



University of
Nottingham
Rights Lab

Beyond the walls:

Microdata on domestic workers in
North East India

2020



Contents

Preface	4
1. Introduction to the data analysis	5
1.1 Research methods	5
1.2 Key findings from the data analysis	8
2. Comprehensive data analysis	10
2.1 Descriptive statistics	10
2.2 Correlation analysis	18
2.3 Factor analysis	22
2.4 Spatial variation and state comparisons	22
3. The surveys in context	28
3.1 A Catholic tradition of human rights data-gathering	28
3.2 Datasets on domestic workers	29
3.3 Data-gathering innovations for modern slavery	29
3.4 Data-gathering by front-line organisations	33
4. Recommendations	37
4.1 Suggestions for future research	37
4.2 Policy recommendations	42
References	44
Appendix: Blank copy of survey	46

The nearly **12,000 surveys** in this new Domestic Workers Dataset offer a unique opportunity to understand people at high risk of labour exploitation whose lives are hidden from view.

Preface

Between 2015 and 2019, a network of Missionary Sisters of Mary Help of Christians (MSMHC) surveyed nearly 12,000 domestic workers in North East India. The surveying work was led by Centre for Development Initiatives (CDI) Director Sister Rose Paite. CDI is a frontline service provider to at-risk populations and a registered non-profit organisation. Based in Guwahati, the largest city in Assam (one of the states where the surveys were conducted), it works across all North East states of India. The surveys were conducted across six states in North East India in order to generate baseline data for the impact assessment of a CDI project supported by an NGO called Don Bosco Mondo (DBM). The project is organising domestic workers into union structures.

The survey respondents were approximately 12,000 of at least 4 million domestic workers in India. The country's official National Sample Survey estimates there are 4.2 million domestic workers in India, while the International Labour Organisation (ILO) notes that the real figure is likely higher (2013a). Globally, the ILO estimates there are at least 67 million domestic workers over the age of 15 worldwide, 80% of whom are women (2013a). The ILO also estimates the number of child domestic workers who are between 5-14 years old at 11.2 million globally (ILO, 2013b), and India's census puts the country's figure for child domestic workers at 0.52 million (ILO, 2017).

People in domestic labour are one of the most vulnerable categories of workers. In India they are part of an unregulated, informal sector, hidden from view in private homes. They have no legal protection as workers under India's labour laws, which do not recognise domestic work as work, and limited social protections. Many are from the most socially discriminated populations, are part of the "Other Backward Classes" category, and have migrated to cities from poor rural areas. Common working conditions include long and unregulated work hours, confinement, physical violence, sexual assault, under payment or no payment.

Many are victims of human trafficking, recruited to cities through offers of work. When domestic workers migrate to cities, they are victimised by labour agents who charge placement and travel fees that place workers in situations that meet the international definitions of trafficking and bonded labour. Globally, of the 16 million people identified by the ILO as being in forced labour (in the private sector rather than state-sponsored), by far the largest share at 24% were domestic workers (ILO, 2017).

The nearly 12,000 surveys in this new Domestic Workers Dataset offer a unique opportunity to understand people at high risk of labour exploitation whose lives are hidden from view—whose work was previously unobservable in such a detailed, large-scale way. In 2019-20, the Rights Lab at the University of Nottingham conducted an analysis on these microdata (information at the level of individual respondents). The dataset opens a window into not only these workers' lives and circumstances, but also a pressing global problem of exploited labour; one that sits at the intersection of sustainable development, human rights, labour rights, and criminal justice.

This report introduces the major findings of the analysis, and presents and discusses the data in detail. It places the Domestic Workers Dataset in the contexts of social justice data-gathering, data innovations for the anti-slavery field, the data-gathering initiatives of front-line organisations, and datasets on domestic work, and it makes suggestions for policy responses and for follow-on research in this area.

The report was prepared by the Rights Lab (University of Nottingham) for the Adivasi Students Association of Assam with the support of the Arise Foundation, DBM and CDI. The report was authored with contributions by members of the Rights Lab research team: James Goulding, Catherine Waite, Zoe Trodd, Doreen Boyd, Todd Landman, Emily Wyman, Laoise Ní Bhriain, and Caroline Emberson.

Section one

Introduction to the data analysis

1.1 Research methods

The surveys analysed in this report were undertaken as baseline data collection for the impact assessment of a CDI-led project to create union structures for domestic workers in six states of North East India. Beginning in 2015, CDI created pre-union groups that connect through the Ferrando Domestic Workers Alliance (FDWA). It aims to organise 30,000 domestic workers via these local groups and the FDWA across 12 cities: Agartala, Aizawl, Barpeta, Bongaigaon, Guwahati, Imphal, Kohima, Sarupathar, Shillong, Tezpur, Tinsukia, Tura.

By 2018, CDI had identified 18,531 domestic workers in the 12 cities, and had registered 13,668 with the FDWA into 534 groups. By 2019, over 18,000 workers were registered in more than 600 groups. Registration with the FDWA involves a payment by each worker of Rs10. The FDWA maintains a register of individuals, and each local group maintains a group register. As part of the CDI project, the registered individuals receive capacity-building support, rights training, and a platform for advocacy and campaigning.

The individual survey respondents are domestic workers who had registered with the FDWA as part of the CDI project. Members of CDI's network of Sisters conducted interviews with the domestic workers between 2015 and 2019 in the project's 12 cities. The interviews took place off-site from respondents' work-places and in a one-to-one setting. The interviewers completed the surveys by hand. A blank copy of the survey is included as an appendix to this report.

As the survey respondents were not randomly selected, the data cannot be considered representative of the general population of domestic workers in North East India. For example the average age of respondents is 36 (standard deviation [SD] 12), the average age when starting work is 26 (SD 10), and the mean number of children that each domestic worker has is 1 (SD 1). This seems unlikely to be representative of the wider population of domestic workers and may reflect the demographics of the respondents (those who chose to participate in the CDI empowerment programme). Similarly, the nature of the respondent pool has implications for the findings on child labour. The CDI project aims to empower domestic workers through creating union structures and over 99% of the individuals who chose to participate in the programme and register with the FDWA (and were therefore surveyed) were age 14 or older. The figures for child labour in the dataset are therefore very likely to underestimate the rate of child labour in North East India's domestic work sector.

The data and this report's findings also cannot be generalised to the national level. The CDI empowerment programme is designed for six states in North East India, a region with unique socio-economic features. However, in terms of the geospatial patterns we identified across the six states themselves, particular concentrations are not simply a function of where the CDI staff members were based. The interviewers travelled widely to achieve geographical coverage.

After receiving and cleaning the data, including duplicate identification, we identified a total of 11,759 respondents who were surveyed throughout the six states of North East India (Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura). The surveys cover demographic characteristics, cultural factors, education status, and work background. They contain information on both work history and current work undertaken, including terms of current employment and working conditions. The data is set in a geospatial context, with worker and employer address recorded. This enabled us to observe patterns within and between the six Indian States.

There is missing data in the full dataset collected from the surveys. The missing data are most often found for variables providing space for additional (i) responses or reasons related to the previous question, (ii) shift times, and (iii) family members, which many respondents did not use. There also are missing data points for answers that depend on a positive response to a previous answer. For example, for any response to be relevant to the question: “reason for salary deduction?” the respondent must have answered yes to the question: “do you have any salary deductions?”

The data analysis stages can be summarised as:

Stage 1	Data cleansing, including duplicate identification, anomaly detection and text coding.
Stage 2	General statistical analysis organised by category—socio-demographics (including family demographics), occupational information, employer, terms of employment, working conditions and social security.
Stage 3	Analysis of variable inter-relationships (e.g. correlation analysis) including topic modelling, to establish underlying and explanatory themes within the data.
Stage 4	Geospatial analysis of the dataset to show how key variables are distributed.
Stage 5	Examination of how the geospatial patterns and themes established in Stage 4 are distributed, with a focus on child labour.
Stage 6	Establishment of target ‘concept’ variables (likely a proxy for “quality of working/employment situation”) and refinement of hypotheses (age, salary and working hours were variables of interest).

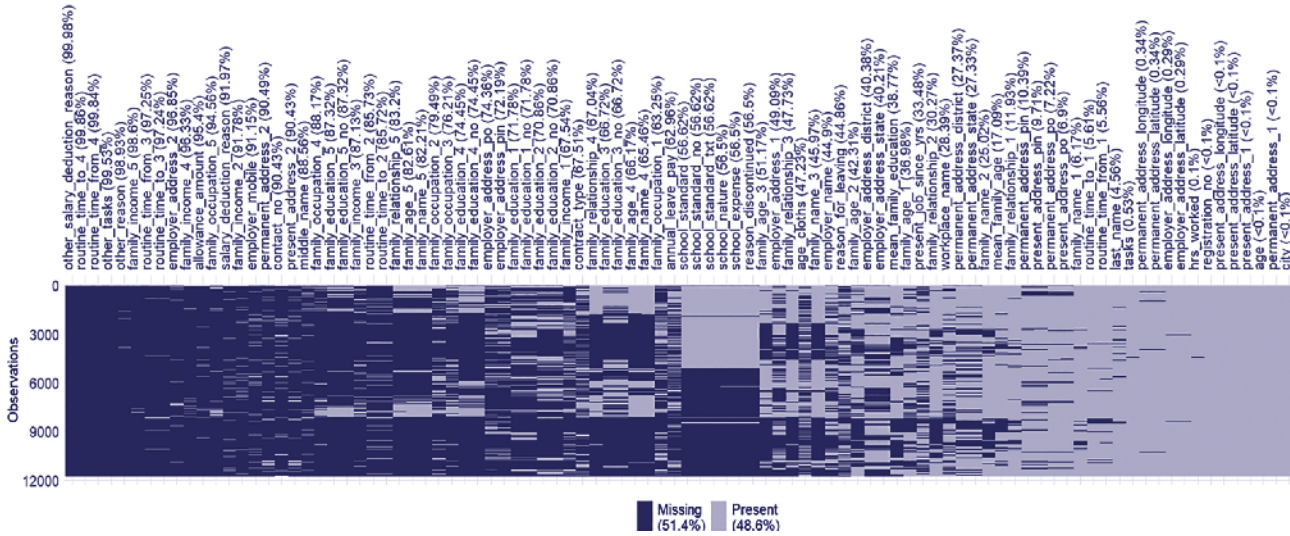


Figure 1. The variables (n = 86) which have at least one missing data point. Those variables with no missing data (n = 30) have been excluded. Variable names are shown along the top with the % of missing data for each.

Our data analysis explored and quantified the proportion of workers who are in decent vs poor quality employment within the dataset. Employment quality was assessed through statistical analyses on working conditions with reference to pay rates, hours worked per day, timing of work shifts, whether paid leave arrangements are in place, and whether any social security arrangements are in place. Using the fine-scale geospatial data inherent in the dataset, spatial mapping was conducted, allowing insights about the spatial distribution of socio-demographic and cultural factors.

Analysis of the dataset included examinations of how quality of employment varies across geographical regions, by type of employment (such as cleaning, kitchen help, childcare, gardening), by demographic characteristics (such as age, marriage status and educational status) and by cultural factors (such as caste, religion and ethnic background).

The data analysis process included general data analysis, geospatial analysis and advanced/targeted analysis. A cluster analysis (to establish a measure of statistical distance between individuals in the dataset, uncover ‘archetypes’ within the data, and examine variability of features across archetypes) was not completed in detail as the data were too variable for this kind of analysis.



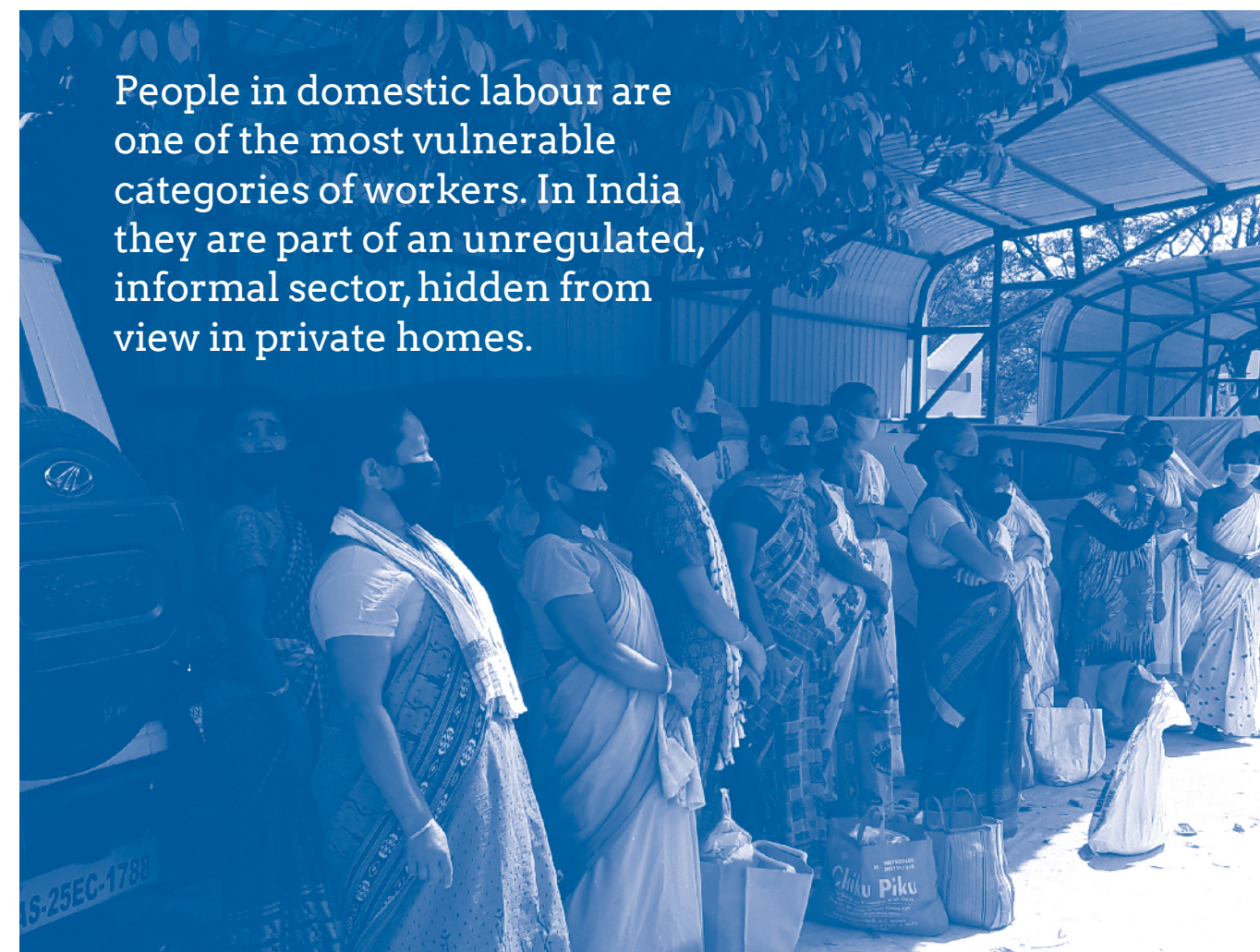
1.2 Key findings from the data analysis

1. **85% of domestic workers were from marginalised social groups.** This includes Scheduled Castes/Dalits. Scheduled Castes, Scheduled Tribes and Other Backward Classes (OBCs) are distinguished from each other but grouped to distinguish these socially marginalised and disadvantaged groups from the four ‘forward’ castes of Brahmin, Kshatriya, Vaishya and Shudra.
2. **Poor working conditions were identified.** The vast majority of workers had only oral contracts (99%). A majority had no weekly holiday (63%) or annual leave (63%). A majority had no access to medical services (74%). Annual leave was unpaid for over half of respondents (51%).
3. **The forms of work undertaken are consistent.** Respondents were engaged in cleaning tasks (82%), washing (80%), and cooking/kitchen work (51%).
4. **Drivers for taking on domestic work are consistent.** The main drivers include supporting the household and paying for education.
5. **There is a distinct demographic for domestic workers.** Domestic workers in the sample were typically young married women, with at least one child. 99% of the domestic workers interviewed were female and 73% were married.
6. **Poverty seems to be inhibiting education.** There is a general trend in low educational attainment that ceases at primary and middle school level, with poverty as the main driver for school dropout rates. A quarter of respondents attended lower primary, half attended upper primary/middle school, and one quarter attended secondary but 74% discontinued study.
7. **Significant differences were found between states.** Key variables (mean age, mean salary) differed considerably between states, and there were high levels of standard deviation within states. The table below indicates that:
 - a. Mizoram and Nagaland had a significantly younger average age for workers surveyed (and average age at which work was started) than the other states.
 - b. Mizoram had significantly higher average number of hours worked, potentially indicating exploitation. However, it also had the highest mean salary, potentially indicating that longer hours are reflected in better wages.
 - c. Nagaland had the lowest wages of all states, potentially indicating exploitation. But it also had the lowest average number of hours worked, potentially suggesting that lower wages are associated with fewer working hours.

