A Bayesian Estimation of Child Labour in India

Jihye Kim[1] · Wendy Olsen [2]· Arkadiusz Wiśniowski[3]

Abstract

Child labour in India involves the largest number of children in the world. In 2011, this number was estimated to be 11.8 million children for ages 5 to 17 according to the latest Indian Census. However, our estimation on child labour using a combined-data approach is higher than that of the Indian Census, which is about 16 million for ages 5 to 17. How to measure the prevalence of child labour varies according to divergent opinions across international agencies such as the International Labour Organization (ILO) and the United Nations Children's Fund (UNICEF). In this study, we use the ILO's methodology to define hazardousness of work and UNICEF's time threshold for domestic work. The specific aims of this study are to estimate the prevalence of child labour in the age group 5 to 17 and to suggest a combined-data approach using a Bayesian method to improve the estimation of child labour. This study uses the most recent National Sample Survey on Employment and Unemployment and the India Human Development Survey, comparing and combining them with the reported figure of child labour from the Indian Census. The combined-data approach provides a way to improve accuracy and potentially reduce measurement error. This method also smooths the variation between ages and provides more reliable estimates of child labourers.

Keywords: Child labour, Bayesian Estimation, Combining data, India

[1] Jihye Kim, jihye.kim@postgrad.manchester.ac.uk
[2] Wendy Olsen, wendy.olsen@manchester.ac.uk
[3] Arkadiusz Wiśniowski, a.wisniowski@manchester.ac.uk

1 Introduction

India has the largest number of child labourers in the world, with an estimated 12.7 million children aged 5 to 14 in 2001 and 4.5 million in 2011[4]. However, the figure is varied according to datasets or definitions. For example, the number of child labourers reached 11.8% among children ages 5 to 14 in 2012, calculated by UNICEF, which is about 29 million in total (UNICEF, 2013).[5] The reduction over time arises through increased household income, but many industries and farm work are still related to child labour. Domestic labour is one of the most prevalent types of child work but whether it should be included as a child labourer is still debatable.

There has been an extensive discussion to define child labour. The major issues are the differentiation between child labour and child work and setting up age boundaries. The ILO and UNICEF have made achievements to lead an international agreement, but their definitions still have some distance from the definition at the national level. In 2016, the Indian government amended the Child Labour Act to adopt a strict banning policy for any children under the age of 18, following the international standard. However, the Amendment Act still permits domestic work for long hours, which is regarded as hazardous activities from the international regulations.

The first focus of this paper is on how different definitions of child labour affect measurements of child labour. We would like to see the differences of definitions on child labour of the two major stakeholders – the ILO and UNICEF– and modify them to meet the situation of India. We define child labour as any work that is harmful to children's development, including domestic and non-domestic work that requires considerable time. Then, we explore the application of these definitions to measure the number of child labourers.

Besides being a matter of definition, there is a high demand to address how to measure the number of children in labouring status with accuracy. A measurement error, a departure from the true value of measurement and the value provided (Groves et al., 2011, p.52), is highly related with the undercounting of child labour. A possible undercounting matter that might have occurred is intentional or unintentional misresponse, for example, parents' non-response due to their increasing awareness about the illegality of child labour (Basu, 1999) and also children who are involved in labour but are not recognised as child labour (Chaudhri et al. 2003; Chaudhri and Wilson 2000). A measurement error might be raised by the limitation of variables that are necessary to estimate child labour such as lack of information about precise working hours or working conditions.

Furthermore, child labour is sparse, especially among younger age groups. There could be overrepresentation when we calculate its number only using a sampling weight of a survey. A model-

[4] India Census; calculated with the number of main workers.

[5] The population of children aged 5 to 14 is estimated at about 253 million by the Indian Census, 2011

based approach can provide more accurate information regarding the number of child labour. The National Sample Survey (NSS) and India Human Development Survey (IHDS) provide qualified datasets relating to child labour, with the most recent data sources in 2011/12 and have different strengths in explaining child labour. We prefer to use the Indian Census 2011 as auxiliary information. Despite its large coverage of the population, it provides only aggregated number of child labour based on the two broad categories - main (working more than six months) or marginal (working more than three months but less than six months) workers.

This paper intends to address how to reduce discrepancies between the truth and estimated number of child labour. As a way of accurately estimating child labour, we would like to suggest a combination of these two datasets – the IHDS and the NSS. A Bayesian hierarchical model can be used to combine different datasets and provides more precise estimates for an unknown parameter.

This research will make a significant contribution to child labour studies in several ways. It will provide an accurate number of child labourers based on an appropriate definition and estimate by a proper model, overcoming the limitations of using a single dataset. This study is the first trial to apply a Bayesian model to measure child labour by providing an advantage to see the relationship between age and child labour.

This study aims to apply the international definition of child labourers, with which we can consider children working in hazardous industries or occupatioins, or child labourers in the domestic sector. Using the definition, we would like to provide accurate estimations of child labour in the age group 5 to 17 and carefully look at the age pattern of child labour in India. Lastly, this study aims to reveal whether the Bayesian combined-data approach is efficient in reducing survey errors and measurement errors regarding estimates of child labour. Section 2 introduces backgrounds of the research and discussions on the definition of child labour, which will justify our definition of child labour. Section 3 explains the methodology of this study including a review on a Bayesian statistical method and describes our key models. Lastly, we summarise the result of this study in section 4 and provide a key political implication in section 5.

2 Backgrounds on Child Labour Debates

Earlier research on child labour has defined child labour usually in an extensive way to put a stress on legislative intervention. Weiner (1991) suggests that because children who are not in school are potential child labourers (1991, p.3007). Grootaert and Kanbur (1995) admit that how to define exploitation of child labour is the main question for a policy towards child labour and point out the unclear definition of child labour as a challenge for study (1995, p.188).

Basu and Van (1998) define child labour as any economic activity. Basu (1999) keeps the broad definition of child labour and includes even part-time workers. In a later study, he also includes domestic labour as child labour (Basu et al., 2010). Ray analyses wage and child labour hours together for children who are in full-time labour outside a home (Ray, 2000, p.350), which is narrower than Basu's definition of 'economically active children'. In many studies, child labour is defined by a child's working as a principal activity [6] (Kambhampati and Rajan, 2008; Das and Mukherjee, 2007). Using a principal status as a definition of child labour might exclude any significant types of child labour if they are recorded as a second status.

In spite of much literature regarding the concept of child labour, it is difficult to find literature that focuses on the measurement of the number of child labourers. The recent attempt to count child labourers is meaningful as it integrates economic activities and domestic work, and also includes 'nowhere children', children who are neither in school nor work, as child labourers (Giri and Singh, 2016). However, considering all 'nowhere children' as potential child labourers might bring misunderstanding about reality (Lieten, 2002). Furthermore, the use of a simple weighted calculation of a survey data is not free from a sampling problem of sub-strata (Chaudhri and Wilson (2000, p.13). We investigate why applying weight is problematic in our case in the later section (See Section 4.1).

2.1 Review of International Definitions

There has been an effort to define child labour by international agencies. The wide-used definitions are described – specifically, those of the ILO and UNICEF. They provide essential implication to consider child labour in a global standard, but they have slightly different strengths in capturing child labour. Hence, we prefer to make our definition of child labour by integrating them.

2.1.1 ILO Definition of Child Labour

In the ILO definition, child labour means children in employment excluding children who are in permitted light work and those above the minimum age (ILO, 2017). The ILO's focus is on the hazardousness of work that children are engaged with (Omoike, 2010). The ILO recognises the significance of hazardous unpaid household activities as well, but it does not provide an explicit method to contain domestic chores for calculation of child labour.

According to the ILO's minimum age standard[7], the minimum age should not be less than the age of completion of compulsory schooling and in any case, no less than 15 years old (14 for developing countries), which is agreed by 166 countries in 2014. Furthermore, the ILO shows a concern for children who are 16 or 17 years old. The ILO's Worst Forms of Child Labour Convention clarifies

---

[6] For the NSS, a principal activity means a status that people spend a relatively longer period during the 365 days (http://mail.mospi.gov.in/index.php/catalog/143/datafile/F5/V209, accessed 22 May 2018).

[7] Minimum Age Convention in 1973, No. 138

that the minimum age for any work of the worst forms of child labour shall not be less than 18 years, which is approved by 179 countries in 2014.

Child labour is regarded as employment harmful to children's health and development, prejudicing their school attendance or participation in training programmes. Later, hazardous work is defined as a sort of worst forms of child labour, which is any work, by its nature and circumstances, harming children's health, safety or morality[8]. The worst forms of child labour include slavery work, prostitution, illicit activities and harmful work to health and safety. Among them, hazardous work is defined as any work, by its nature and circumstances, harming children's health, safety or morality.

