(11): p. 1524-1536.
8. Harris-Fry, H.A., et al., *Participatory Women's Groups with Cash Transfers Can Increase Dietary Diversity and Micronutrient Adequacy during Pregnancy, whereas Women's Groups with Food Transfers Can Increase Equity in Intrahousehold Energy Allocation.* J Nutr, 2018. **148**(9): p. 1472-1483.

9. Gram, L., et al., *Validating an Agency-based Tool for Measuring Women's Empowerment in a Complex Public Health Trial in Rural Nepal.* Journal of human development and capabilities, 2017. **18**(1): p. 107-135.
10. Gram, L., et al., *Do Participatory Learning and Action Women's Groups Alone or Combined with Cash or Food Transfers Expand Women's Agency in Rural Nepal?* J Dev Stud, 2019. **55**(8): p. 1670-1686.
11. Gram, L., *Women's Empowerment in a Complex Public Health Intervention in Rural Nepal*, in *Institute for Global Health*. 2018, University College London PhD Thesis.
12. Gram, L., et al., *Revisiting the patriarchal bargain: The intergenerational power dynamics of household money management in rural Nepal.* World Development, 2018. **112**: p. 193-204.
13. Gram, L., et al., *'There is no point giving cash to women who don't spend it the way they are told to spend it' - Exploring women's agency over cash in a combined participatory women's groups and cash transfer programme to improve low birthweight in rural Nepal.* Soc Sci Med, 2019. **221**: p. 9-18.
14. Morrison, J., et al., *Formative qualitative research to develop community-based interventions addressing low birth weight in the plains of Nepal.* Public Health Nutr, 2018. **21**(2): p. 377-384.
15. Morrison, J., et al., *A report of the Low Birth Weight South Asia Trial process evaluation exploring mechanisms of how interventions may have worked*. 2020, Univsersity College London: UCL discovery.
16. Marphatia, A.A., et al., *How Much Education Is Needed to Delay Women's Age at Marriage and First Pregnancy?* Frontiers in Public Health, 2020. **7**.
17. Saville, N.M., et al., *Comprehensive analysis of the association of seasonal variability with maternal and neonatal nutrition in lowland Nepal.* PlosOne, under review.
18. Saville, N.M., et al., *Do mothers have larger energy budgets for investing in sons compared to daughters? Evidence from lowland Nepal.* American Journal of Human Biology, under review.
19. Acharya, D.R., et al., *Women's autonomy in household decision-making: a demographic study in Nepal.* Reprod Health, 2010. **7**: p. 15.
20. Acharya, M. and L. Bennett, *Women and the subsistence sector.* World Bank Staff Working Papers, 1983(526).
21. Human Rights Watch, *"Our Time to Sing and Play" - Child Marriage in Nepal*. 2016.
22. Bista, D.B., *People of Nepal*. 1972: Ratna Pustak Bhandar Kathmandu.
23. Sob, D., *The Situation of the Dalits in Nepal: Prospects in a New Political Reality.* Contemporary Voice of Dalit, 2012. **5**(1): p. 57-62.
24. Gellner, D.N., *Caste, Ethnicity and Inequality in Nepal.* Economic and Political Weekly, 2007. **42**(20): p. 1823-1828.
25. World Bank Group, *Migration and Remittances Factbook 2016*. 2016.

26. Ministry of Health Nepal, *Nepal Demographic and Health Survey 2016: Key Indicators.* , N.E.a. ICF, Editor. 2017: Kathmandu, Nepal: Ministry of Health, Nepal.
27. Gram, L., et al., *Validating an Agency-based Tool for Measuring Women's Empowerment in a Complex Public Health Trial in Rural Nepal.* J Human Dev Capabil, 2017. **18**(1): p. 107-135.
28. Ministry of Health Nepal, New ERA, and ICF, *Nepal Demographic and Health Survey 2016.* . 2017, Ministry of Health, Nepal.