State	# responses	Mean age (SD)	Mean age when started work (SD)	# hours worked (SD)	Mean salary (SD)
Assam	7453	38 (11)	28 (10)	6 (3)	2284 (3656)
Manipur	766	39 (11)	29 (10)	6 (3)	3404 (1391)
Meghalaya	2194	36 (11)	22 (9)	6 (3)	2900 (1653)
Mizoram	369	22 (6)	19 (5)	13 (3)	3427 (886)
Nagaland	374	22 (10)	18 (9)	7 (2)	1763 (3039)
Tripura	603	39 (11)	29 (10)	5 (3)	2306 (1386)

8. **Child labour rates are high.** A small percentage of the sample were children under 14 (0.8%, a total of 91 respondents) at the time of the survey. This is because the focus of the surveys was not child labour. The surveys were targeted at individuals who had signed up to CDI’s “Domestic Workers Union Structure” project and so underrepresent the incidence of child labour in domestic work. Nonetheless, 6% (a total of 816 respondents) had begun work before the age of 14. This is significant for six states that are not normally considered to have high incidences of child labour. Five other states—Uttar Pradesh, Bihar, Rajasthan, Maharashtra, and Madhya Pradesh—are understood to constitute nearly 55% of total working children in India (ILO, 2017). Even the figure of 6% may imply that official estimates of the number of child domestic workers in India are too low.

People in domestic labour are one of the most vulnerable categories of workers. In India they are part of an unregulated, informal sector, hidden from view in private homes.



Comprehensive data analysis

2.1 Descriptive statistics

The purpose of this section is to provide background to the data, presenting statistics describing each of the variables from the surveys. The continuous and categorical variables included in the surveys are summarised in Tables 1 and 2, respectively. These variables are then presented in more detail in the following text, with associated visualisations, in the sub-sections: 1.1 Personal information, 1.2 Education details, 1.3 Work history, 1.4 Current terms of employment, and 1.5 Current working conditions.

Table 1. Every continuous variable in the survey is listed. The mean, minimum, maximum, and standard deviation values for each are given, along with the total number of respondents for which data are available (n).

Variable	Mean	Minimum	Maximum	Standard Deviation	n
Age	36.5	4	95	11.70	11,756
Number of family members	2.6	0	5	1.58	11,759
Number of children	1.1	0	5	1.15	11,759
School standard	6.6	0	13	2.85	5,101
Age started work	26.5	1	80	10.27	11,758
Number years working	10.0	0	64	9.30	11,759
Number previous workplaces	1.4	0	29	1.78	11,759
Present job since	4.9	1	40	5.44	7,822
Monthly salary	Rs3417.30	Rs0	Rs250000	6803.76	11,759
Extra allowance	Rs260.0	Rs10	Rs5000	435.82	541
Number hours worked	6.2	0.5	24	3.06	11,747
Number social security	0.02	0	3	0.18	11,759

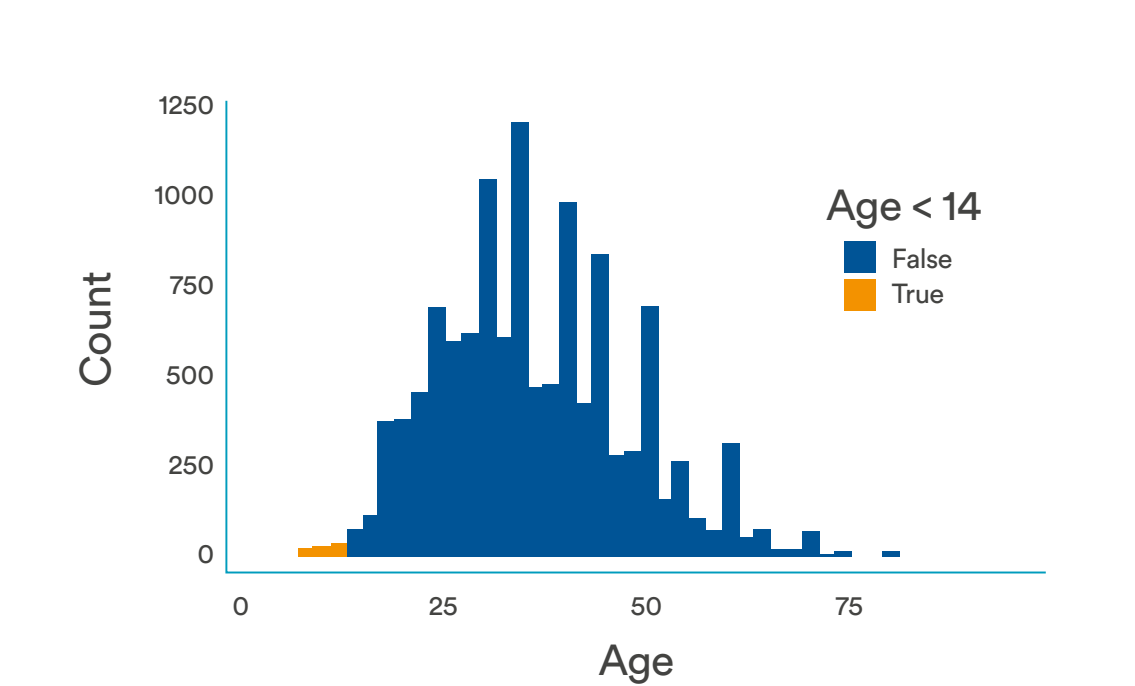
Table 2. Every categorical variable in the survey is listed. The most frequently occurring response is indicated (mode), along with the percentage of respondents who gave that response (%) and the total number of respondents for which data are available (n).

Variable	Mode	%	n
Sex	Female	99.3	11,759
Caste	OBC	33.9	11,759
Religion	Hinduism	60.0	11,759
Marital status	Married	73.4	11,759
Attended school	Did not	56.5	11,759
School type	Government	93.4	5,115
School expense paid	Self	61.7	5,115
Reason discontinued studies	Poverty	73.5	5,115
Reason for changing jobs	Financial	87.5	6,484
Purpose of work	Working livelihood	94.0	11,759
Employment type	Part time	66.4	11,759
Work contract?	No	67.5	11,759
Contract type	Oral	98.9	3,821
Payment frequency	Monthly	93.0	11,759
Salary deductions?	No	92.0	11,759
Reason for salary deduction	Leave taken	85.0	944
Extra allowance?	No	95.4	11,013
Weekly holiday	No	62.6	11,759
Annual leave	No	63.0	11,759
Annual leave pay	Without pay	50.6	4,355
Tasks	Cleaning	82.1	11,697
Access to medical facilities?	No	74.4	11,759
Social security type	BPL	54.3	230

Personal information

The mean respondent age was 36.5 (4–95, SD = 11.69; Fig. 3) with a mean number of family members and mean number of children of 2.6 (0–5, SD = 1.58) and 1.1 (0–5, SD = 1.15), respectively. Of the respondents, 91 (0.8%) were under the age of 14 (Fig. 2), and therefore classed as children for the purposes of work under the Child and Adolescent Labour (Prohibition and Regulation) Act, 1986.

Figure 2. Histogram showing the ages of all respondents for which data were collected (n = 11,756). Any respondents aged 14 or over (n = 11,665) are shown in blue. Respondents aged below 14 years of age (n = 91) are shown in orange.



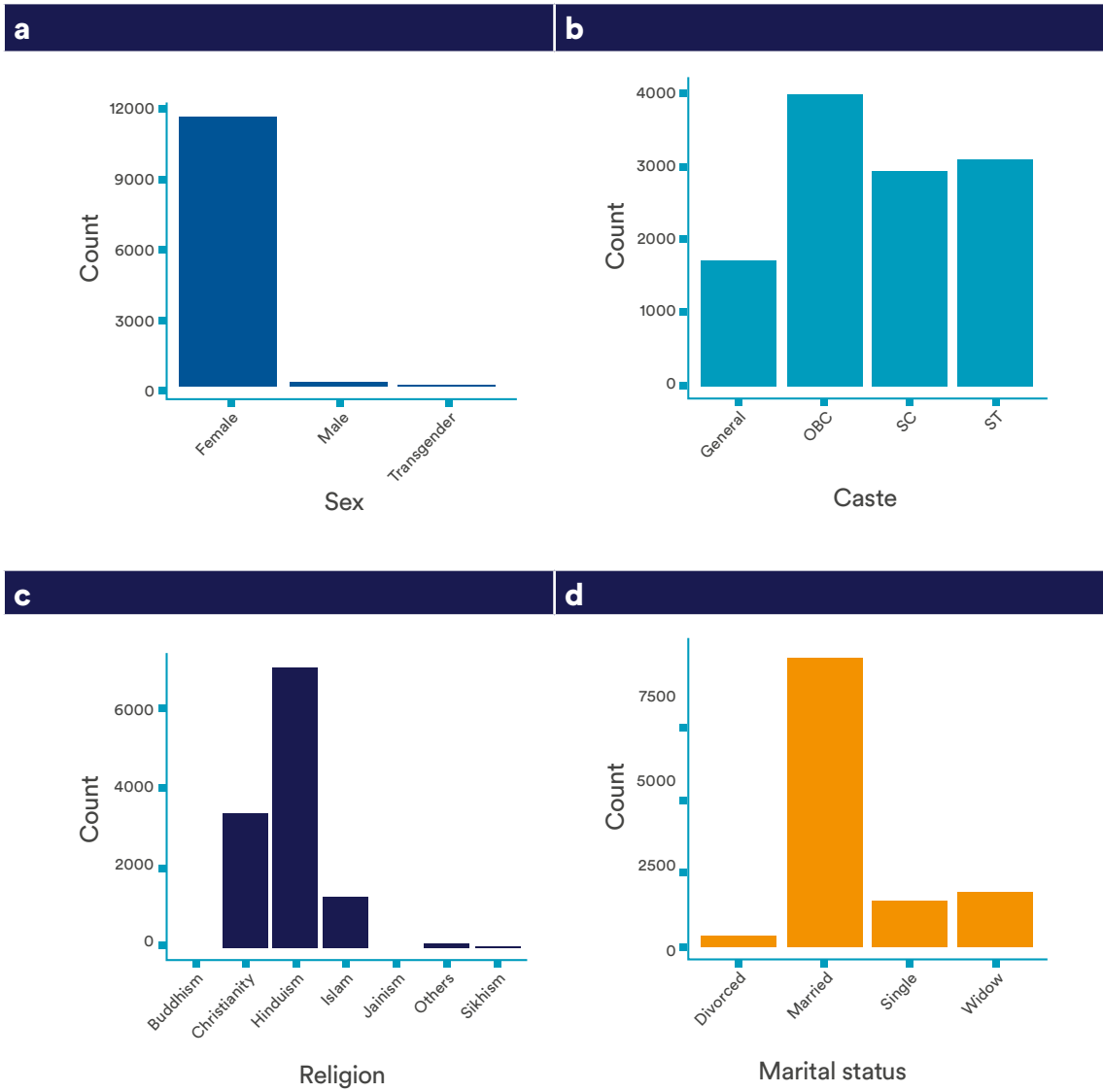
The vast majority of respondents identified as female (99.3%), 0.7% as male and 0.04% as transgender (Fig. 3).

The most common caste represented was Other Backward Class (OBC, 33.9% of respondents) followed by Scheduled Tribes (ST) and Scheduled Castes (SC) (26.4% and 25.1%, respectively), and General (14.6%) (Fig. 3). OBC is a collective term used by the Government of India to classify castes which are educationally or socially disadvantaged, and is one of several official classifications of the population of India, along with SCs and STs.

The majority of the respondents identified their religion as Hinduism (60.0%), with Christianity (28.6%) and Islam (10.6%) the second two most common religions. The remaining 0.8% of respondents were split between other religions (72 respondents), Sikhism (12 respondents), Buddhism (4 respondents) and Jainism (1 respondent) (Fig. 3).

The majority of respondents were married (73.4%) with fairly even numbers of widowed (13.4%) or single (10.9%) respondents and 2.3% of divorced respondents (Fig. 3).

Figure 3. Bar charts showing (a) sex, (b) caste, (c) religion and (d) marital status of all respondents (n = 11,759).



Education details

Of the 11,759 respondents, there was a fairly even split between those that did not (56.5%) and did (43.5%) attend school. Of those that did attend school (n = 5,115), the vast majority attended government schools rather than private schools (93.4% vs 6.6%).

The expense of attending school was met, for the majority of respondents, by themselves (61.7%). Guardians or employers met the expenses for 32.3% and 5.9% of respondents, respectively (Fig. 4a).

The most common reason for discontinuing education was poverty (73.5% of respondents), followed by not being interested (14.1%) and death of parents (7.0%) (Fig. 4b). The reason of “death of parents” also points to economic constraints due to loss of financial support. The remaining respondents listed “Any Other Reason” for leaving school, with around half (45.3%) providing further details. 82.3% of these respondents identified marriage as the reason for leaving education and 8.1% identified sickness.

Of those that did attend school, only 0.1% of respondents did not give further details on the level of school education obtained. For the rest, their school standard is measured

on a scale from 0-13. This scale covers Pre-Primary (0), Lower Primary (1-4), Upper Primary/Middle School (5-8), Secondary (9-10), Senior/Higher Secondary (11-12) and University/Higher (13) education. The mean school standard for respondents was 6.6 (0-13, SD = 2.85, n = 5,100) with Upper Primary/Middle School levels 5-8 the most common school standard obtained (44.1% of respondents) (Fig. 5).