The below table is a summarisation of ILO estimation on child labour. The ILO applies a different level of working category such as children aged 5-11 in any work, 12-14 who are in more than light work and children 15-17 who are in hazardous work. Hazardousness is specified by industrial and occupational types, working conditions including long-hour work and hazardous domestic work as well.

Table 1 Estimation of child labour by the ILO criteria

<table>
<tr><td colspan="2" rowspan="2">Category of child work</td><td colspan="3">Age group</td></tr>
<tr><td>5-11</td><td>12-14</td><td>15-17</td></tr>
<tr><td rowspan="4">Hazardousness</td><td>Hazardous industries</td><td colspan="3" rowspan="4">Child labour</td></tr>
<tr><td>Hazardous occupations</td></tr>
<tr><td>Hazardous working conditions
(43 hrs or more, night work, etc.)</td></tr>
<tr><td>Hazardous unpaid household activities
(No methodology)</td></tr>
<tr><td colspan="2">More than light work (14 hours or more)</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Any employment</td><td></td><td>Non-Child</td><td>labour</td></tr>
</table>

Source: Summarised from the ILO conceptual framework (ILO, 2017, p.56)

However, the ILO does not provide specific working hours to estimate child labour in the domestic sector. Hazardous unpaid household activities mean children's involvement in domestic work for long hours, in an unhealthy environment and dangerous locations (ILO, 2016, p.55), but there are no specific criteria for measurement (ILO, 2016, p.57). Many countries have missing counts about child labourers in the domestic sector, so an indirect imputation method is used to estimate them. For example, the ILO measures one country's proportion of child labourers in domestic work using the average of the geographical sub-region (ILO, 2016, p.73).

[8] Worst Forms of Child Labour Convention initiated in 1999, No. 182

2.1.2 UNICEF Definition of Child Labour

UNICEF's definition is similar to the ILO's, but it brings more interest on child's domestic work. UNICEF emphasises the importance of domestic work by children, which is measured by different time bound for ages 5 -11, 12 -14 and 15-17 (Chaubey et al., 2007, p.2). As a result, the number of child labourers in UNICEF's standard shows significant extension than in the ILO's.

UNICEF's time boundaries are not consistent, as there are changes by countries or projects. We use UNICEF's most recent time boundaries of child labour for each age group (Table 2). According to the current database of UNICEF (2017), child labour is defined by : (a) children 5–11 years old who did at least one hour of economic activity or at least 28 hours of household chores per week, (b) children 12–14 years old who did at least 14 hours of economic activity or at least 28 hours of household chores per week, (c) children 15–17 years old who did at least 43 hours of economic activity or household chores per week, and (d) children aged 5–17 years old in hazardous working conditions.

There are some concerns related to a time threshold. Firstly, it does not count the children who work both in economic and domestic work for ages 5 to 14. For example, between ages 12 and 14, working for 10 hours in economic activity and 18 hours in domestic work is regarded as non-child labour despite significant workload. Secondly, 28 hours in domestic work (4 hours each day) seems too high standard. Domestic work more than 21 hours is reported to be harmful for children's education (ILO, 2016).

Table 2 Estimation of child labour by the UNICEF criteria

Time-use	Age group		
	5-11	12-14	15-17
a) Any work in hazardous working conditions			
b) At least 43 hours of economic activity or household chores	Child labour		
c)14 hours or more in economic activity or at least 28 hours in domestic work			
d) At least 1 hour in economic activity or at least 28 hours in domestic work		Non-Child labour	

Source: UNICEF, 2017

2.2 Definition of Child Labour in this Study

Child labour, in this paper, means children between 5 to 17 years of age, who are engaged in any work that is harmful to their development as well as domestic work that requires considerable time. Time thresholds are applied differently for different age groups depending on the types of work. We follow the ILO's minimum ages: age 15 for basic work and age 18 for hazardous work.

Our definition is based on Human Right and Capability Approach, following the definition of international agencies, that includes any types of child labour that hamper children's total development such as physical, intellectual, and mental development (Weiner, 1991; Weston ed., 2005). Child labour is not limited to working as a principal activity but engaging in any types – formal and informal work - of hazardous industries and occupations, so we use both children principal and subsidiary status, which means any work for 30 days or more. A mixed status of working in non-domestic and domestic work should also be considered according to the amount of time spent.

Our definition can be measured by several steps. Firstly, child labourers are screened by the measurement of the ILO that is calculated by three criteria – hazardous industry, hazardous occupations and working for long hours (43 hours or more). Next, the UNICEF time thresholds are applied. We keep time thresholds for the economic activity (43 hours for 15-17, 14 hours for ages 12-14, and 1 hour for ages for 5-11). However, we modify time thresholds to consider children's working in domestic and non-domestic work at the same time. Our changes are for ages 5 to 14, at least 28 hours (4 hours each day) in economic activity or household chores. Still, 28 hours scheme for domestic work is kept, but 21 hours can be considered in the future study. We exclude any cases working for less than 30 days a year from child labour.

Table 3 Estimation of child labour by our definition

<table>
<tr><td colspan="2" rowspan="2">Proposed Sub-category of child work</td><td colspan="3">Age groups</td></tr>
<tr><td>5-11</td><td>12-14</td><td>15-17</td></tr>
<tr><td rowspan="3">Hazardousness</td><td>a) Hazardous industries</td><td colspan="3" rowspan="3"></td></tr>
<tr><td>b) Hazardous occupations</td></tr>
<tr><td>c) Hazardous working conditions (working for 43 hrs or more)</td></tr>
<tr><td colspan="2">d) At least 43 hours of economic activity or household chores</td><td colspan="3">Child labour</td></tr>
<tr><td colspan="2">e) At least 28 hours in economic activity or household chores</td><td colspan="2"></td><td></td></tr>
<tr><td colspan="2">f) 14 hours or more in economic activity or at least 28 hours in domestic work</td><td colspan="2"></td><td></td></tr>
<tr><td colspan="2">g) At least 1 hour in economic activity or at least 28 hours in domestic work</td><td></td><td colspan="2">Non-Child Labour</td></tr>
</table>

Note: The ILO scheme shown in a), b) and c); The UNICEF scheme shown in d), f) and g); Our modifications are in e); excluded working less than 30 days a year

3 Methodology

3.1 Review of Methods

There are several advantages of using the Bayesian approach in measuring posterior probability, or, in this research, the number of child labourers. Firstly, the Bayesian approach quantifies the uncertainty of conditions that may not be observable, taking it as a prior distribution (Gelman et al., 2013, p.8). Secondly, the Bayesian approach provides direct interpretations of the range of posterior probability. While the frequentist approach uses confidence intervals that are the ranges of chances to include the true value of parameters (i.e. 95% confidence), the Bayesian approach uses predictive intervals (PI), which are the ranges of the true values that lie within. The predictive intervals allow us a clear interpretation of the ranges of the numbers of child labourers.

Moreover, the Bayesian method provides an efficient way to combine different datasets. Bayesian statistics transform uncertainty into parameters, making it easy to fit models with many parameters and with multi-layered probabilities (Gelman et al., 2013, p.4). This way becomes more useful when necessary information overspreads in different datasets as we can use multi-parameters from various datasets to fit our Bayesian model to estimate the value of interest. A Poisson model is often used under an assumption of exchangeability, for example, in the epidemiological study (Gelman, 2013, p.45). In this study, we also assume exchangeability between the datasets.

3.2 Data

This research maximises the accuracy of measurements on child labourers in India, combining both the NSS 2011 and the IHDS 2011, and using the Indian Census 2011 as auxiliary information.

The NSS - Employment and Unemployment Survey is a most commonly used dataset regarding employment as it provides details of working types and industrial categories. The NSS 68th round (July 2011-June 2012) is a large sample survey (the number of respondents for ages 5-17: 122,630), covering all the states in India except Andaman and Nicobar Islands. A stratified multi-stage design is applied in the survey; the first stratum is decided by urban- rural relationships and the second stratum is decided by household wealth.

The Indian Human Development Survey (IHDS) is panel data for two rounds – wave 1 in 2004/05 and wave 2 in 2011/12- and we use only the second wave for this study to match the year with the NSS. The sample size of the IHDS is half of the NSS's (the number of respondents for ages 5-17: 51,556). It covers 33 states in India except for Andaman and Nicobar Islands and Lakshadweep. The samples in rural areas were partly drawn from previous participants in the Human Development Profile of India (HDPI) and the samples in urban areas were selected by proportional to population (PPP).