Figure 4. Bar charts showing, for all respondents who attended school (n = 5,115) (a) who was responsible for meeting the schooling expense of the respondent and (b) the reason the respondent discontinued their studies.

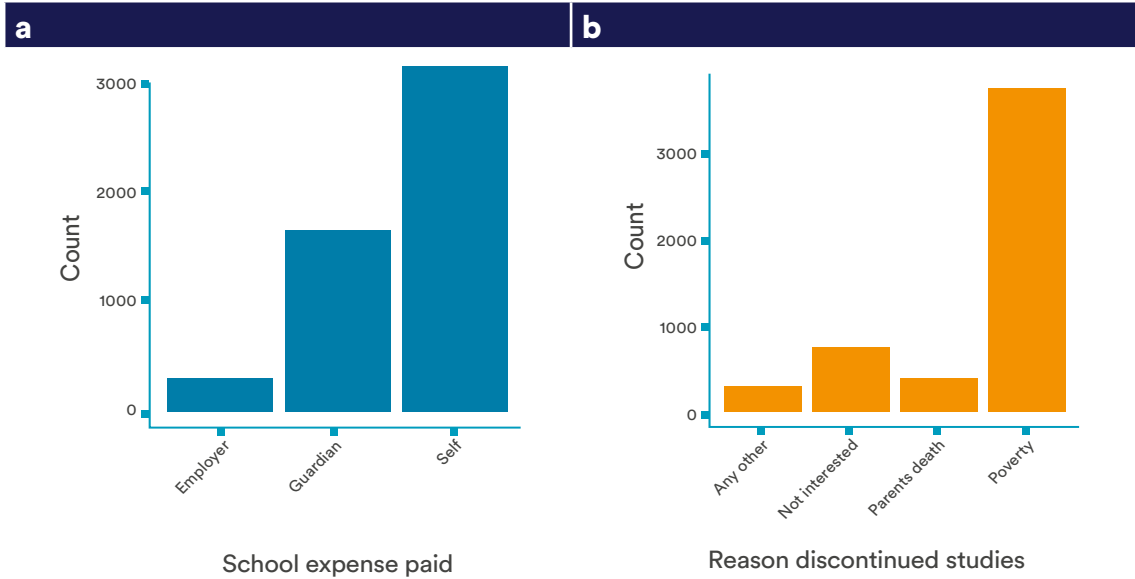
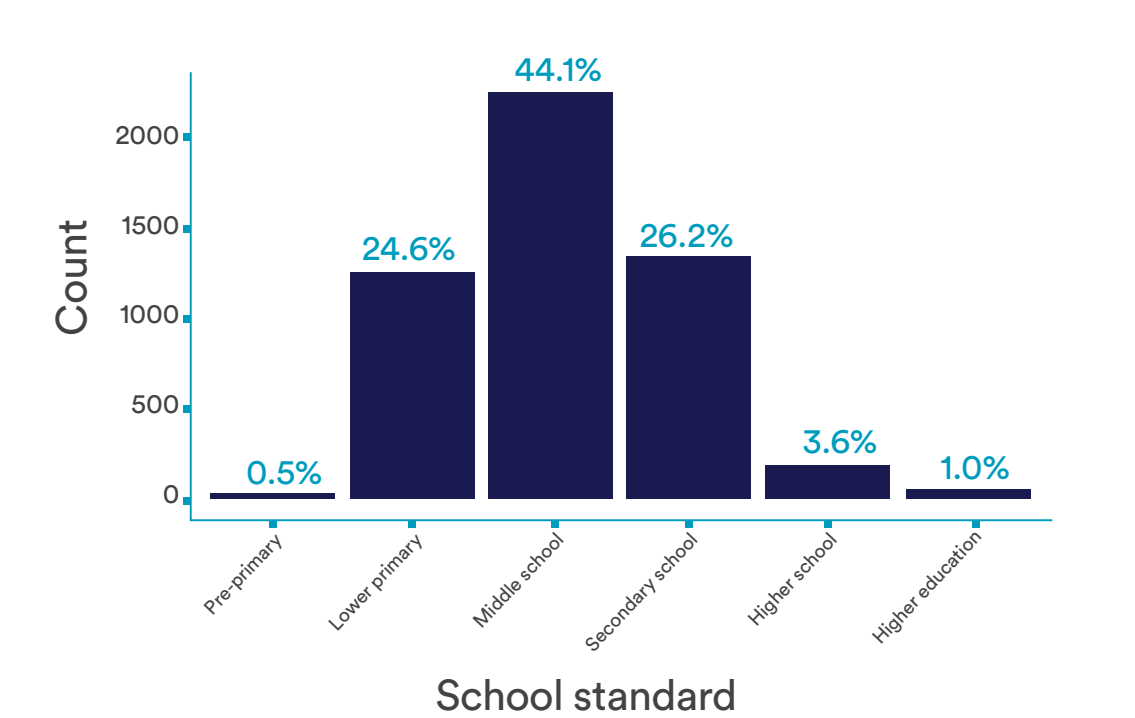


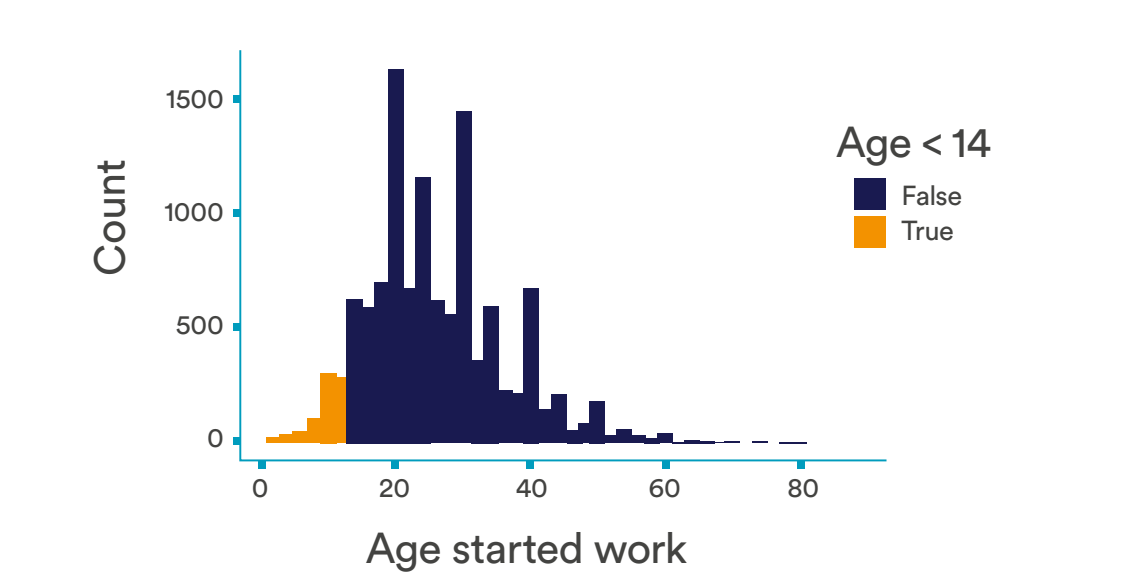
Figure 5. Bar charts showing, for those respondents who attended school, the level of education obtained (n = 5,101). The count of individuals in each schooling standard is indicated by the height of the bars, with the percentage of respondents who attended school indicated at the top of each bar.



Work history

The mean age for starting work for the respondents was 26.5 (1-80, SD = 10.27; Fig. 6). When examining the prevalence of child labour among the surveys, 816 people (6.4% of respondents) began work below the age of 14 (Fig. 6). The mean number of years working was 10.0 (0-64, SD = 9.30) with a mean number of previous workplaces of 1.4 (0-29, SD = 1.78).

Figure 6. Histogram showing the age at which respondents started work (n = 11,755). Any respondents aged 14 or over (n = 10,939) are shown in blue. Respondents who started work aged below 14 years of age (n = 816) are shown in orange.



Of the total 11,759 respondents, 3,981 (33.9%) have not changed jobs in the past. Of those that have changed jobs (n = 7,778) 66.1% gave reasons for this. The top 5 categories for reasons given are listed below, along with the percentage of respondents who gave reasons in each.

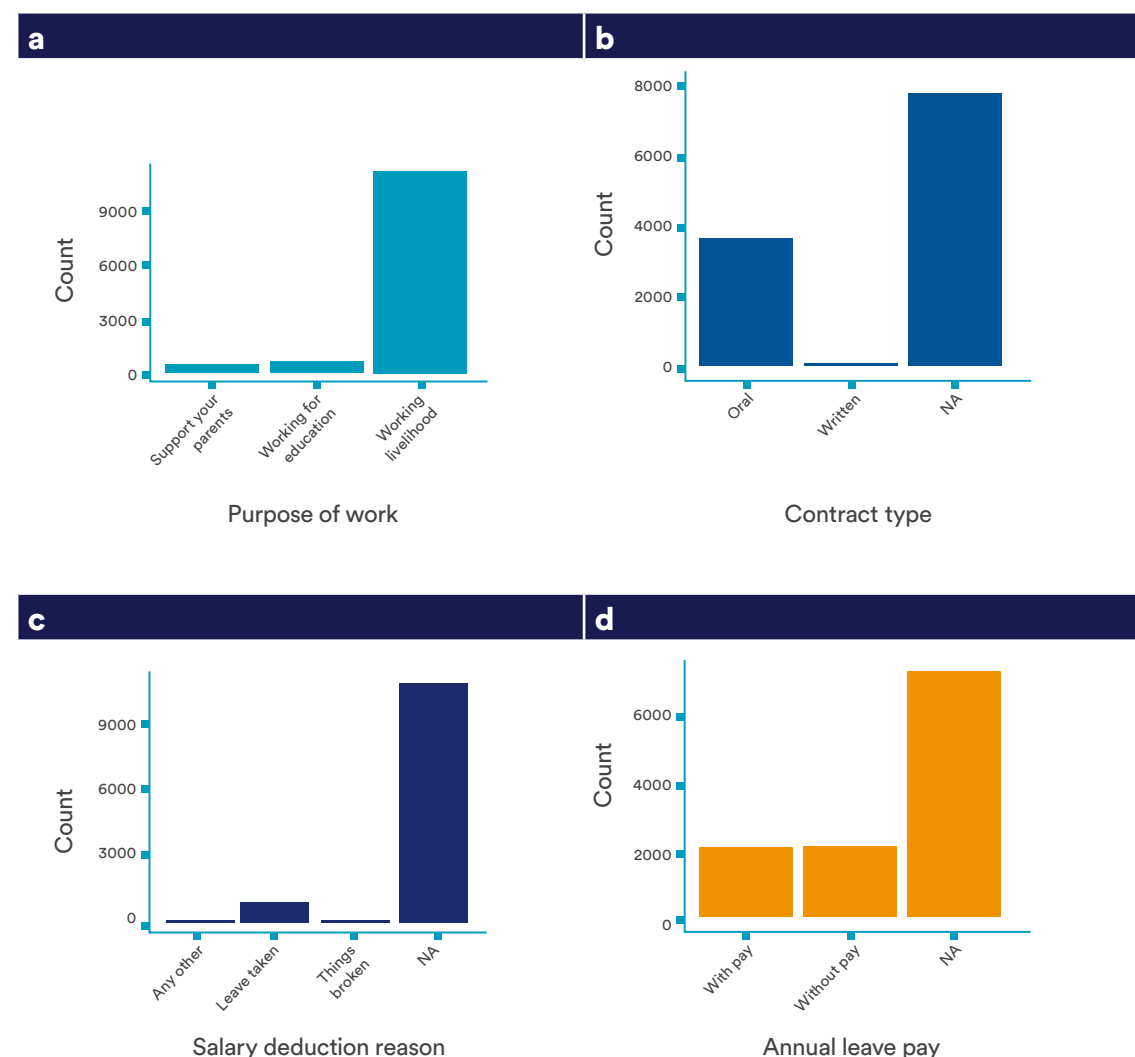
- 1. Financial (respondents were in poverty and looking for a better wage): 87.5%
- 2. Geographical (their or their employers address changed): 3.1%
- 3. Family/personal problems: 2.3%
- 4. Marriage: 1.2%
- 5. Illness: 0.9%

Other reasons given (by <0.4% of respondents) include (i) pregnancy or caring for small children, (ii) termination of previous job by employer, (iii) behaviour of the employer (not paying wages on time) (iv) behaviour of the employer (cruel/strict), and (v) death of a family member.

Current terms of employment

The respondents were asked to identify the purpose of work from three options. Respondents were able to select multiple purposes, so percentages do not add up to 100. Working livelihood was the main purpose selected by 94.0% of respondents. Supporting parents and working for education were identified as purposes behind work by 3.6% and 4.6% of respondents, respectively (Fig. 7a).

Figure 7. Bar charts showing for each respondent (n = 11,759) (a) purpose of work, (b) the type of contract held, (c) any reasons for salary deductions (no salary deduction is denoted with NA), and (d) whether annual leave is granted (no leave granted is denoted with NA), and if granted, whether paid or unpaid. Respondents were able to select more than one purpose for work, and more than one reason for salary deduction so the total counts for these variables do not equal the total number of respondents.



The mean number of years respondents had held their current job is 4.9 (1-40, SD = 5.44). Of all 11,759 respondents, 66.4% identified their employment as part time. The 33.6% who are full time are live-in domestic workers. 32.5% of respondents had work contracts, only 1.1% of which were written, the rest being oral contracts (Fig. 7b).

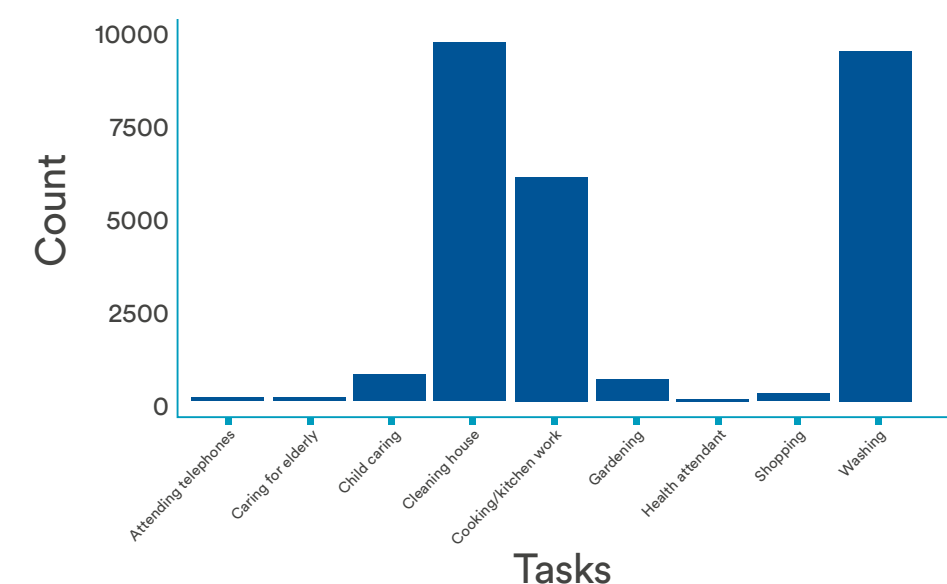
The mean monthly salary reported was Rs3,417.30/£36.53 (Rs0/£0 - Rs250,000/2672.16, SD = 6803.76). The majority of respondents are paid monthly (93.0%) with no salary deductions (92.0%) (Figs. 7c and 7d). Respondents that do experience salary deductions were asked to select the reason for this from 4 options; the most common being (i) leave taken (n = 803), followed by (ii) any other (n = 94) and (iii) things broken (n = 93) (Fig. 7d). Only two respondents who selected “any other” provided further reasons for leave taken, both being “illness.” Only 541 respondents (4.6%) stated that they received extra allowance, the value of which ranged from Rs10/£0.11 - Rs5000/£53.45 (mean = Rs260.0/£2.78, SD = 435.82).

The majority of respondents received no weekly holiday (62.6%) and no annual leave (63.0%). Where annual leave was given, it was fairly evenly split between being given with or without pay (Fig. 7d).

Current working conditions

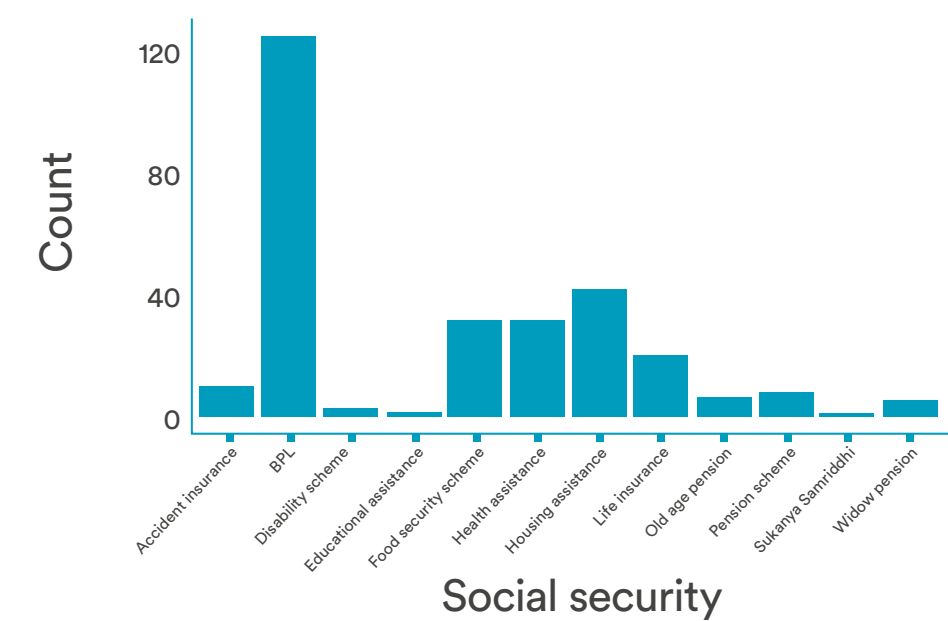
Respondents were asked to identify the tasks that they undertook during their work; they were able to identify more than one task so the percentages do not add up to 100 (Fig. 8). The most common tasks undertaken by the respondents were Cleaning (9653 people, 82.1%), Washing (9436, 80.2%) and Cooking and Kitchen Work (6030, 51.3%). The average number of hours worked by respondents was 6.2 (0.5–24, SD = 3.06).

Figure 8. Bar charts showing the main tasks that the respondents undertake in their work. Respondents were able to identify more than one task so the total numbers of people undertaking each task do not equal the number of respondents who gave details on this (n = 11,697).



The majority of respondents had no access to medical facilities (74.4%) and only 230 respondents (2%) had any level of social security in place. Of a potential list of social security types, respondents were asked to select which they had (Fig. 9). The maximum number of social security schemes held by any one person was 3 (0 – 3, mean = 0.02, SD = 0.18), the most common being BPL (54.3%).

Figure 9. Bar charts showing the social security schemes that respondents had in place. Respondents were able to select more than one social security scheme if they had multiple in place, so the counts for each social security scheme do not equal the total number of respondents who had social security in place (n = 230).



2.2 Correlation analysis

The purpose of this section is to conduct an examination of numerical variable inter-relationships through correlation analysis to establish underlying and explanatory themes within the data. Prior to the correlation analysis presented below on a subset of the variables, correlation analysis was conducted on all variables to establish inter-correlations in the dataset. These variables were removed from the subsequent correlation analysis in order to aid interpretability of the results presented (Fig. 10).

The variables removed include: multiple replicates of variables for individual family members (i.e. family age, family education, and family income). These variables are summarised with ‘mean family age’, ‘mean family education level’, ‘number of children in the family’, ‘number of family members’, and ‘family income’. This ‘family income’ variable is combined with the respondent’s ‘salary’ to produce a ‘total income’ variable. The correlation analysis is presented visually below in Fig. 10, and explicated in more detail in Table 3.

Figure 10. Figure showing the results of the correlation analysis. Correlations in blue indicate positive relationships, while those in red indicate negative relationships. The gradient of colour indicates the strength of the relationship (darker = higher r2). Only correlations that achieved statistical significance are shown, any correlations that were not significant (p = >0.05) are left blank.

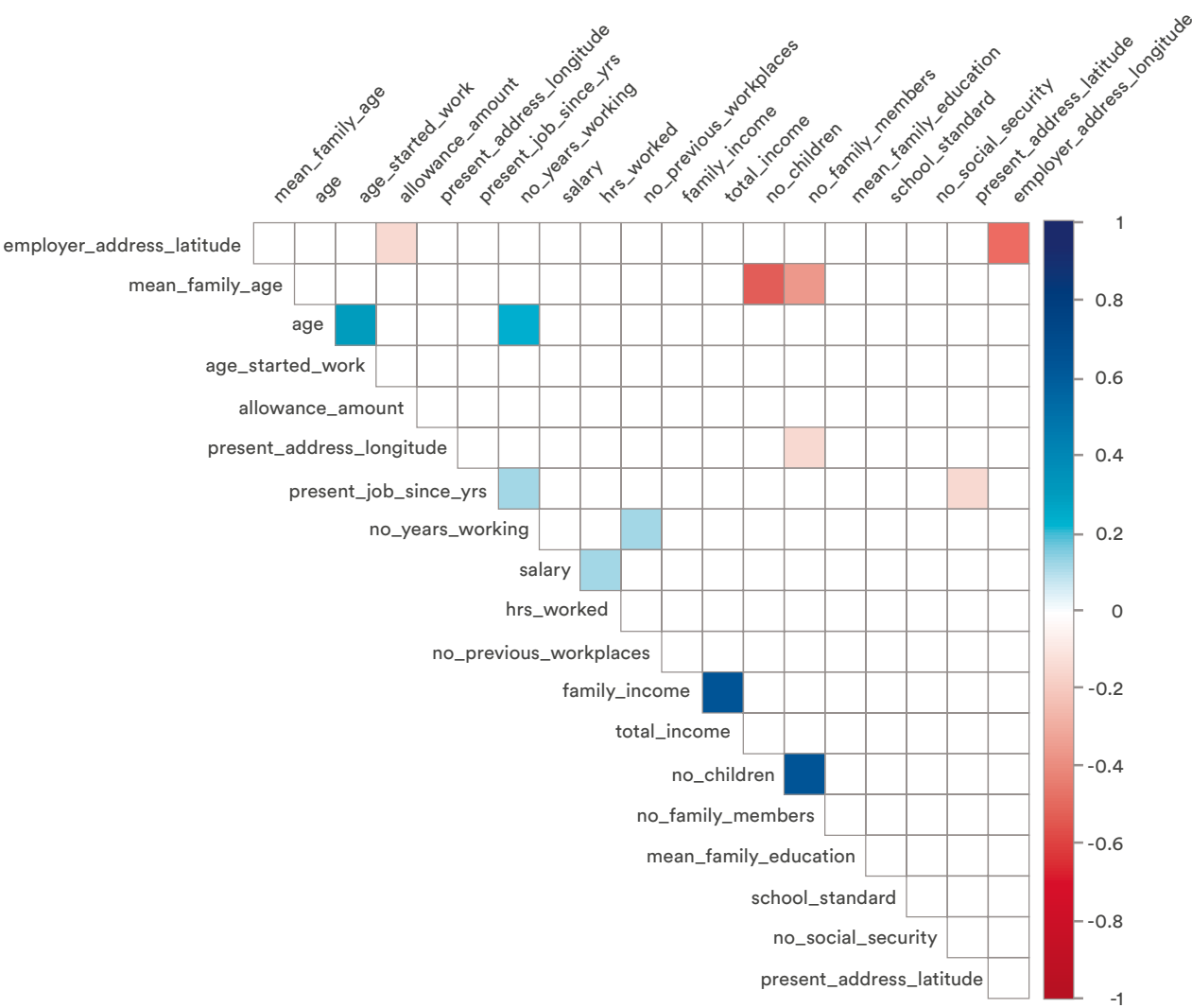


Table 3. Table detailing the key correlations in the dataset. The correlations are numbered, with the two variables comprising each named. The r^2 of the correlations are reported and additional details on the correlation and possible indications noted. Potentially obvious and/or uninteresting correlations have no highlight and potentially important and/or interesting correlations are marked with blue highlight.

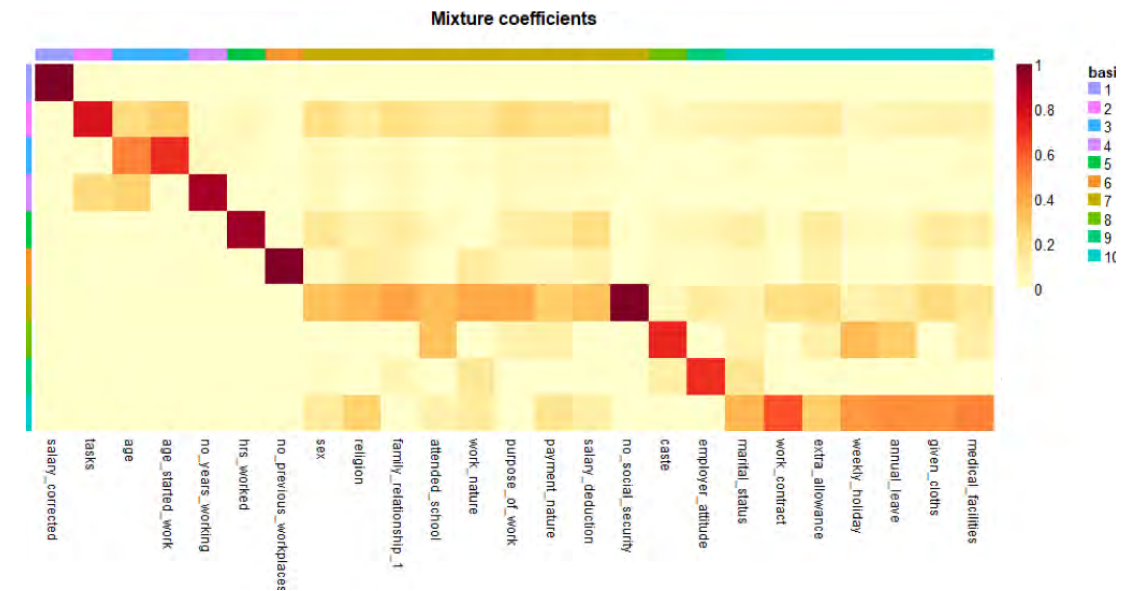
No.	Variable	Variable 2	r^2	Details
1	Family income	Total income	0.86	An obvious correlation: the higher the income earned by family members (excluding the respondent), the higher the total income when including the respondent's income
2	No children	No family members	0.80	An obvious correlation: the higher the number of children, the higher the number of family members.
3	Age	Age started work	0.63	The higher the respondent age, the higher the age of the respondent when starting work. This implies older respondents started work later in their lives. This may indicate recall bias, or may indicate that the age of starting work was older in the past.
4	Mean family age	No children	-0.58	An obvious correlation: the higher the mean family age, the lower the number of children as a higher number of children will necessarily decrease the mean family age.
5	Age	No years working	0.51	An obvious correlation: the higher the respondent age, the higher the number of years working. A greater age indicates to more years available to have been working.
6	Employer address latitude	Employer address longitude	-0.49	An unimportant correlation: the higher the employer address latitude, the lower the employer address longitude. A by-product of the spatial patterning in the data.
7	Present job since (yrs)	Number years working	0.40	The longer the respondent has been in their present job, the longer they have been working. This potentially indicates that respondents are more likely to change jobs more frequently at the beginning of their career, and retain the same job for longer periods later on. Potentially this could be related to changing jobs in pursuit of better wages and/or as a result of life changes (marriage/relocating) which happen less often later in life (equivalent to a greater number of years working).

No.	Variable	Variable 2	r^2	Details
8	No years working	No previous workplaces	0.35	An obvious correlation: the more years you have been working, the higher number of previous workplaces. This seems related to the previous correlation.
9	Salary	Hours worked	0.33	The higher the respondent's salary, the greater the number of hours worked by the respondent. This makes logical sense, and may indicate that, for some people, a fair equivalency of increased hours equates to an increased salary is being worked out. The relatively low r^2 , however, indicates that for many people, the number of hours worked does not have a strong bearing on their salary.
10	Mean family age	No family members	-0.33	An obvious correlation: the higher the mean family age, the lower the number of family members. This is related to correlation 4 (above). Number of family members is increased by increased numbers of children which reduces mean family age.
11	Employer address latitude	Allowance amount	-0.27	As latitude decreases, allowance amount increases. This potentially indicates that allowance amounts are higher in the south of the region covered by the surveys. This low r^2 , however, may indicated this relationship only holds for few respondents.
12	Present address longitude	No family members	-0.19	As longitude increases, the number of family members decreases. This potentially indicates that there are fewer family members in families living in the east of the region covered by the surveys. This low r^2 , however, may indicated this relationship only holds for few respondents.
13	Present job since (yrs)	Present address latitude	-0.18	As the number of years a respondent has held their current job increases, the latitude of their present address decreases. This potentially indicates that people work for longer in one place in the south of the region covered by the surveys. This low r^2 , however, may indicated this relationship only holds for few respondents.

2.3 Factor analysis

Non-negative matrix factorisation (NMF) is a group of algorithms based on analysing the dataset as matrices. NMF has become a widely used tool for the analysis of high dimensional data as it automatically extracts sparse and meaningful features from a set of data vectors. NMF analysis has been used here to extract seven key features from the dataset, and to indicate which variables contribute to these features (Fig. 11).

Figure 11. Output from NMF showing the 7 key bases for the dataset. The bases are indicated by the colour ramp at the top and left of the figure. All of the variables upon which the NMF analysis was carried out are identified along the bottom of the figure. The colour ramp from red to yellow shows the strength of the relationship between each variable and the bases (darker red = stronger, lighter yellow = weaker).