The Indian census 2011 does not include specific data which requires measuring our definition of child labour, such as industries or working hours, but it only offers the number of main workers who work more than 6 months or marginal workers who work for 3 to 6 months. We can use its information only as an auxiliary variable. Also, we can obtain a population size by age and state from the Indian census 2011.

3.3 Matching the two datasets

To combine the datasets, we have critically reviewed the datasets and tried to match the time-use information and types of work (industries and occupations). Regarding time-use, the IHDS asks respondents how many hours a day they usually work, while the NSS asks a daily time disposition of activity, based on a one-week recall. The IHDS provides natural working hours (0 to 24 hrs); however, the NSS offers a categorised intensity of each activity (None-0, half-0.5 and full-1.0) for last seven days (Max points are 7.0 per activity).

This study uses time-thresholds on a weekly basis, so working hours of the IHDS needs to be multiplied by seven days, assuming that children's working hours is consistent for a week. On the other hand, the NSS needs a complicated process to match time-use with the IHD. The NSS sets the max working hours at 7 points in a week, which is converted to 70 points after being multiplied by 10. 70 points in the NSS are regarded as equal to 43 hours in the IHDS. Thus, in the NSS, time thresholds should be multiplied by the ratio of the NSS to the IHDS (e.g. 43 hours *(70/43)=70; 28 hours*(70/43)= 45.58).

The NSS provides 5-digit codes of industries and occupations, corresponding to the NIC (National Industrial Classification) 2008 and the 3-digit codes of NCO (National Classification of Occupation) 2004. However, the IHDS gives broad categories for industries and occupations which do not follow the national standards. Thus, through a careful comparison, we match the IHDS's industrial and occupational codes with the NIC and NCO.

3.4 Model Framework

Under our definition of child labour, we estimate the number of child labour as a mean of the Bayesian hierarchical Poisson log-normal model. We will count the number of child labourers in each state and each age group in the datasets. Then, using our model with grouped data (13 age group * 35 states), we can obtain the new average mean of child labour by age and state.

We use a symbol, μ_{ij}, to represent a key parameter in child labour estimates: the true ratio of children in child labour to all children (i: age, j: states). We use $n1_{ij}$ and $n2_{ij}$ to reflect the number of samples by age and state in the IHDS and the NSS, respectively. Firstly, we estimate μ_{ij} separately for the NSS and the IHDS (Model1 and Model2). Then, we generate a new posterior parameter, μ_{ij}, based on a new distribution constructed by a combination of the IHDS and NSS (Model3). Thus, for each Model, the Poisson distribution's only parameter is the combination μ_{ij} and n_{ij}. Multiplying μ_{ij} by n_{ij} - either $n1_{ij}$ or $n2_{ij}$, we get a suitable parameter for estimating any integer count (Gelman et al., 2013, pp.42-44).

3.5 Prior distributions

We borrow previous knowledge on priors from other literature. Then, we select a prior that is decided by reasonability of posterior distribution (Lynch, 2007, p.72). A Poisson regression model is used in some literature with different priors. In the Poisson regression model, a conjugate prior that results in a posterior distribution of the same distributional family is a gamma distribution. For example, Ntzoufras (2009, p. 245) uses the usual independent normal prior with a large variance, and a small precision at $\tau = \Gamma\ (10^{-4}, 10^{-4})$. Wiśniowski (2016) uses a vaguely informative gamma prior for precision, which assumes $\tau = \Gamma(10^{-6}, 10^{-6})$.

As a prior distribution in this study, we assume a vaguely informative gamma prior for a log-normal distribution of the key parameter, μ_{ij}. The level of the gamma distribution was selected by the simulation study, which is $\tau = \Gamma\ (10^{-3}, 10^{-3})$.

$$\lambda_{ij} \sim \text{normal}(0, \tau),$$

$$\tau \sim \text{gamma}(0.001, 0.001),$$

$$\delta \sim 1/\sqrt{\tau}$$

The priors for the coefficients, α, β, and δ are assumed to be normally distributed with a mean of 0 and a large variance (precision = 10^{-6}).

$$\alpha,\ \beta,\ \delta \sim \text{normal}(0, 10^{-6}),$$

In the combined data model (Model3), we use an over-counting parameter, *oc*, to control over-estimation of child labour in the NSS compared to the IHDS. A systematic over- or under-estimation in one dataset might be solved by applying an over- or under-count parameter (Wiśniowski, 2016). According to a critical analysis of the surveys, especially about time-use data through the IHDS and the NSS, we have found that the NSS shows over-estimation in time-use than the IHDS does. Thus, the over-counting parameter, *oc*, is defined as a uniform distribution between 1 and 100.

$$oc \sim \text{uniform}(1,100)$$

3.6 Model specifications

In our age-state model, i denotes age and j denotes states. $y.ihds_{ij}$ and $y.nss_{ij}$ represent the observed number of child labourers in a group ij (those in one age group in a state) in each survey. $y.ihds_{ij}$ and $y.nss_{ij}$ are described as results of a Poisson distribution of the key parameter, the true ratio of children in child labour to all children (μ_{ij}) multiplied by the number of samples ($n1_{ij}$ from the IHDS or $n2_{ij}$ from the NSS) in a group ij.

$y.pred_{ij}$ is the predicted number of child labourers among the population by age and state. Multiplying the parameter μ_{ij} and the population, N_{ij}, and we can obtain the parameter to predict the number of child labourers by age and state of the whole population, $y.pred_{ij}$. The sum of the result of $y.pred_{ij}$, by age, indicates the predicted number of child labourers by each age ($sum.y_i$).

The log-normal distribution explains the relationship between the parameter, μ_{ij}, the age, x. We use the log-ratio of child labourers to all children (z_{ij}) as an auxiliary variable, which is obtained from the Indian Census 2011 and defined by its broad definition of work (working more than three months). The Indian Census 2011 does not provide specific information to measure our definition of child labour, but it can help to make the results more precise as it covers almost all population. α is the intercept of the model and β and δ are the vector of the coefficients of the covariates x_{ij} and z_{ij}. An over-counting parameter (oc) is applied in Model3 to control over-counting of the NSS compared to the IHDs.

Table 4 summarises the three models

Model 1 (IHDS)	Model 2 (NSS)	Model 3 (Combination)
$y.ihds_j \sim \text{Poisson}(\mu_{ij} * n1_{ij})$	-	$y.ihds_{ij} \sim \text{Poisson}(\mu_{ij} * n1_{ij})$
-	$y.nss_{ij} \sim \text{Poisson}(\mu_{ij} * n2_{ij})$	$y.nss_{ij} \sim \text{Poisson}(oc * \mu_{ij} * n2_{ij})$
$\log(\mu_{ij}) = \alpha + \beta * x_i + \delta * \log(z_{ij}) + \lambda_{ij}$	$\log(\mu_{ij}) = \alpha + \beta * x_i + \delta * \log(z_{ij}) + \lambda_{ij}$	$\log(\mu_{ij}) = \alpha + \beta * x_i + \delta * \log(z_{ij}) + \lambda_{ij}$
$y.pred_{ij} \sim \text{Poisson}(\mu_{ij} * N_{ij})$ $sum.y_i = \Sigma^i\, y.pred_{ij}$	$y.pred_{ij} \sim \text{Poisson}(\mu_{ij} * N_{ij})$ $sum.y_i = \Sigma^i\, y.pred_{ij}$	$y.pred_{ij} \sim \text{Poisson}(\mu_{ij} * N_{ij})$ $sum.y_i = \Sigma^i\, y.pred_{ij}$

Note: i – age; j – state; using grouped data (13 age group * 35 states)

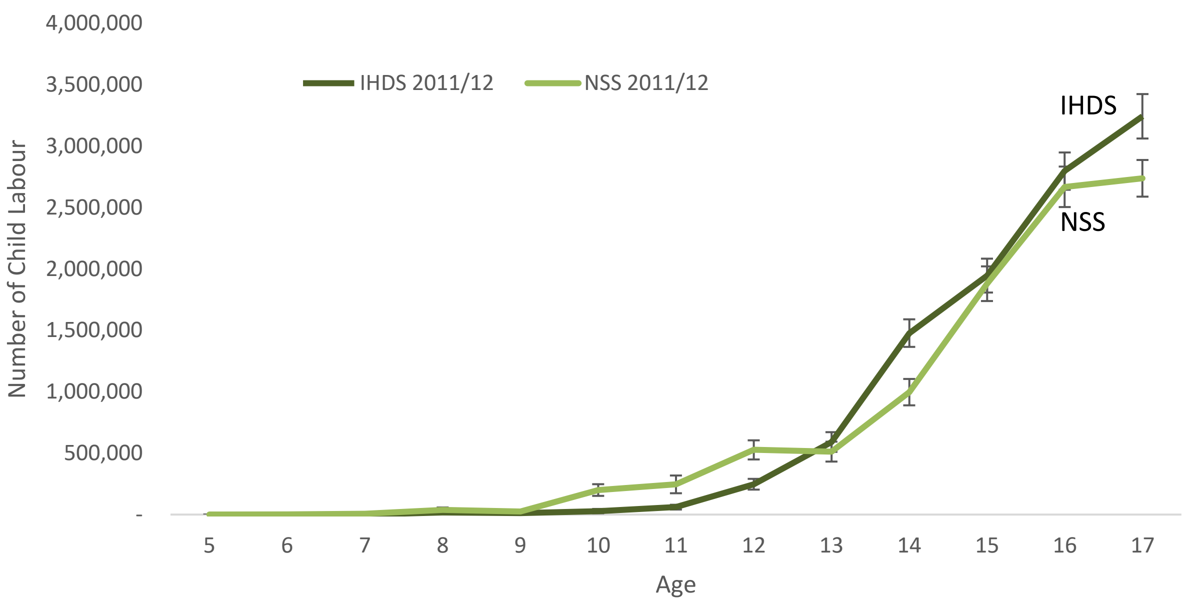
We have run Markov Chain Monte Carlo (MCMC) simulations using R2jags in R. After discarding the first 30,000, we implemented 170,000 iterations and thinned them by 10, producing about 17,000 posterior samples in total.

4 Results

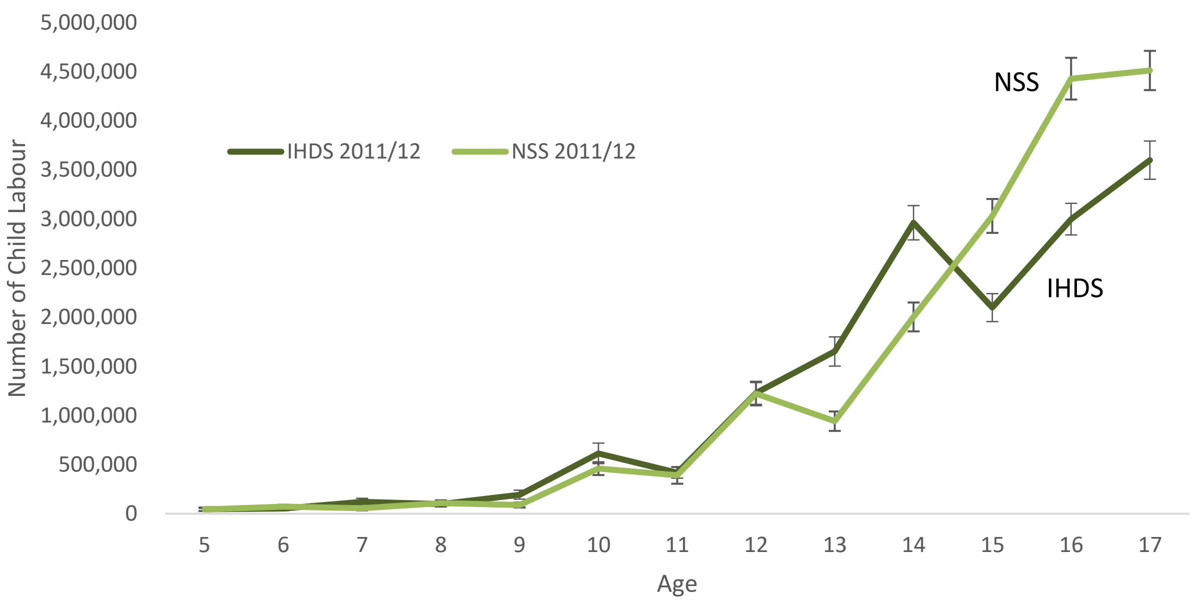
4.1 Findings from Datasets

Figure 1 Weighted Number of Child Labourers Using Different Definitions

a. ILO definition



b. UNICEF definition



Source: NSS 2011/12; IHDS 2011/12

Note: Error bars represent ± 1.96 standard error; Survey weight applied for the purpose of descriptive data analysis

Figure 1 provides a weighted count of children who are deemed as labourers according to two different definitions - the ILO, UNICEF definition - using the NSS and the IHDS, in 2011/12. When applying the ILO standard, the total figure of child labourers is counted as 10 million for the IHDS while the same figure represents 9.8 million for the NSS. The UNICEF standard for those aged 5 to 17 provides higher figures; 16.1 million and 17.3 million respectively, for the IHDS and the NSS.

The patterns of this simple counting results provide a general glance of child labour. As the UNICEF definition adds domestic work for a long hour to the category of child labour, it gives a higher figure than the result of the ILO definition does. The weighted number of child labourers shows an irregular pattern by age, especially for the UNICEF standard. It is the result affected by different population per each age, for example, the population at age 11 and 13 are critically small compared with any other age groups (See appendix 2). Nevertheless, according to our knowledge from the Indian census, there should not be such a decrease in the number of child labourers between ages. The age pattern informs us that multiplying by a sampling weight might not be a suitable solution for this case.

Figure 2 Weighted Number of Child Labourers Using a New Definition

Note: Error bars represent ± 1.96 standard error; Survey weight applied for the purpose of descriptive data analysis

Figure 2 shows the result of the weighted calculation of child labourers and the linearised standard errors using NSS and IHDS 2011/12 when using our definition. The NSS provides the weighted figure for child labour, 18.6 million, and the IHDS offers 15.2 million. The standard error is too large to indicate precise figures for child labour. Another problem with applying sampling weight in our case is that weights do not reflect age proportions. In both datasets, weights are correspondent to the census proportions and are adjusted by urban-rural proportion, not by age proportion. Thus, when we compare the weighted number of child labourers by age, it might misrepresent the true figure. Accordingly, we prefer not to use weighted counting, but to get modelled results for the rate of child labour and to multiply it with the population from the Indian census.

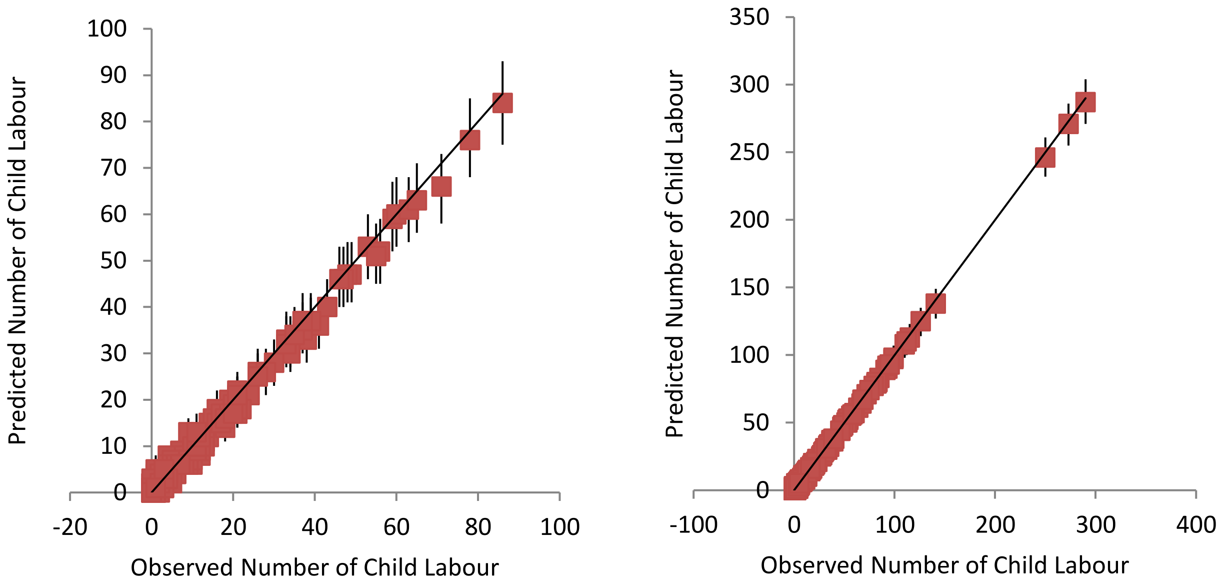
A descriptive analysis informs us that there could be a systematic over-counting of child labourers in NSS comparing the result in the IHDS, especially among ages 15, 16 and 17. The scales of the time-variable of the two surveys are different, which might make a difference in a calculation of child labour. Thus, in our models, we use a parameter to measure a systemic over-counting of the NSS comparing with the IHDS.