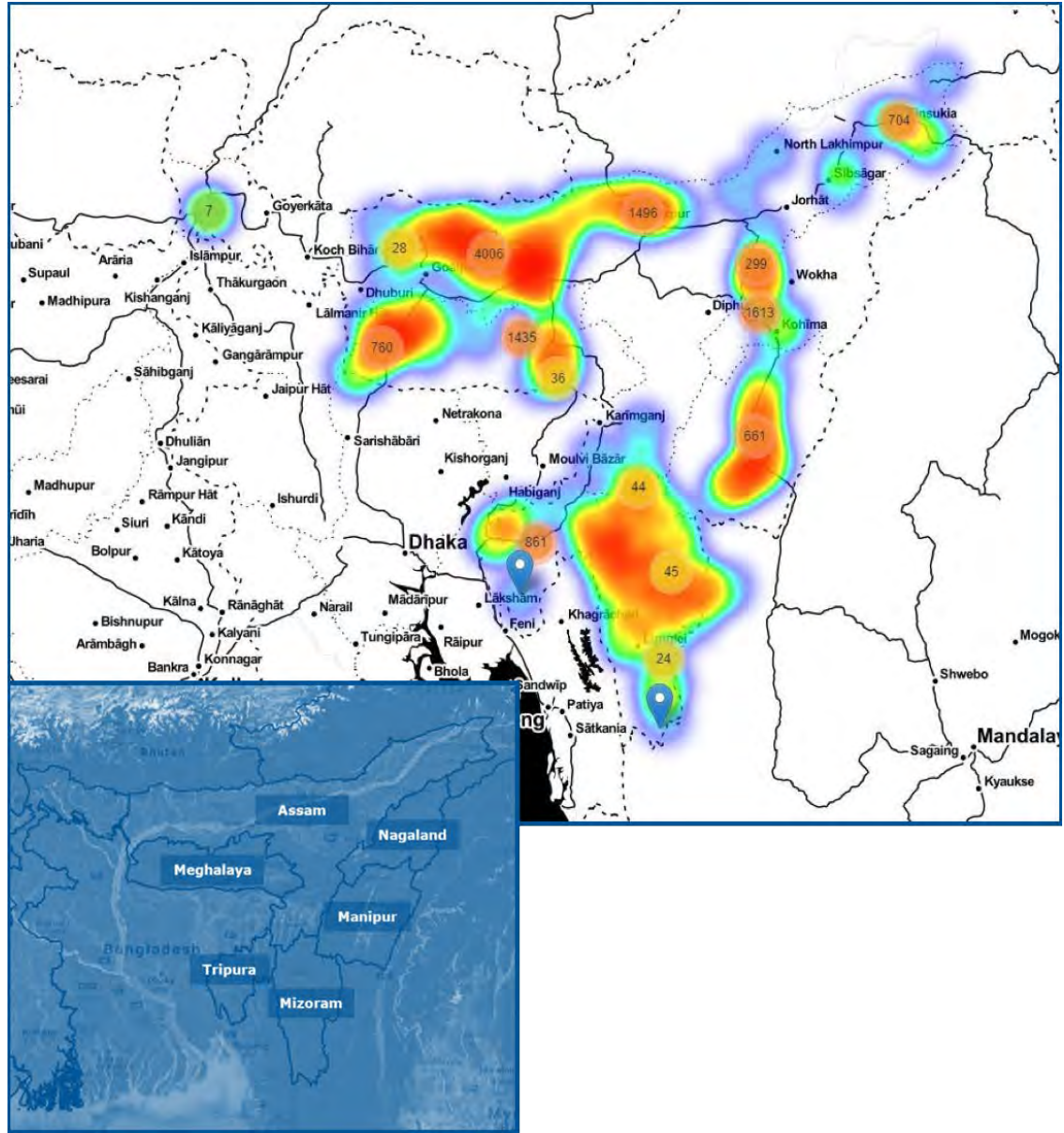
The key features identified from the NMF analysis are given below. The first five of these are retained as the most easily interpreted features and are analysed with respect to spatial variation in the dataset in the next section.

1. Salary
2. Tasks
3. Age of the respondent (at the time of survey and when starting work)
4. Number years working
5. Number hours worked

2.4 Spatial variation and state comparisons

The 11,759 respondents surveyed were located across six states in North East India: Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura (Fig. 12). The number of respondents was not equally spread across the states, with Assam having the highest number ($n = 7,453$) followed by Meghalaya ($n = 2,194$), Manipur ($n = 766$), Tripura ($n = 603$), Nagaland ($n = 374$) and Mizoram ($n = 369$). In this section, we will examine some of the key factors outlined above, and how patterns of these variables vary in a geospatial context.

Figure 12. Main figure (a) shows a heat map covering the extent of the surveying region, comprising six states of India: Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura. The colour ramp indicates the number of respondents in certain areas (blue = fewer respondents, red = more respondents) with numbers of respondents for aggregated areas indicated. The state boundaries are depicted with dashed black lines. To enable better identification of the states, the inset (b) shows the same region with the state boundaries marked with solid black lines, and the states named.



Salary

There was a statistically significant difference in mean salary across the six states in the region surveyed (One-way ANOVA: $p = <0.05$).

Mizoram had the highest mean salary at Rs3427.0. Nagaland had the lowest mean salary at Rs1763.0 (Table 4).

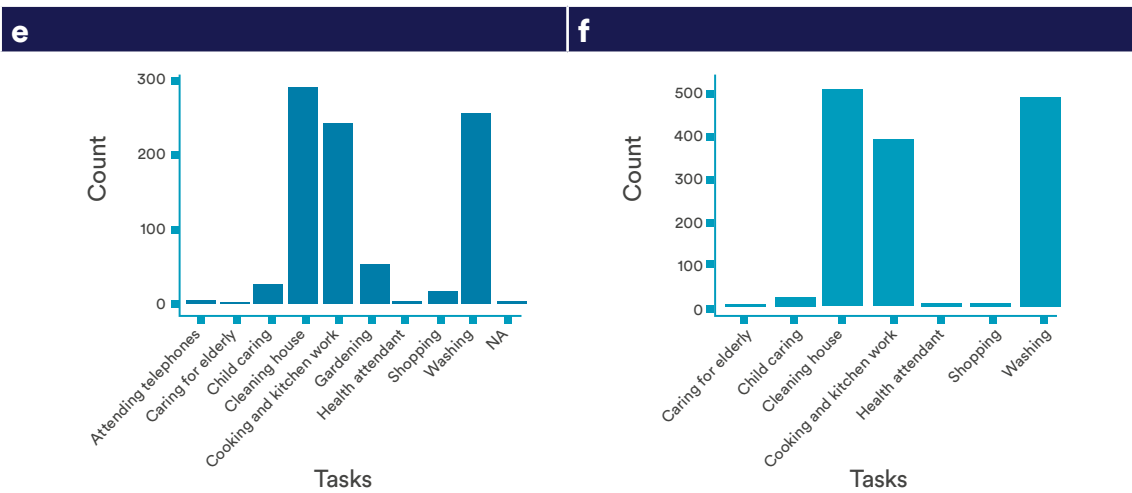
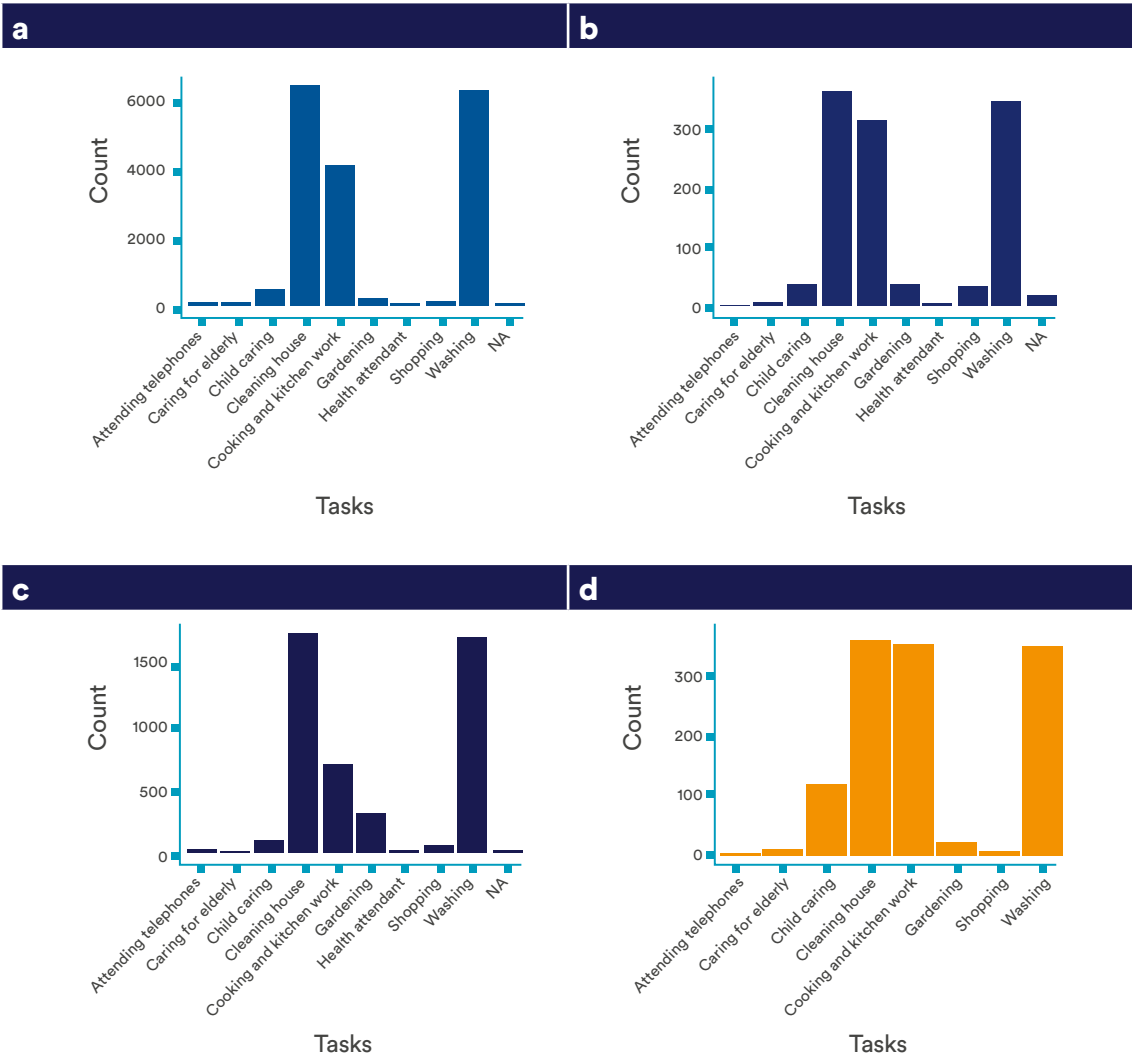
Table 4. Table showing the mean and standard deviation (SD) of the salary received by the respondents across the six states covered by the surveys.

State	Mean	SD
Assam	2284.4	3655.9
Manipur	3403.9	1390.6
Meghalaya	2899.9	1652.8
Mizoram	3427.0	886.4
Nagaland	1763.0	3038.9
Tripura	2306.4	1386.4

Tasks

The tasks undertaken by respondents were fairly consistent across the states. All states had a high prevalence of respondents undertaking (i) cleaning of the house, (ii) cooking and kitchen work, and (iii) washing. In addition, there was a higher prevalence of gardening as a task in Meghalaya when compared to the other states, especially considering Tripura where gardening was never identified as a task undertaken (Fig. 13).

Figure 13. Bar charts showing the main tasks that the respondents undertake in their work for (a) Assam, (b) Manipur, (c) Meghalaya, (d) Mizoram, (e) Nagaland, and (f) Tripura.



Age/age started work

Two of the most important variables to consider in this dataset are (i) the age of the respondents at the time of the survey, and (ii) the age of the respondents when they started work. There was a statistically significant difference in mean age of respondents and mean age of respondents when starting work across the six states in the region (One-way ANOVA: $p < 0.001$). Manipur had the highest mean age (38.6) and mean age when starting work (29.2) while Mizoram had the lowest mean age (21.7) and Nagaland the lowest mean age when starting work (Table 5). This mean age variance may be due to the “working-students” modalities in these states.

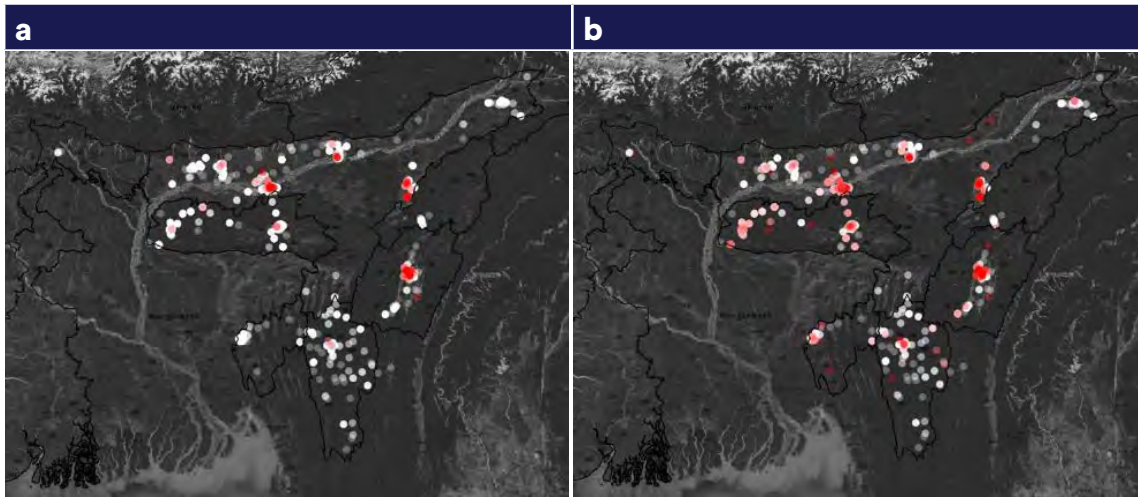
Table 5. Table showing the mean and standard deviation (SD) of the age of respondents at the time of the survey and age when starting work.

Age of respondent at time of survey			Age of respondent when starting work	
State	Mean	SD	Mean	SD
Assam	37.6	11.3	28.1	10.2
Manipur	38.6	11.3	29.2	10.4
Meghalaya	36.3	10.8	22.4	9.1
Mizoram	21.7	6.0	19.4	4.7
Nagaland	22.0	9.7	18.1	9.1
Tripura	38.5	11.3	28.5	9.8

The Child and Adolescent Labour (Prohibition and Regulation) Act of 1986 defines any person under the age of 14 as a child, and the employment of children as illegal. There are respondents included in the non-random sample survey who were under the age of 14, as well as respondents who started working before the age of 14 (Fig. 14). This is indicative of illegal practices in domestic work in the region surveyed.

In this section, the geographical distribution of this child labour is explored. Of all the states included in the surveys’ non-random sample, Nagaland had by far the majority of respondents under the age of 14 ($n = 80$, 87.9%). A further 10 children were located in Assam, and 1 in Mizoram (Fig. 14a). Assam had the majority of respondents who began work below the age of 14 ($n = 385$), followed by Meghalaya ($n = 254$), Nagaland ($n = 139$), Tripura ($n = 190$, Manipur ($n = 14$) and Mizoram ($n = 5$).

Figure 14. Maps showing the (a) age of respondents and (b) the age of respondents when they started work. Any respondents 14 years of age or older are depicted in white, while those below 14 years of age are depicted in red. State boundaries are outlined in black.