4.2 Goodness of fit of the Models

Figure 3 Comparing Observation and Prediction in the Number of Child Labourers

a. IHDS

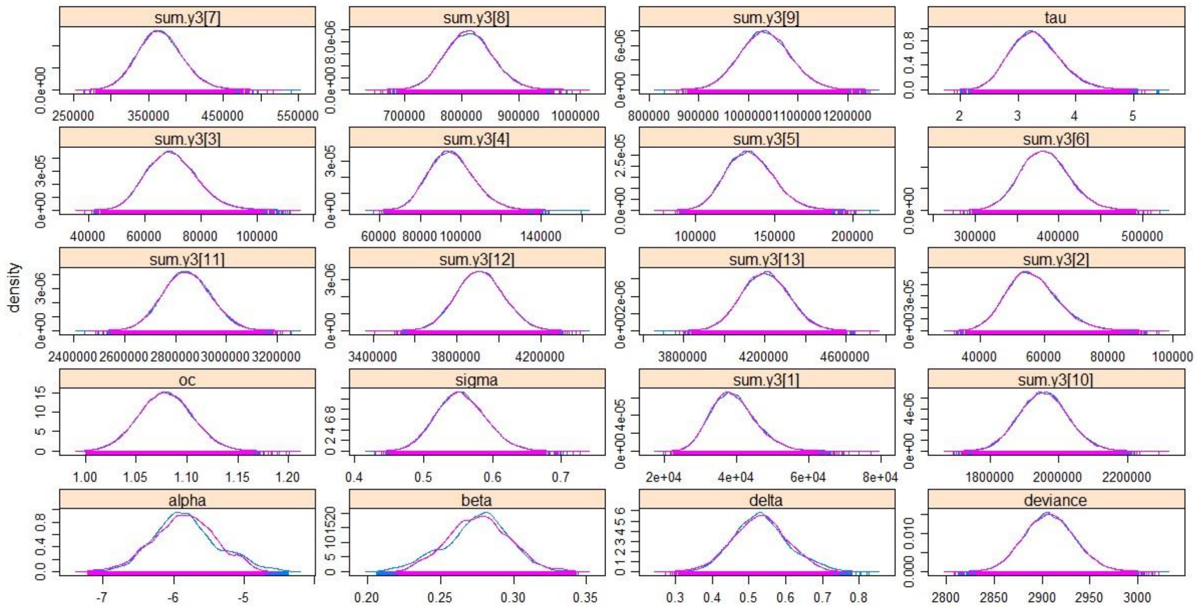
b. NSS



Note: Median and 50% interquartile ranges (25%~75%); N=455 (aggregate number by age and state)

We have reviewed the differences between the observed number of child labour and the predicted number of child labour through Model 1and Model 2. The interquartile ranges between 25% and 75% of posterior parameters, includes the 45-degree line that indicates a perfect prediction, in both IHDS and NSS. The result indicates our models replicate the posteriors values without any systemic errors. As we do not have observations for the combined-data model, the comparison of observation and prediction is not possible for the Model 3.

Figure 4 Density Plot of the Posterior Number of Child Labourers by Age



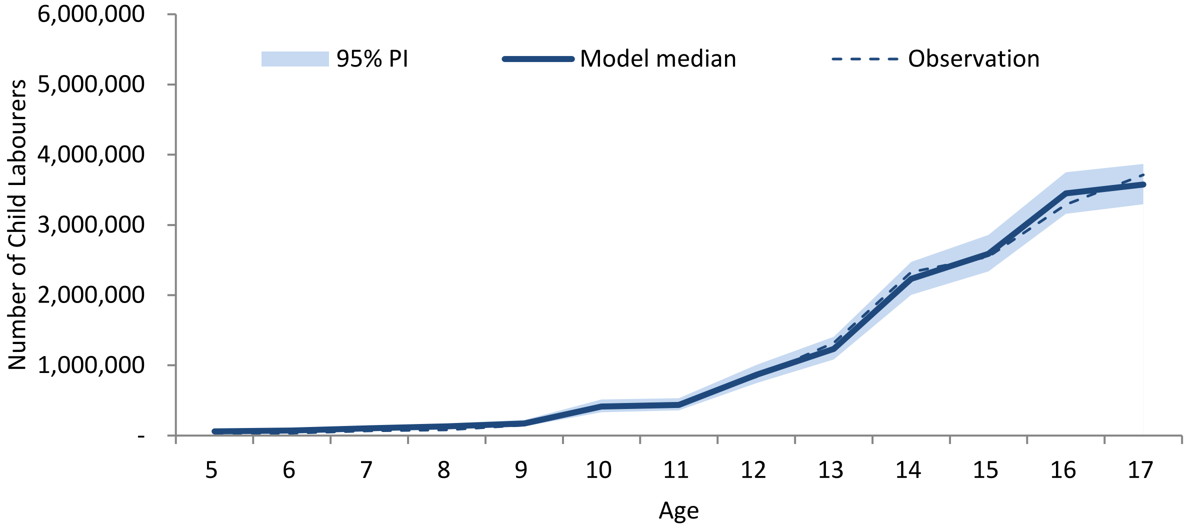
Note: Model3; Burnin-30,000; iterations kept-17,000; thin by 10

The MCMC algorithm shows proper convergence in our models. Sufficient burn-in over 30,000 has been made to avoid any in-convergence problem. The density plot of model 3 indicates that the posterior parameters of interests are based on stable chains.

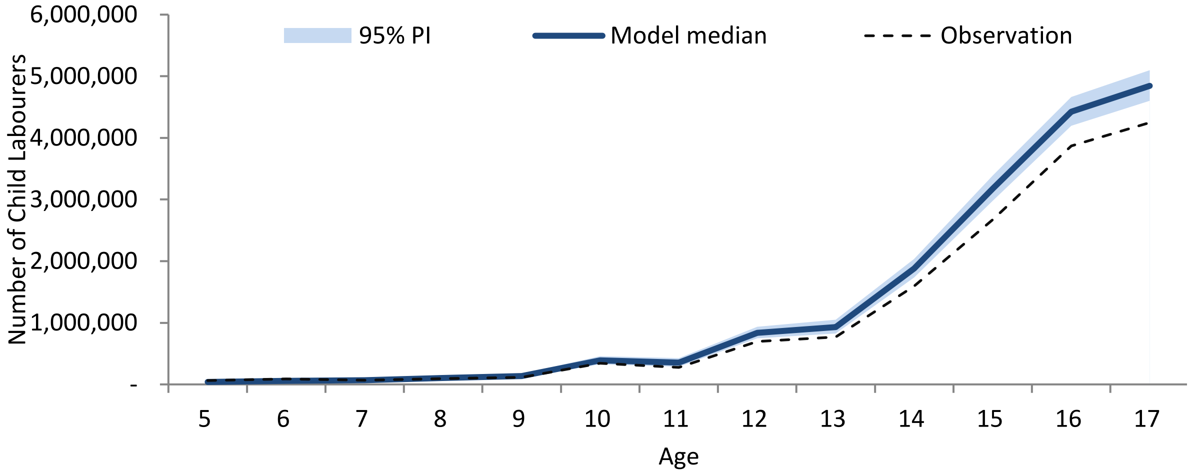
4.3 A Bayesian Hierarchical Poisson Regression Model for a Single Dataset

Figure 5 Posterior Results of Bayesian Poisson Age-State Model Using IHDS 2011-12 and NSS 2011-12 (Model 1 and Model 2)

a. IHDS 2011/12



b. NSS 2011/12



Notes: Observed figures calculated by (No. of child labourers/No of children*Population by age)

Both Model 1 (with the IHDS 2011/12) and Model 2 (with the NSS 2011/12) show a sharp increase of child labour as age increases. However, a critical difference exists in the estimation of the total figures of child labour between them. The IHDS 2011 estimates the number of child labourers at 15.3 million, and the 95% prediction interval (PI) ranges are from 14.7 million to 16 million, while the NSS 2011 provides measures at 17.3 million, with the range 16.8 million to 17.7 million.