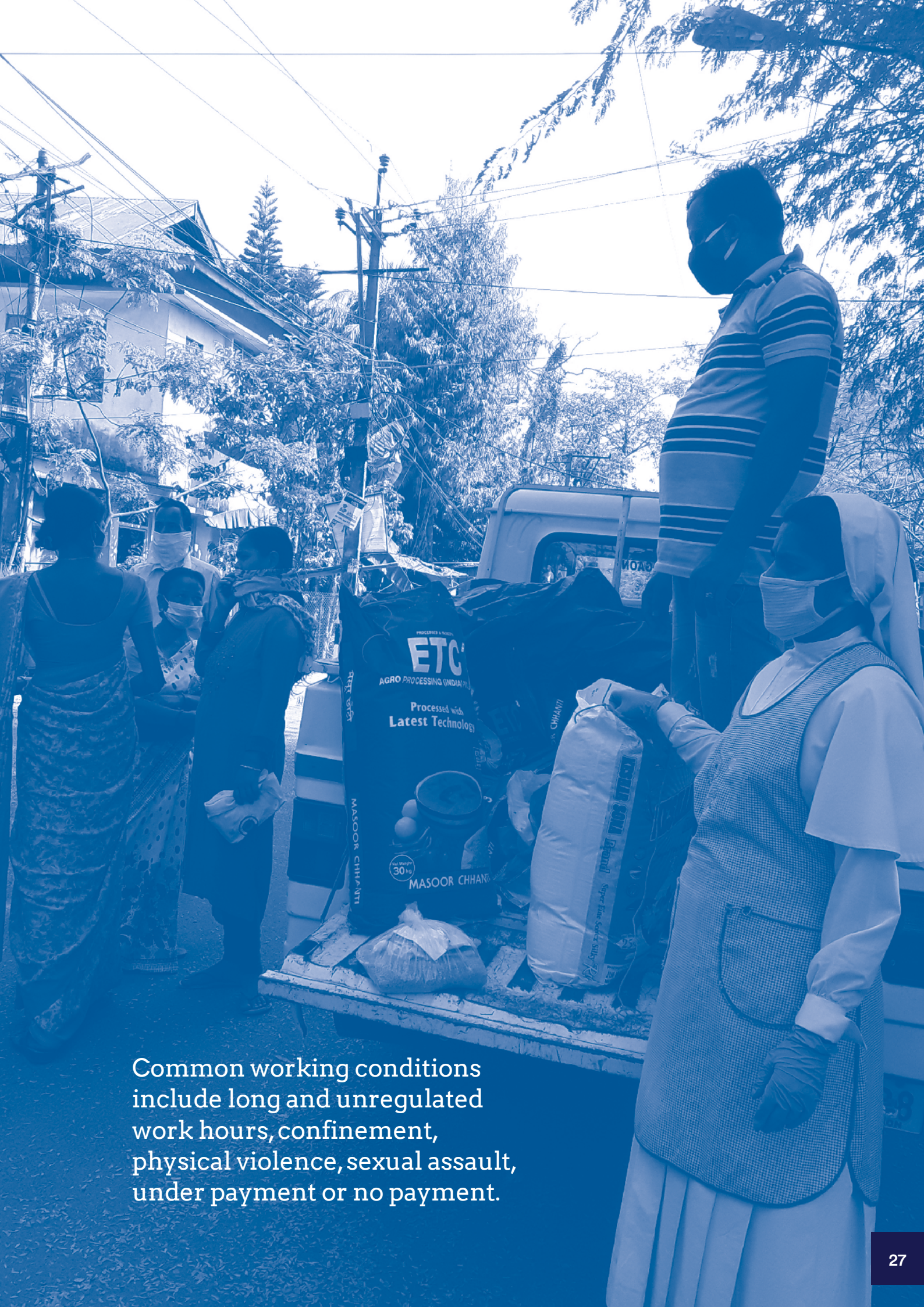


Number of hours worked

There was a statistically significant difference in mean number of hours worked across the six states (One-way ANOVA: $p = <0.001$). The majority of the states had similar mean number of hours worked (~6-6.5), with the exception of Mizoram, which had the highest number of hours worked at 13.0 (Table 6).

Table 6. Table showing the mean and standard deviation (SD) of the number of hours worked by the respondents at the time of the survey.

State	Mean	SD
Assam	6.0	2.8
Manipur	6.2	2.5
Meghalaya	6.2	2.9
Mizoram	13.0	3.1
Nagaland	6.5	2.1
Tripura	4.6	2.8



Common working conditions include long and unregulated work hours, confinement, physical violence, sexual assault, under payment or no payment.

The surveys in context

3.1 A Catholic tradition of human rights data-gathering

One context for the Domestic Workers Dataset is a Catholic tradition of gathering data on human rights and social justice. The Catholic Church is a global organisation that occupies a remarkable position to bring about positive social change through its many networks, communities, and formal Church structures. The 1968 Second Vatican Council signalled the ‘preferential option for the poor’ which found expression in more concerted efforts by the clergy and lay members working on issues related to poverty and social justice. Pope Francis’s encyclical *Laudato Si*, while primarily focussed on the environment, contains many continued Church commitments to social justice more broadly.

Observable manifestations of this social justice turn in the Church are evident in the work of the Church in Latin America. The development of liberation theology, ecclesial base communities, and human rights documentation centres yielded safe spaces for people to receive direct assistance and create collective responses to the negative consequences of different modes of economic development and the turn toward authoritarianism and conflict in the region during the 1970s and 1980s. Priests and Catholic communities mobilised against violence in the civil wars in El Salvador, Guatemala, and Nicaragua in the 1980s, and priests and nuns were targeted by death squads, state agents, and supporters of military and counter-revolutionary movements. The Church in Brazil was critical of the military regime between 1964 and 1985, and works today to help landless people in the rural sector and homeless people in the urban sector.

More broadly, networks of Sisters and chaplains provide a grassroots response to poverty, vulnerability, and the use of forced labour, child labour, bonded labour, and modern slavery in many countries around the world. These active networks include, for example, the Apostleship of the Sea which supports fishermen affected by human trafficking and labour abuses.

This focus on social justice and poverty also includes efforts by networks of Catholics to gather human rights data—on rights violations, labour abuses and the working conditions of the poor. For example, the *Vicaría de la Solidaridad* in Chile was a key agent in the documentation of human rights abuses that occurred during the Pinochet dictatorship between 1973 and 1989. Catholic organisations documented human rights violations during the civil conflict in Peru between 1980 and 2000. The first national survey of domestic workers was conducted on behalf of the Catholic Bishops Conference of India in the late 1970s, and laid out a manifesto for improving conditions for domestic workers in India.

Delivered by CDI as a wing of the Missionary Sisters of Mary Help of Christians (MSMHC), the first indigenous Congregation in North East India (founded in 1942), for the purpose of providing a baseline for an impact-oriented monitoring system, the Domestic Workers Dataset is part of this little-examined tradition of human rights data-gathering by Catholic congregations, organisations and networks.

3.2 Datasets on domestic workers

A second important context for the Domestic Workers Dataset is other efforts to survey domestic workers—including that first national survey of domestic workers completed in India in the late 1970s. We believe that the new dataset is the largest single set of surveys with domestic workers to date.

At a multi-country level, Anti-Slavery International published a report in 2013 that surveyed 1465 child domestic workers in six countries (India, Peru, the Philippines, Togo, Tanzania and Costa Rica). The ILO-DFID Partnership Programme on Fair Recruitment and Decent Work for Women Migrant Workers in South Asia and the Middle East (or “Work in Freedom” project) aimed to reach at least 100,000 women and girls in the sectors of domestic work and garment work across five countries (Bangladesh, India, Jordan, Lebanon and Nepal) between 2013-2018. But its activities, reports and publications did not include large-scale surveying of domestic workers.

At a national level, a baseline study to establish the prevalence of child domestic work in Ghana (CLADHO-IDAY, 2015) surveyed 2480 households. The first national survey of domestic workers in the United States interviewed 2086 domestic workers from 71 countries (Burnham and Theodore, 2012). And in India, the late 1970s survey of several thousand domestic workers conducted for the Catholic Bishops Conference of India (published as the National Socio-Economic Survey of Domestic Workers in 1980) covered 12 cities and determined that 17% of the domestic workers interviewed were under the age of 15.

The ILO also has produced a compendium of national domestic worker surveys (2013x). Its report highlights that many countries compile national and regional statistics on the number of domestic workers. The sources used are national labour force surveys (for example, LABORSTA) and other household sample surveys. These national surveys capture data that is related, for example, to domestic worker-reported hours of work and employment. Such surveys may be task-based, using the International Standard Classification of Occupations (ISCO-88); adapted from the International Classification by Status in Employment (ICSE-93), which is popular in Latin America (see Tokman, 2010); based on household rosters, which identify live-in domestic work in labour force surveys such as those carried out in the Philippines; or industry-based approaches which use the International Standard Industrial Classification of all economic activities. Despite that fact that it excludes domestic workers employed via agencies, the ILO considers this industry-based approach to be the most comprehensive mechanism for data collection.

Finally, at the local or regional levels, there have been multiple small-scale, bespoke domestic work surveys. For example, Agaya (2013) surveyed 114 domestic workers in Nairobi, Kenya. Gurtoo (2015) surveyed 487 domestic workers across two cities in Karnataka, India. And Tariq et al. (2020) surveyed 406 domestic workers in Karachi, Pakistan. Relative to all these studies—multi-country, national and local—the bespoke survey data in the Domestic Workers Dataset of nearly 12,000 surveys is very significant in its scale.

3.3 Data-gathering innovations for modern slavery

A third context for the dataset is the set of efforts to document modern slavery (including forced labour and forced marriage) at global, national and local levels, using a range of techniques. At the micro level, there are measurement strategies and data available for a range of dimensions relevant to slavery and exploitative labour. These include individual acts, violations, events, perceptions, attitudes, experiences, and feelings. At the macro level, available measurement strategies and data include those

for state violence, economic structures, demographics, third party violations, and political institutions. In the broader field of human rights, these micro and macro data strategies are in four main categories: (1) events-based data, (2) standards-based data, (3) survey-based data, and (4) new forms of data that have emerged with the advent of 'big data' and the use of computational social science, machine learning, and artificial intelligence.

Most relevant to the Domestic Workers Dataset is the survey-based approach—the use of structured, semi-structured and open survey tools to uncover perceptions, attitudes, and experiences of individuals. This approach can be used for revealing human rights abuses and has been adopted in work estimating the prevalence of modern slavery. Most notably, the ILO's Global Estimates of Modern Slavery (2017), delivered in collaboration with the Walk Free Initiative and the International Organisation for Migration (IOM), estimated that 40.3 million people were in modern slavery. Behind this estimate, and the Global Slavery Index (GSI) by Walk Free (2018) that breaks down the global figure of 40.3 million at the level of countries, are surveys administered by Gallup that collect data on individual vulnerability to modern slavery in high prevalence countries. The GSI uses respondent-level survey data and country-level predictions to provide estimates of modern slavery prevalence for 167 countries, including an estimate of 8 million enslaved people in India.

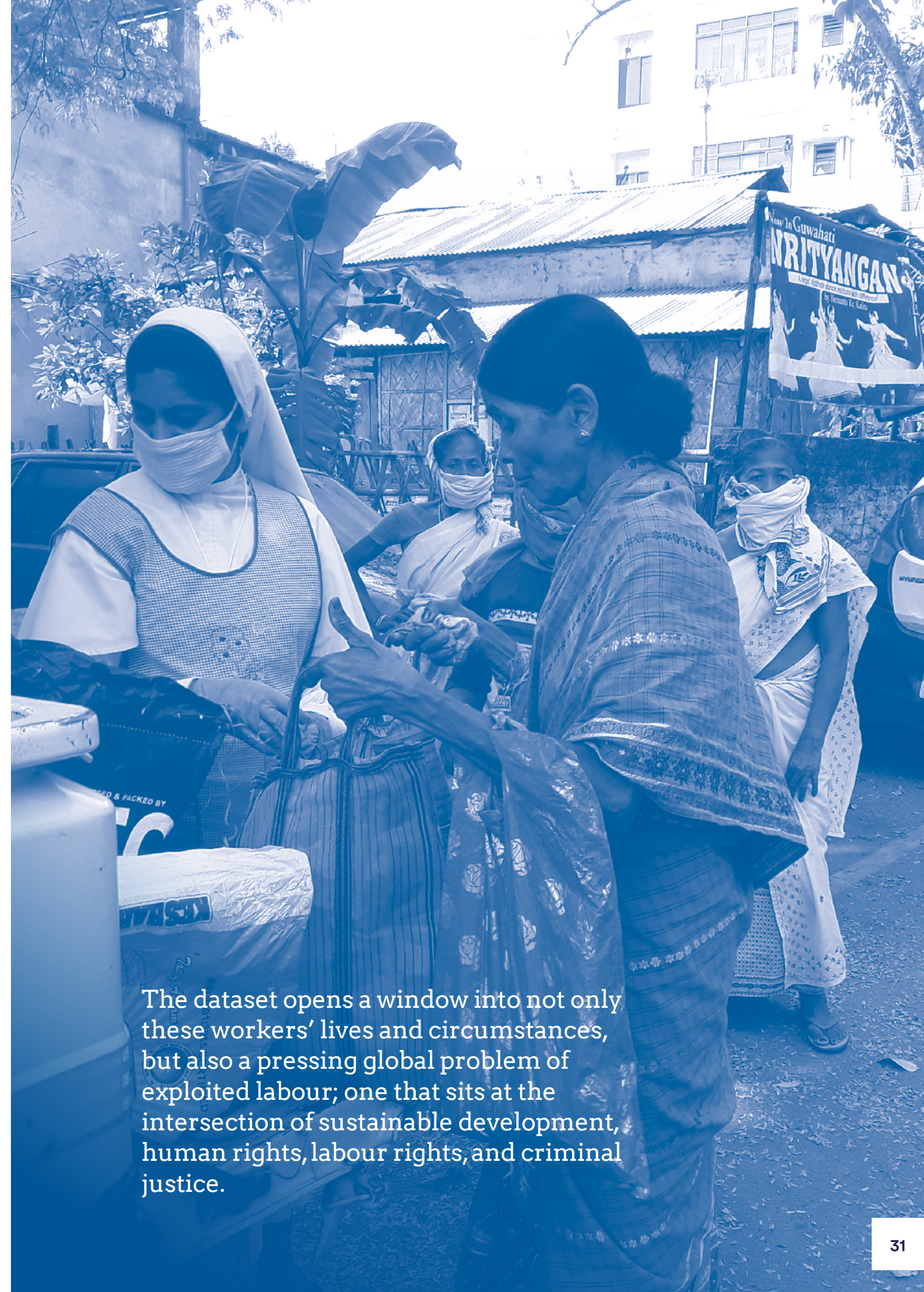
Other survey-based approaches to estimating slavery prevalence include cases of trafficking in San Diego (Zhang et al, 2014), forced marriage and child bearing of Myanmar women in China (Robinson and Branchini, 2018), and minors exploited in the adult entertainment sector in Kathmandu, Nepal (Dank et al, 2019). In these cases, the methodologies combine qualitative in-depth interview data with quantitative household survey data. The China study found that 39.8% of respondents experienced forced marriage, where respondents answered yes to at least one question relating to them being trafficked. The Nepal study found that 1650 minors (± 23) are working in adult entertainment venues in Kathmandu, an estimation based on 50 in-depth interviews and surveys from a sample size of 600 workers. The research on trafficking among migrant workers in San Diego adopted a similar methodology. It developed a legally and theoretically grounded survey instrument, GPS-enabled sampling strategies of households, and systematic data collection to estimate trafficking prevalence (Landman, 2020).

Seeking a more fine-grained and layered approach to data collection, a team of data scientists gathered street-level data in Dar es Salaam, Tanzania, in 2019 and was able through a series of data innovations to generate a forced labour heat-map for the city (Goulding et al, 2020). Dar es Salaam has a population of over 6 million across 90 administrative wards. But the research team believed there were too many people in each ward for the model to be informative. They used a community-generated map produced in collaboration with the dLab, a local NGO that promotes data literacy. This map sees the city as over four hundred hyperlocal sub-wards, divided by decision-making structures called 'shinas.'

Each shina is administered by a 'mjumbe', a community-appointed and trusted point of contact for local households on issues of public services and resource allocation. These individuals represent anywhere from 30-200 households to the government. Working with these vernacular geographies, the team trained local volunteers to survey people

Most notably, the ILO's Global Estimates of Modern Slavery (2017), delivered in collaboration with the Walk Free Initiative and the International Organisation for Migration (IOM), estimated that 40.3 million people were in modern slavery.

The dataset opens a window into not only these workers' lives and circumstances, but also a pressing global problem of exploited labour; one that sits at the intersection of sustainable development, human rights, labour rights, and criminal justice.



in each sub-ward about its features. The process involved 30 team leaders and 163 local participants with hand-held devices who surveyed more than 5,000 respondents across the 443 sub-wards of Dar es Salaam. The survey [see figure 15] sought information on 30 known indicators of child labour, forced labour and forced marriage, and included the question: “I know there are some people in this sub-ward being forced to work against their will” (answers could range from “strongly agree” to “strongly disagree”). By combining the survey data with layers of non-standard data that act as fine-grained proxies for vulnerability to slavery (including telecommunications and transport data), the team built a predictive model that can now visualise any of the city’s blocks through what they term ‘insight tiles.’ They moved beyond national-level estimates to understand what predicts the presence of slavery locally.

Figure 15: Sample questions from Tanzania street surveys (Goulding et al, 2020).

Another anti-slavery measurement innovation, as well as these recent innovations in hyperlocal surveying, is the use of Multiple Systems Estimation (MSE). This approach uses multiple-samples and a 19th-century statistical technique called “capture-recapture.” Silverman et al (2015) used MSE to estimate that the total number of people in conditions of modern slavery in the UK was between 10,000 and 13,000. This analysis was based on six different lists of people reported as experiencing modern slavery, including the UK government’s own National Referral Mechanism (NRM). They used different sets of lists and fit a series of models across them to make the best estimate possible, given the sparse coverage of data across the different sources.

Then in 2019, Silverman et al carried out the same kind of estimations for the US city of New Orleans. This found the estimated total number of enslaved people was somewhere between 650 and 1,600. This used multiple sources, and the probability of victims being captured by one or more lists versus the ratio of the probability of not being captured by these lists. The relative overlap of sources and the ratio of probabilities of appearing in these sources enabled an estimation of the total number of victims (known and unknown victims).

The New Orleans study was one of the first attempts to quantify modern slavery or human trafficking at the city level. Key was the use of de-identified data provided by local organisations. The eight participating agencies represented a multi-disciplinary array of law enforcement, social service providers, housing providers, and legal assistance providers. It demonstrated that a culture of data collection and sharing can be built by anti-trafficking collaboratives over time, and the research team recommended that local agencies should receive training on data collection and employ

data analysts so as to be able to analyse their data, report findings, and use the data for the growth and improvement of victim services.

Understanding the true nature and extent of modern slavery in any given local or national context is fraught with methodological difficulties, stemming from the fundamental problem that victims of modern slavery are a hard-to-find population. But these recent innovations have delivered the most reliable estimates of modern slavery’s scale to date: multi-level modelling based on random sample survey instruments (Walk Free), hyper-local street surveys combined with non-standard data into insight tiles (Goulding) and MSE (Silverman et al).

The Domestic Workers Dataset is part of this recent global effort to gather data that illuminates the nature and scale of slavery, forced labour and exploitation. It combines the large-scale surveying approach of Walk Free (71,000 respondents across 48 countries for the GSI) with the locally-embedded approach of Goulding et al (163 local volunteers conducting surveys in Dar es Salaam) and the emphasis on front-line data of Silverman et al (eight datasets from front-line service providers in New Orleans).

3.4 Data-gathering by front-line organisations

A fourth context for the Domestic Workers Dataset is front-line data-gathering. As was the case with the data collected by the eight front-line organisations in New Orleans, which formed the basis for a city-wide estimate of human trafficking, the data of service-providers offers rich and nuanced information. With analysis and interpretation, it can help to shape policies at local, national and international levels. Data availability and quality remains an issue, especially in developing countries. As the UN has observed, “too many countries still have poor data, data arrives too late and too many issues are still barely covered by existing data” (2014). But the data gathered by front-line organisations may be a route towards a fuller understanding of the scale, nature, causes and consequences of modern slavery and labour exploitation. High quality, localised data can enable the implementation of effective policies and programmes, and the evaluation of their impact.

At the international level, the IOM leads the Counter-Trafficking Data Collaborative (CTDC), which has a collection of data on over 100,000 cases of trafficked people from across the world. This case-level data is provided by front-line organisations on victims of human trafficking who they have identified or assisted. The collection aims to provide capacity for cross-border, inter-agency data analysis and therefore improved evidence for policy and programming.

In national contexts, Polaris is an anti-trafficking charity based in the US that operates the national trafficking helpline. The highly structured helpline call data has information on over 30,000 victims of slavery in the US. It then actively conducts analysis on the data in order to shape state, federal and local laws and policies, including on temporary work visas and trafficking-specific bills. In the UK, the charity Unseen operates a national modern slavery helpline and has helpline data on nearly 6,000 victims of modern slavery. It shares observations from the helpline data with relevant statutory agencies to support their anti-slavery efforts.

In Kenya, the front-line organisation HAART has developed its own capacity on data collection and analysis. Founded in 2010, it is the only organisation in Kenya focused exclusively on counter-trafficking. Its leadership includes Sophie Otiende, a trafficking

In the UK, the charity Unseen operates a national modern slavery helpline and has helpline data on nearly 6,000 victims of modern slavery.

survivor and anti-trafficking advocate. HAART's research reports use qualitative and quantitative methods of data collection and analysis. In one example, it used a cross-sectional survey with 308 respondents for quantitative research and also interviewed 21 individuals and hosted eight focus group discussions. This data-gathering and analysis forms the basis for its policy work at the national level and its programming work locally (including its work running a shelter for girls aged 6-18 years).

The front-line data work of HAART may also be considered a "Small Data" approach. Coined as a term in 2014 by D'Ignazio et al, Small Data is "a practice owned and directed by those who are contributing the data.... The essence of Small Data is that such communities may not just participate in, but can actually initiate and drive such data investigations towards the better understanding of an important local issue." It is "a bottom-up, participatory, grassroots approach to... data collection [that] addresses the key issues of inclusion, accountability, and credibility, by building public participation into the data lifecycle."

Small Data is not necessarily about datasets that are small in size. Its approach can include medium or large datasets, and its data can be quantitative, geographic, qualitative—numbers, maps, interviews, narratives, images and surveys. Small Data also does not imply a lack of rigour or complexity, rather a focus on matching data-gathering and analysis to the interests and needs of a community (including its training needs on data-gathering and analysis). While "Big Data" traditionally implies the use of specialist computer science skills to analyse datasets in ways that exclude the people referenced in the data itself, Small Data takes a participatory approach that makes community members the data-owners and data-users. It empowers individuals and local actors with actionable insights.

Led by survivors, HAART surveys survivors and at-risk communities in order to then feed those insights back into grassroots awareness toolkits, training manuals. Its participatory, "small data" approach means it works closely with survivors, considers what survivors would find useful in terms of data, and tries to understand how community actors themselves can better examine and tackle slavery locally and regionally.

Another example of a Small Data approach to modern slavery and labour exploitation is Apprise, a mobile application created by the United Nations University Institute in Macau and deployed in Thailand in fishing, seafood processing, and sexual exploitation sectors. The app supports communication between frontline responders and vulnerable workers. Workers control their own data collection: they select their preferred language for the interview while answering the questions in privacy and anonymity, and can ask for help to leave their current situation. The app then reports any indications of vulnerability to the frontline responder. It has improved the identification of victims of human trafficking and forced labour, highlighted the full range of migrant workers' experiences, and provided microdata that can inform migration policy.

Small Data can also be sourced via the crowd, leveraging citizen-generated data. The use of citizen scientists—usually non-expert volunteers—to gather and interpret data has dramatically increased in the past five years. The proliferation of citizen science and crowdsourcing platforms make it relatively easier to organise projects, but citizen interests are highly dynamic. Initiatives that are highly topical can generate

high volumes of data but risk quickly losing public interest. The 'non-expert' nature of participants can also risk data quality and accuracy. Data collected via citizen science approaches can be 'noisy' because of the redundancies and gaps arising from human behaviour, and it can be difficult to establish a formalised process to robustly operationalise ad-hoc voluntary inputs.

In the field of modern slavery and labour exploitation, a successful example of citizen science was the "Slavery from Space" project focused on brick kilns led by Boyd et al (2018). This took a citizen science approach and used the online platform Zooniverse. The project asked volunteers to help identify brick kilns—known sites of bonded labour—in satellite imagery of the Brick Belt across India, Bangladesh, Nepal and Pakistan. 120 volunteers took part and the final result was the first rigorous estimate of the number of brick kilns: 55,387.


The work of Sisters to survey domestic workers in India is part of the tradition of front-line data-gathering. Like the helpline data gathered by Polaris and Unseen, the case data gathered by the contributors to the CTDC, and the surveys conducted by HAART, the Domestic Worker Dataset represents data gathered by people offering front-line services—a context of seeking information towards further liberatory action (in the case of Polaris, Unseen and some CTDC contributors) or the provision of recovery services (in the case of HAART and some CTDC contributors). In addition, the quality of the data generated by the Sisters who administered the domestic worker surveys has the potential to be higher still than some of these other examples of front-line data-gathering. Embedded in the communities where they surveyed, providing front-line health, economic and educational services to domestic workers, they may have a trusted status that goes beyond that of the helpline operators at Polaris and Unseen. Their community presence may provide the potential for rich, deep data that offers insights beyond those that community outsiders can gain through interviews or surveys.

A major example of front-line data-gathering, the Domestic Workers Dataset can also be understood as an example of citizen science—a network of volunteers, not data professionals, who gathered data on a scale and over a time period that would not have been achievable by time- and budget-limited visiting researchers. As local, embedded and volunteer researchers, they gathered this data in contexts (the vulnerability and invisibility of domestic workers in private households) that would not have been accessible by outsiders. Though not explicitly designed as such, the surveying initiative may represent a new form of the now 25-year-old practise of citizen science.

A front-line, citizen science dataset, the Domestic Worker Dataset has the potential to be a pioneering example of "Small Data"—participatory data—too. As we noted in this report's section on future survey design and new directions, its findings and on-going design and delivery processes could now be handed back to domestic worker—for example to shape co-design work on data-driven action with the National Domestic Workers' Movement. Then the dataset would be a major example of a truly grassroots, participatory data-impact cycle.

In the field of modern slavery and labour exploitation, a successful example of citizen science was the "Slavery from Space" project focused on brick kilns led by Boyd et al (2018).

Small Data also does not imply a lack of rigour or complexity, rather a focus on matching data-gathering and analysis to the interests and needs of a community (including its training needs on data-gathering and analysis).



Globally, of the 16 million people identified by the ILO as being in forced labour (in the private sector rather than state-sponsored), by far the largest share at 24% were domestic workers (ILO, 2017).

Section four

Recommendations

4.1 Suggestions for future research

There are several possible avenues for new research that would extend our existing analysis of the Domestic Workers Dataset.

Firstly, further work could focus on answering questions that can specifically **inform and help to prioritise the front-line work by the Sisters** and other actors on empowering domestic workers and improving their working conditions. For example:

- a. Additional analysis on wages could determine whether **underpayment** is a factor. Some workers are paid very little, even for the region. Examining the cultural and demographic profiles of these workers may reveal particular vulnerability factors for low wages.
- b. A deeper wage analysis could also examine relationships between official minimum wage legislation for domestic workers and paid wages.
- c. The mean number of hours worked is 6 (SD 3). This may indicate under-employment and requires further investigation. A confirmed finding of under-employment would necessarily inform very different responses from NGOs and other actors than one of long hours.
- d. The variable of “part-time” in the dataset indicates that the respondent is not a live-in domestic worker. Exploitation threat tends to be higher for **live-in domestic workers**. Additional analysis that controls for this variable while examining working hours and wages may provide data to help design a specific programme aimed at the empowerment of live-in domestic workers.
- e. Types of work do not vary significantly between states, but statistically significant differences indicate **some states are potentially higher risk** for poor working conditions than others. Additional contextual information might help to either explain the high daily hours worked in Mizoram or additional analysis may provide information to help target the empowerment programme in particular states.
- f. Under child labour law in India, children under 14 are not allowed to work as domestic workers. The low percentage of children under 14 in the sample is likely because the survey was part of a project aimed at empowering domestic workers via union structures. A higher number (6%) had begun work before the age of 14, and 88% of the instances of child labour recorded in the surveys were in Nagaland.

The data (summarised in the table below) seems to show that Nagaland and Mizoram have a lower average age for domestic workers along with higher hours worked: the mean age of the respondents varied from 36-39 for all states except Mizoram and Nagaland, where the average age of respondents was 22, and the mean age when individuals started work is lower in Mizoram and Nagaland.

Nagaland has a high SD (indicating that there is a substantial range. It may be valuable to focus on these two states for any **additional analysis of child labour**

risks. Additional targeted analyses may be able to reveal indicators for vulnerability to illegal child labour practices, especially in Mizoram and Nagaland. Further analysis may also be able to show a relationship between age and hours worked per day.