The model better fits with the IHDS than with the NSS. The 95% PI of the IHDS suitably includes the observed number of child labour (the observed rate of child labour multiplied by the population), while the 95% PI of the NSS does not. This represents that the observation of child labour through the NSS is more erratic than the IHDS. As age increases, the difference between the posterior results and the observation extends in the NSS. The overcounts of Model 2 is apparent for children who are aged 15, 16 and 17.

Comparing the result with the result from a simple weighted counting, we recognise a large reduction in standard errors in our Bayesian hierarchical Poisson model. Also, the model smooths the difference between ages comparing the weighted number of child labourers. The result shows that the sudden decrease in the ratio of child labourers between ages is almost removed in a model. There is only a slight reduction between ages 10 and 11 in the NSS, which is related to the small size of the population in the age 11.

There are still critical problems that should be addressed regarding counting child labourers through a model – firstly, a comparably large measurement error of the IHDS, which is caused by relatively small sample size. The 95% PI is wider in the IHDS than the results in the NSS. The number of child labour suggested by the IHDS is still quite broadly ranged from 14.7 million to 16 million.

Secondly, in the NSS, overestimation of the number of child labourers is found, especially among ages 15,16, and 17. The reason for this inflation can be traced to the time-use scale. Average time-use in both domestic and non-domestic work is much higher in the NSS than in the IHDS. The number of child labourers, especially among ages 15, 16 and 17, seems inflated in the NSS. This problem could be corrected by using an over-counting parameter in the next model.

4.4 A Bayesian Hierarchical Poisson Regression Model for Combination of Two Datasets

Figure 6 Results of Bayesian Poisson Age-State Model Using a Combination of Datasets (Model 3)

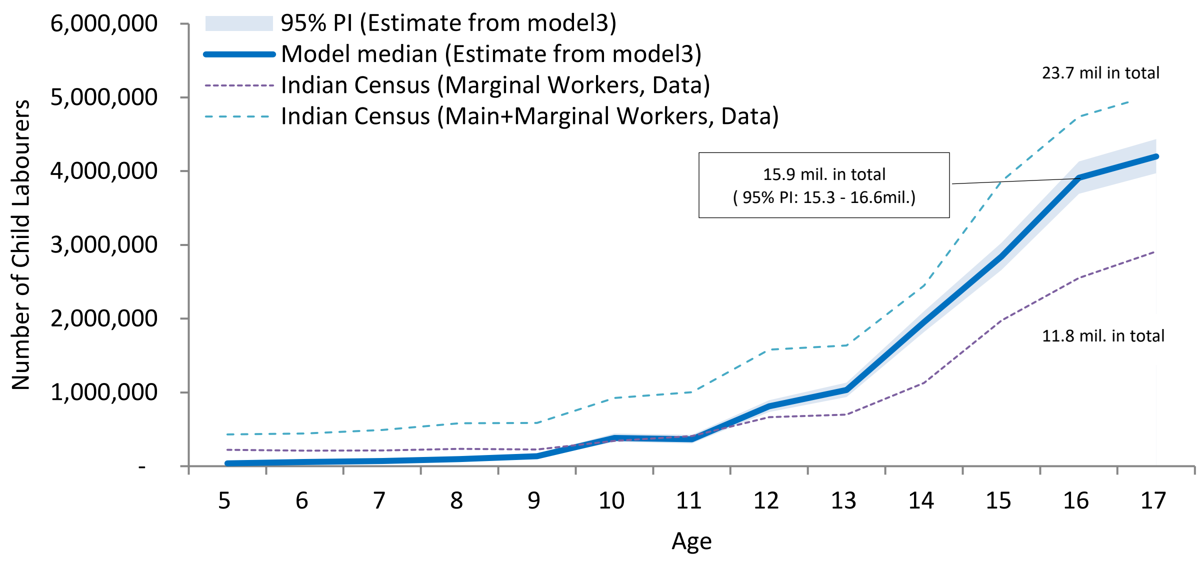


Table 5 Aggregate Child Labour Estimation

Age groups	Age 5-14		Age 5-17	
Models	No. of Child Labourers 1) (95% PI)	%2)	No. of Child Labourers3) (95% PI)	%4)
IHDS 2011/12	5,734,638 (5,350,407-6,143,590)	2.2	15,348,199 (14,705,441-16,009,437)	4.6
NSS 2011/12	4,800,984 (4,557,793-5,054,943)	1.8	17,256,310 (16,783,535-17,737,416)	5.2
Model 3	4,974,050 (4,713,925-5,242,877)	1.9	15,906,458 (15,267,842 – 16,554,824)	4.8

Notes: See Appendix-Table1; 1), 3) Median; 2), 4) % of the population from the Indian Census 2011/12

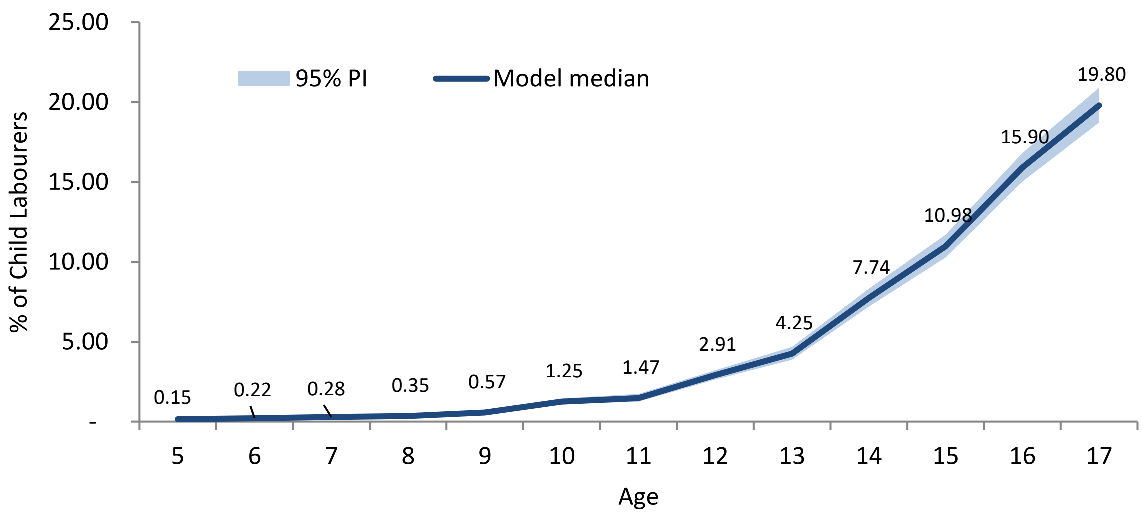
According to the results of our final model, the number of child labourers (ages 5-17) is estimated at 15.9 million in 2011. The 95% prediction interval indicates that a range of the number of child labourers is from 15.3 million to 16.6 million. The figure for ages 5-14 is estimated at around 5 million and ranged from 4.7 million to 5.2 million.

Our over-counting parameter indicates the mean value of 1.08 (sd. 0.026), which explains that the NSS has a systemic over-counting in the child labour than the IHDS. After control of the over-counting of the NSS, the combination model provides a smaller number of child labour in ages between 5 to 17, which tells that the NSS has overall inflation in the figures. The number of child labourers was estimated at 17.2 million for the NSS only model, and it adjusted to 15.9 million in the combined-data model.

The number of child labourers is estimated to be higher than the figure proposed by the Indian census (Figure 5). The number of child labourers suggested by the Indian Census is 11.8 million among ages 5-17, which is much smaller than our estimation. The figure for the main and marginal workers of ages 5-17 is around 23.7 million, but this is not the number of child labourers but more likely child workers.

Our final model does not adequately capture the child labourers under age ten; although, the Indian census informs of the existence of a large number of young-aged child workers and child labourers. It is because our models considerably rely on the datasets, and many children in early childhood are categorised as either ‘other’ or ‘too young’ in a survey. We believe children who are under age ten are underestimated in a close relationship with ‘nowhere children’, but we do not cover them in this research.

Figure 7 Posterior % of Child Labour by Age using the Model 3



A large extension in the number of child labourers appears at age 12 and age 14 when much more children are likely to be involved in labour compared to the ages before that. The age pattern is similar in the proportion of child labourer to the population of each age (Figure 5). Accordingly, we can assume more children begin their labour, or change from child worker to child labourer, when they turn 12 and 14, which might be related with the education system in India.

5 Conclusion and Policy Implications

Given our definition of child labour using hazardousness concept of the ILO and a time threshold of UNICEF, the probability distribution represented an upper bound (4.7 ~ 5.2 million for ages 5 to 14, 15.3 ~ 16.6 million for ages 5 to 17) of child labour compared to other available definitions of child labour. The estimated number of child labour in India using the combined data approach, indicates higher values than the number of child labour announced by the Ministry of Labour and Employment, the Government of India, 4.35 million for ages 5 to 14[9], which is calculated by the Indian census 2011, only.

A Bayesian combined data approach can overcome the limitation of the use of a single dataset, providing a precise estimate. A Bayesian hierarchical model provides an efficient way to incorporate uncertainty raised by a small number of observations of child labour as well as potential under- or over-count of child labourers, as it is measured by multi-dimensional indicators. The posterior probability distribution allows precise estimation as it maximises the use of information using different datasets. Also, in our case, a prediction by age is smoothed by borrowing knowledge from the Census data.

Our results recognise that child labour in the domestic sector is a non-ignorable part of child labour in India. The Indian Government's Amendment Act, which came into effect from September 2016, excludes "helping families or working in family enterprises" from a category of child labour. Considering the large number of child labourers engaged in domestic work, the next step should be made for the inclusion of domestic work for long hours in the definition of child labour as well as institutional support for them.

The probability distribution of our age-state model shows a clear age trend when children decide to be labourers. It is found that children might decide before they enter secondary school at age 12 or after completing primary school at age 14. This result reinforces the importance of secondary education in demotivating children from becoming full-time workers (Chaudhri et al. 2003; Chaudhri and Wilson 2000). Secondary education is not yet compulsory, although the law (the Right of Children to Free and Compulsory Education Act 2009) defines education as free for children 6 to 14 years of age or up to class 8, which is at age 12. Although further investigation about the relationship between education and child labour is needed, our findings support that robust political interventions are demanded around those ages.

The outcome of the models lacks in capturing child labour in early age groups, such as age 10, as it largely relies on the observation from the datasets. We could improve the prediction using more

[9] Ministry of Labour and Employment, web address: labour.gov.in/childlabour/census-data-child-labour, accessed on 08/06/2018

informative priors in the next study. Also, ‘nowhere children’, who are neither in school nor labour, can be considered as a separated category in further research, as a way to seize child labour in early childhood.

Reference

Basu, K. (1999). Child labor: cause, consequence, and cure, with remarks on international labor standards. Journal of economic literature, xxxvii(3), pp.1083–1119.

Basu, K. and Van, P.H. (1998). The Economics of Child Labor. The American Economic Review, 88(3), pp.412–427.

Chaudhri, D. et al. (2003). Determinants of child labour in Indian states: Some empirical explorations (1961-1991). Journal of Quantitative Economics, 1(1), p.1.

Chaudhri, D. P. and Wilson, E. J. (2000). The Challenge of Child Labour in Rural India: A Multi-Dimensional Problem in Need of an Orchestrated Policy Response, Department of Economics, University of Wollongong.

Chaubey, J. et al. (2007). Child Labour, Education and Policy Options. UNICEF: New York.

Das, S. and Mukherjee, D. (2007). Role of women in schooling and child labour decision: The case of urban boys in India. Social Indicators Research, 82(3), pp.463–486.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. Boca Raton, FL: CRC press.

Giri, A.K. and Singh, S. (2016). Child Labour and ‘Nowhere’Children in Post-reforms India. Indian Journal of Human Development, 10(1), pp.97–110.

Grootaert, C. and Kanbur, R. (1995). Child labour: an economic perspective. International labour review, 134(2), pp.187–204.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). Survey methodology (Vol. 561). John Wiley & Sons.

Kambhampati, U.S. and Rajan, R. (2008). The `Nowhere’ Children: Patriarchy and the Role of Girls in India’s Rural Economy. Journal of Development Studies, 44(9), pp.1309–1341.

Lynch, S.M. editor. (2007). Introduction to Applied Bayesian Statistics and Estimation for Social Scientists. New York, NY : Springer New York,.

Ntzoufras, I. (2009). Bayesian modeling using WinBUGS (Wiley series in computational statistics). Hoboken: John Wiley & Sons.

ILO. (2017). Global estimate of child labour: Result and trends, 2012-2016. International Labour Office: Geneva.

ILO. (2016). Methodology: Global estimate of child labour, 2012-2016. International Labour Office: Geneva.

Omoike, E. (2010). Child domestic labour: fostering in transition? Child Slavery Now: A Contemporary Reader, p.203.

Ray, R. (2000). Child labor, child schooling, and their interaction with adult labor: Empirical evidence for Peru and Pakistan. World Bank Economic Review, 14(2), pp.347–367.

UNICEF (2013). [online] India Statistics. Available at: https://www.unicef.org/infobycountry/india_statistics.html [Accessed 22 May 2018].

UNICEF (2017). [online] Access the data - Child Labour. Available at: https://data.unicef.org/wp-content/uploads/2015/12/Child-labour-database_Nov-2017.xls [Accessed 22 May 2018].

Weston, B. (2005). Child labor and human rights : Making children matter. Boulder, Colo.: Lynne Rienner.

Wiśniowski, A. (2017). Combining Labour Force Survey data to estimate migration flows: the case of migration from Poland to the UK. Journal of the Royal Statistical Society: Series A (Statistics in Society), 180(1), pp.185–202.

Wisniowski, A. et al. (2016). Integrated modelling of age and sex patterns of European migration. Journal of the Royal Statistical Society: Series A (Statistics in Society), 179(4), pp.1007–1024.

Affiliations

Jihye Kim[1)] · Wendy Olsen[2)] · Arkadiusz Wiśniowski[3)]

1) Jihye Kim, jihye.kim@postgrad.manchester.ac.uk

School of Social Science, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

2) Wendy Olsen, wendy.olsen@manchester.ac.uk

School of Social Science, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

3) Arkadiusz Wiśniowski, a.wisniowski@manchester.ac.uk

School of Social Science, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

Appendix

Figure 1 The Framework of a Bayesian Data Combination

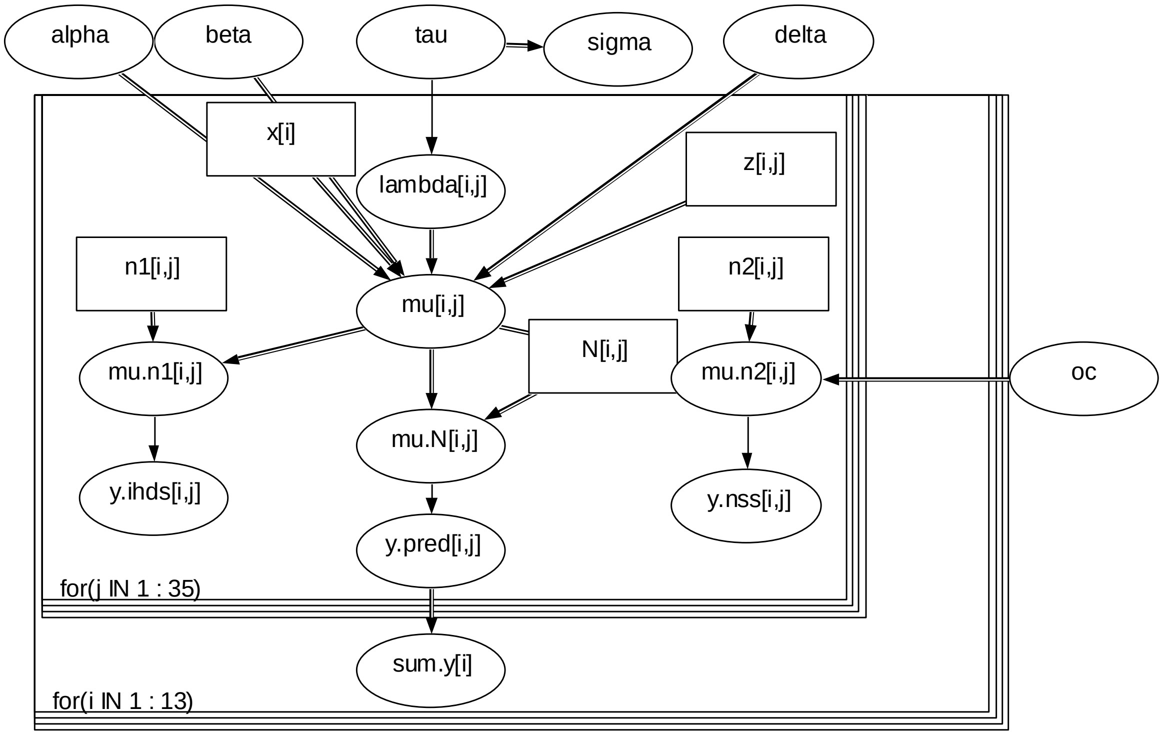
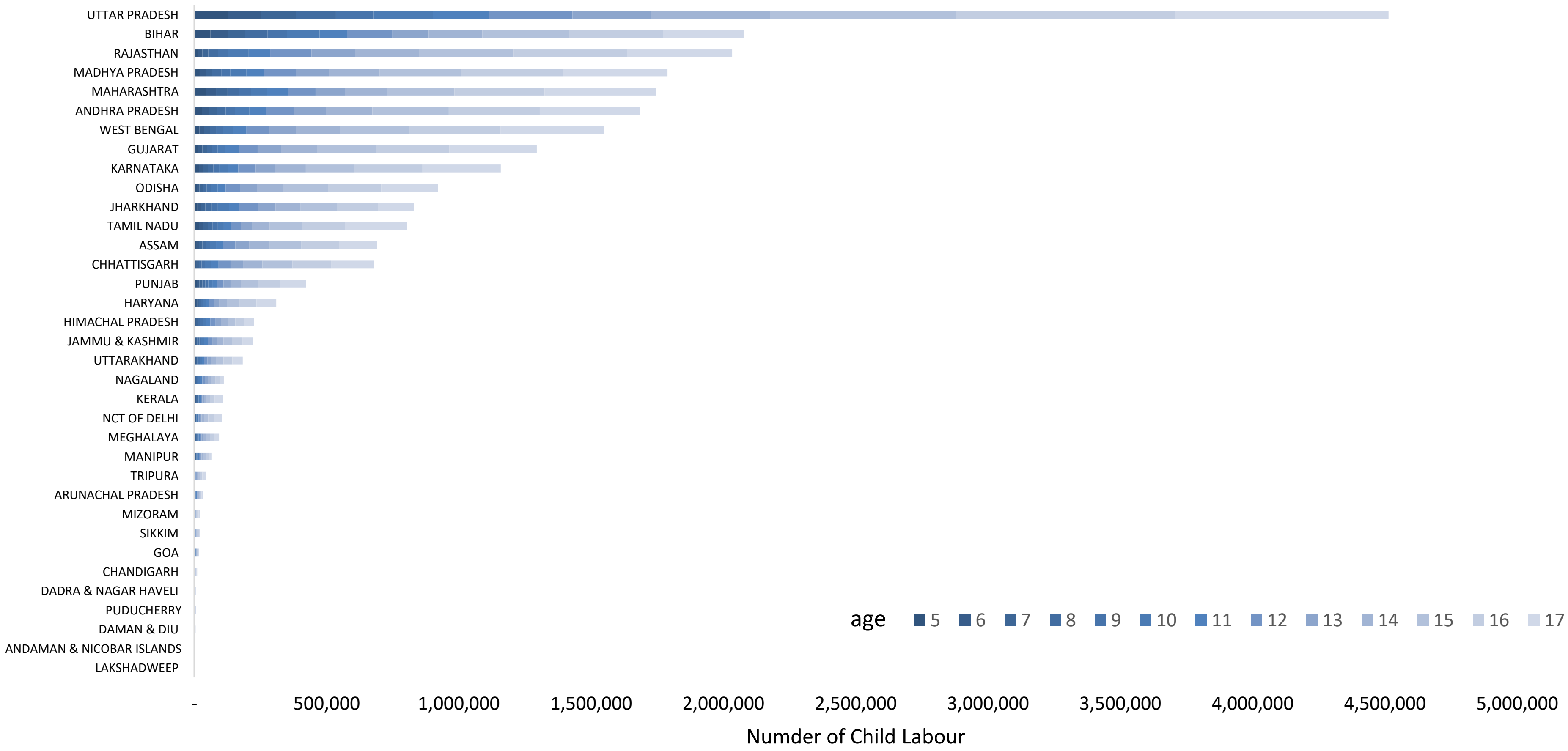


Figure 2 The number of child labour by age and state using the Indian Census 2011



Note: Main workers (working more than 6months) + marginal workers (working 3 – 6 months)

Source: Indian Census, 2011(available on www.censusindia.gov.in/2011census/population_enumeration.html, data file : C-13 Appendix Single Year Age Returns by Residence, Sex and Literacy Status (India & States/UTs), DDW-C13APPENDIXB—0000.xlsx, accessed on 08/06/2018)

Table 1 Summary of Child Labour Data in the IHDS, the NSS and the Indian Census

	IHDS[1)]			NSS[2)]			Indian Census		
Age	No. of Child Labourers (Y_{ihds})	No. of Children in Sample (n1)	Weighted No. of Child Labourer[3)]	No. of Child Labourers (Y_{nss})	No. of Children in Sample (n2)	Weighted No. of Child Labourer[4)]	No. of Child Labourers with narrow definition[5)]	No. of Child Labourers with broad definition[6)]	Population
5	5	3,785	38,748	21	8,736	43,069	223,354	430,785	26,048,171
6	5	3,760	29,385	31	9,190	70,479	211,068	442,565	25,647,854
7	12	4,312	106,699	24	8,658	51,049	214,041	491,150	24,820,355
8	11	3,605	64,745	37	10,416	100,603	234,439	583,419	26,961,440
9	22	3,406	138,090	35	7,226	88,526	225,906	585,719	23,418,444
10	59	4,346	471,300	131	11,717	474,522	344,651	925,032	30,544,351
11	62	3,345	358,629	84	7,478	396,039	409,365	1,003,678	24,733,883
12	154	5,133	1,052,417	296	11,812	1,254,192	663,856	1,579,741	27,869,538
13	236	4,110	1,461,701	267	8,457	947,902	699,458	1,636,946	24,273,967
14	415	4,537	2,632,954	633	10,041	2,025,760	1,127,109	2,449,628	25,250,481
15	359	3,657	2,167,182	1,009	9,778	3,342,412	1,975,126	3,865,154	25,891,864
16	518	3,829	3,102,188	1,532	9,738	4,858,874	2,552,054	4,738,080	24,584,341
17	614	3,731	3,642,489	1,604	8,011	4,944,608	2,911,827	5,047,586	21,210,681
Total	2,472	51,556	16,131,877	5,704	121,258	18,642,214	11,792,254	23,779,483	331,255,370

Note: 1), 2) Our definition of child labour is applied; 3), 4) Sampling weight for the household; 5) including main workers; 6) including main and marginal workers

Table 2 Summary of the Result of a Bayesian Hierarchical Poisson Model

	Model1 (IHDS)			Model2 (NSS)			Model3 (Combination of IHDS and NSS)		
Age	No. of Child Labourers[1)]	95% PI		No. of Child Labourers[1)]	95% PI		No. of Child Labourers[1)]	95% PI	
		HL[2)]	LL[3)]		HL[2)]	LL[3)]		HL[2)]	LL[3)]
5	59,338	85,099	41,006	39,687	56,556	27,795	38,658	52,604	28,328
6	73,110	102,150	51,940	59,823	82,168	43,528	55,617	73,334	41,916
7	104,882	141,920	76,843	66,296	88,442	49,426	68,888	88,831	53,412
8	129,016	170,330	96,544	102,062	131,792	78,551	94,747	118,430	75,337
9	170,060	220,815	130,535	134,547	173,965	103,808	133,436	166,068	107,030
10	411,869	513,652	330,964	388,107	458,694	326,385	382,152	441,706	329,561
11	437,473	533,264	357,844	351,045	424,677	288,117	364,472	426,852	310,049
12	866,672	1,007,516	745,360	837,488	937,073	746,028	811,437	894,317	734,478
13	1,234,579	1,408,704	1,083,176	932,381	1,048,952	827,296	1,032,732	1,134,049	938,876
14	2,230,443	2,474,560	2,006,033	1,881,518	2,038,050	1,734,090	1,953,820	2,099,855	1,818,617
15	2,585,414	2,855,638	2,337,707	3,179,919	3,391,393	2,979,508	2,842,799	3,029,663	2,660,716
16	3,447,667	3,751,588	3,159,285	4,426,425	4,662,358	4,197,419	3,910,144	4,133,208	3,692,803
17	3,574,853	3,868,104	3,294,026	4,844,503	5,098,871	4,602,384	4,200,088	4,435,323	3,972,953
5-14	5,734,638	5,350,407	6,143,590	4,800,984	4,557,793-	5,054,943	4,974,050	4,713,925	5,242,877
5-17	15,348,199	14,705,441	16,009,437	17,256,310	16,783,535	17,737,416	15,906,458	15,267,842	16,554,824

Note:1) median; 2) HL refers to high level (97.5%); 3), LL refers to low level (2.5%)