	Mean age of respondent (SD)	Mean age at which began working (SD)
Assam	38 (11)	28 (10)
Manipur	39 (11)	29 (10)
Meghalaya	36 (11)	22 (9)
Mizoram	22 (6)	19 (5)
Nagaland	22 (10)	18 (9)
Tripura	39 (11)	29 (10)

Secondly, CDI’s surveying work could be expanded with **supplementary elements**, to **other areas of the world** and/or **combined with additional forms of evidence** in order to expand and enrich the available data on labour exploitation in our world. For example:

- g. The scale of the existing dataset suggests that by adjusting its survey design, CDI and its network of Sisters potentially has the capacity to generate reliable **prevalence estimates** for exploited, forced and child labour in domestic work. The 11,759 individual surveys in this dataset were conducted as part of a baseline study for project impact assessment. The responses represent a convenience sample. A new survey with a clear sampling frame that specifies the population of interest, takes a random sample, and seeks balance in sample numbers between states (so as to minimise the variation in the number of respondents between states) would allow generalisations to be made.
- h. CDI’s unique access to domestic workers via its front-line network of Sisters suggests that it could adjust its survey design to elicit key potential **vulnerability factors** that explain the prevalence of labour exploitation. Additional questions could include details about the proximity of the individuals’ family homes to a school (to understand if distance from school correlates to lack of education), and about the individuals’ proximity to and awareness of community support groups (to understand their current sense of isolation versus connection to a wider community of domestic workers). Survey questions could aim to capture workers’ experiences, perceptions, and understandings of conditions, as well as core demographic and geographic information.
- i. The existence of the FSWA and its 600+ local groups suggests that CDI could now **engage domestic workers themselves** in responding to the dataset and designing a new survey instrument. It could share the findings of this report, reflect on those findings from the perspective of the workers, and discuss what questions a new iteration of the survey should include. Domestic workers themselves may suggest phrasing that can uncover vital insights. By then running the same analysis on the second dataset, drawn from a new round of co-designed surveys, we would be able to identify any differences between the two datasets.

One model for this kind of co-designed data-collection is the work of Walk Free on its analysis of government responses to the problem of modern slavery and forced labour. For its 2019 Measurement, Action, Freedom report, which evaluated the performance of all governments on tackling modern slavery, it ran a number of workshops with survivors of slavery to assess the indicators of the government response framework (for example, specific mechanisms in the areas of criminal

justice, supply chains and victim support). One workshop was in the UK, hosted at the Rights Lab with the Survivor Alliance, a global network of slavery and trafficking survivors. Another was in India, hosted with the survivor leader collective Uttham and the Survivor Alliance. Each two or three-day workshop with survivors reviewed Walk Free’s conceptualisation of a government response and asked survivor leaders what was missing from its current framework.

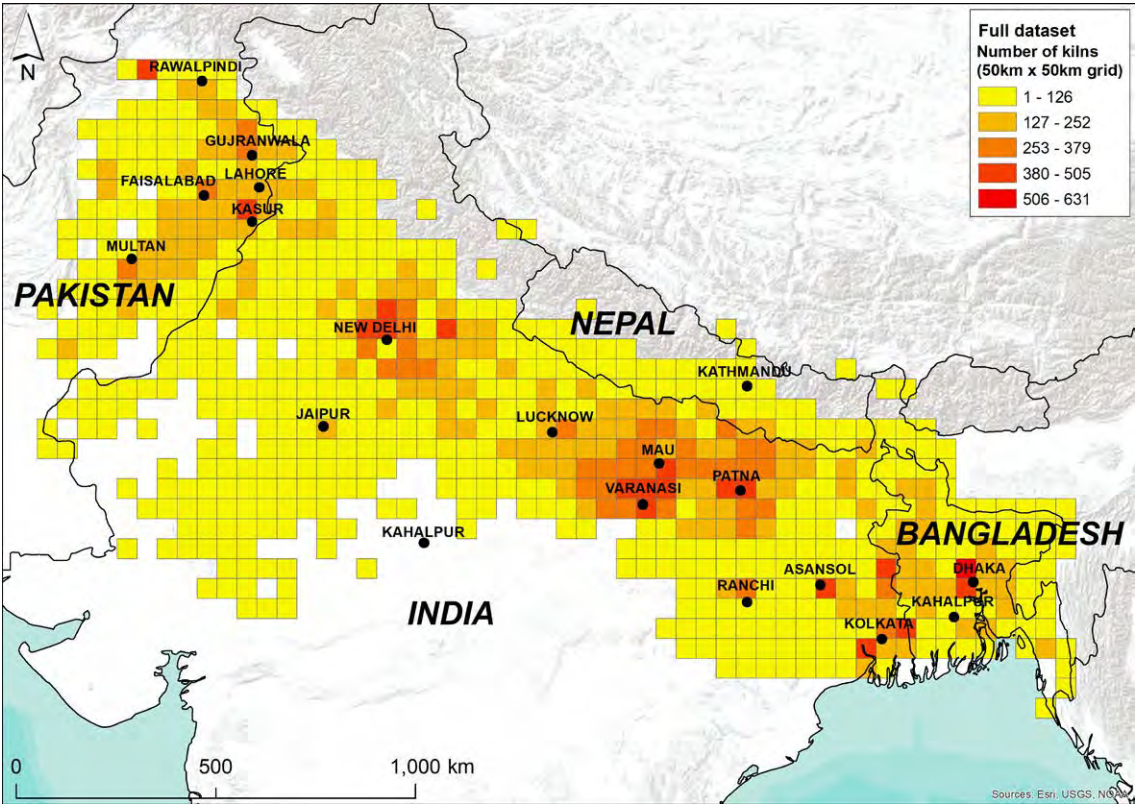
Walk Free incorporated the findings from these workshops into the conceptual framework in order to gather data against the new indicators (Walk Free, 2019). The repetition of the data analysis, on a new dataset of surveys co-designed with domestic workers, would offer an opportunity that is unprecedented, to our knowledge. Part natural experiment and part randomized controlled trial, a comparison of one dataset created without participant input and a second created with that input has the potential to make the strongest case to date for the importance of participatory, grassroots data work in the area of modern slavery and labour exploitation.

- j. The number of survey respondents and the comprehensive nature of the data already gathered is a key demonstration of what is possible. As a data-gathering approach, the deployment of surveys by networks of Sisters could be **expanded to other countries** where similar networks exist. Where else in the world could large networks of Sisters, already embedded in vulnerable communities, gather data to help understand patterns of labour exploitation, forced labour and modern slavery?
- k. The geolocation data from the surveys can potentially provide **complementary evidence** to data being collected through other means, such as the Rights Lab’s Slavery from Space programme. This programme uses satellite remote sensing data that are routinely collected to map the infrastructure associated with slavery. We estimate that we can map over a third of all slavery activity from Space. For example, we have completed a mapping of brick kilns across the Brick Belt of south Asia (where a high proportion of labour is bonded) [see figure 16], and have also mapped brick kilns in Iraq, forced labour in deforestation in Mozambique, child slavery in fish camps in Bangladesh, and slavery in mines in the Democratic Republic of the Congo, among other projects.

A combination of our geospatial data on slavery sites in India with the geolocation data for the Domestic Worker Dataset would build a new, layered picture of labour exploitation hotspots. The addition of earth observation (EO) techniques would mean that patterns within and between states can be detected and compared, and may help to guide the network of Sisters towards working alongside exploited labourers in particular hotspots.

The capacity of the Sisters for data collection suggests that they also are a unique network for potential ‘ground-truthing’ of Slavery from Space data. Remote sensing for EO data can help to fill the data gap in developing countries, but to optimally use the information carried in EO data requires ground data—the verification of what is identified from Space, and the initial rich description of what to look for in satellite imagery in the first place. The areas across which high-risk industries are spread are vast—for example, the areal extent of the ‘Brick Belt’ is 1,551,997 km². The network of Sisters could effectively ground-truth that EO data: it is a network of thousands of people with a ground-truthing capacity that goes way beyond what even the largest in-country NGO could achieve. The Sisters are also uniquely well-positioned as a network of active, embedded volunteers who work discreetly alongside people in some of the most high-risk and vulnerable areas of the world. We would suggest exploring the idea of the Sisters as vital citizen scientists who work alongside EO specialists to verify sites of slavery from Space, in India and other countries.

Figure 16: Map of brick kilns across the Brick Belt of India, Pakistan, Nepal, Bangladesh (Boyd et al, 2020).



As these ideas indicate, we consider this report’s analysis of the Domestic Workers Dataset to be a starting point. The dataset suggests new opportunities for capacity-building on data-gathering for networks of Sisters, including on new survey methods and approaches, in more countries, and potentially on new directions like EO ground-truthing.

There is a strong potential for the Sisters to move from being a community of interest to become a community of intent with regard to data collection—in ways that greatly develop the global knowledge base on labour exploitation and modern slavery.



4.2 Policy recommendations

After sharing this report's data analysis and initial findings with CDI, we invited suggestions from CDI on the policy implications of the data and findings. CDI identified the following eight key recommendations, all of which map onto the report's findings and analysis. The following recommendations are therefore generated by CDI as a local front-line organisation, but in response to—and checked against—the report's analysis and findings. They are included as a co-designed and jointly endorsed set of recommendations by the UK-based team that authored this report (from the Rights Lab) and the India-based team that gathered this report's data (from CDI).

- 1. Ratify ILO Convention 189.** In June 2011, the government, worker and employer delegates of the ILO adopted Convention 189 and Recommendation 201 on Decent Work for Domestic Workers, aimed at improving the working and living conditions of tens of millions of domestic workers worldwide. India is one of 156 countries yet to ratify ILO Convention 189. We recommend that India sign and ratify the Convention and implement Recommendation 201. Doing so would recognise the economic and social value, rights and dignity of domestic workers, and would support calls for action to address the existing exclusions of domestic workers from labour and social protections.
- 2. Registration of domestic workers as workers.** Domestic workers have no legal protection as workers under India's labour laws, which do not recognise domestic work as work. We recommend that all states open their registration provision to domestic workers and provide labour cards.
- 3. Include domestic workers in the schedule of employment.** Of the eight states in North East India, only two (Assam and Tripura) have fixed minimum wages for domestic workers. A just wage would ensure dignity of work. We recommend that every state in North East India include domestic workers in the schedule of employment and fix minimum wages specific to domestic workers.
- 4. Constitute functional welfare boards.** In States where a minimum wage is fixed for domestic workers, we recommend the constitution of a corresponding functional Welfare Board for domestic workers in order to implement the minimum wage in practice.
- 5. Inclusion of domestic workers within social security schemes.** Even without ratifying ILO Convention 189, a country can commit to improve the working and living conditions of domestic workers in light of relevant provisions. We recommend that India establish a provision for the welfare of domestic workers through social security schemes (including pensions, maternity benefits, medical claims, educational support for children).

- 6. Enable grievance and dispute redressal.** Grievance redressal and dispute resolution mechanisms help to protect the rights and welfare of domestic workers. We recommend that all eight states in North East India provide a single mechanism that is accessible to domestic workers for their working needs, including for welfare matters, social protection, social security, protection from abuse, harassment and violence, and that this mechanism include the addressing of grievances against employers or placement agencies, and the settling of disputes (for example through alternate dispute resolution or the facilitation of free legal aid).
- 7. Skills training.** The introduction of skills training and qualifications would help to ensure respect and dignity for domestic work as a profession. We recommend that NGOs develop and promote skills training and that states introduce a sector qualification scheme.
- 8. Stop child labour.** The children engaged in domestic work are often working to support their education. We recommend that India create provisions to ensure access to quality education for all children in India.



References

Agaya, B.M.O. (2013). Report of a Baseline Survey of Women Domestic Workers in Mukuru Informal Settlement, Nairobi Kenya. Nairobi: Oxfam.

Bales, K., Hesketh, O., Silverman, B. (2015). Modern Slavery in the UK: How Many Victims? *Significance* 12.3: 16-21.

Bales, K., Murphy, L.T., Silverman, B. (2019). How Many Trafficked People Are There in Greater New Orleans? Lessons in Measurement. *Journal of Human Trafficking*. <https://doi.org/10.1080/23322705.2019.1634936>.

Boyd, D. et al (2018). Slavery from Space: Demonstrating the Role for Satellite Remote Sensing to Inform Evidence-Based Action Related to UN SDG Number 8. *ISPRS Journal of Photogrammetry and Remote Sensing* 142: 380-388.

Boyd, D. et al (2020). Informing Action for United Nations SDG Target 8.7 and Interdependent SDGs: Slavery from Space. Under Review.

Burnham, L., Theodore, N. (2012). Home Economics: The Invisible and Unregulated World of Domestic Work. New York: National Domestic Workers Alliance.

Catholic Bishop's Conference of India (1980). *A National Socio-Economic Survey of Domestic Workers*. Mangalore: School of Social Work.

CLADHO-IDAY International (2015). Domestic Workers' Baseline Survey Focusing on Child Domestic Workers and Employers in Rwanda. www.invisibleworkers.eu/wp-content/uploads/2016/03/DWs-Baseline-Survey-Final-Report-07-015.pdf.

Dank, M. et al (2019). Prevalence of Minors in Kathmandu's Adult Entertainment Sector. London: Freedom Fund.

D'Ignazio, C. et al (2014). Less is More: The Role of Small Data for Governance in the 21st Century. *Governança Digital*: 115-129.

Gamlin, J., et al (2013). Home Truths: Wellbeing and Vulnerabilities of Child Domestic Workers. London: Anti-Slavery International.

Goulding, J. et al (2020a). SDG Vulnerability Estimation in Tanzania: A Novel Bayesian Nonparametric Approach. In Progress.

Goulding, J. et al (2020b). Theory vs. Machine Learning: Predicting Modern Slavery at Hyper-Local levels in Africa. Under Review.

Gurtoo, A. (2015). Workplace Conditions and Employer Relationships as Predictors of Economic Well-Being: Female Domestic Work in India. *Asian Social Work and Policy Review* 10.1: 61-75.

ILO (2013a). Domestic Workers Across the World: Global and Regional Statistics and the Extent of Legal Protection. Geneva: ILO.

ILO (2013b). Child Domestic Work: Global Estimates 2012. Geneva: ILO.

ILO (2013c). Domestic Workers: Global and Regional Statistics and the Extent of Legal Protection. Geneva: ILO.

ILO and Walk Free (2017). Global Estimates of Modern Slavery. Geneva: ILO.

ILO (2017). Child Labour in India. New Delhi: ILO Decent Work Team for South Asia.

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014). *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. New York: United Nations.

Landman, T. (2020). Measuring Modern Slavery: Law, Human Rights, and New Forms of Data. *Human Rights Quarterly* 42.2: 303-331

Robinson, W.C., Branchini, C. (2018). Estimating Trafficking of Myanmar Women for Forced Marriage and Childbearing in China. Baltimore: Johns Hopkins Bloomberg School of Public Health.

Tariq, H., Saikh, S., Musharaf, M. (2020). Working Conditions and Work-Related Health Issues of Female Domestic Workers in 4 districts of Karachi. *Annals of Work Exposures and Health* 64.4: 378-386.

Thinyane, H., Sassetti, F. (2019). Apprise: Using Sentinel Surveillance for Human Trafficking and Labor Exploitation. Macau: United Nations University.

Tokman, V. (2010). Domestic Workers in Latin America: Statistics for New Policies: *WIEGO Working Paper No. 17*. Cambridge, MA: WIEGO.

Walk Free (2019). Measurement, Action, Freedom: An Independent Assessment of Government Progress Towards Achieving UN Sustainable Development Goal 8.7. Dalkeith: Minderoo.

Zhang, S.X. et al (2014). Estimating Labour Trafficking Among Unauthorized Migrant Workers in San Diego. *Annals of the American Academy of Political and Social Science* 653.1: 65-86.

Appendix: Blank copy of survey

FERRANDO DOMESTIC WORKERS' ALLIANCE (FDWA) Baseline Survey on Domestic Workers of North East India Survey done by Centre for Development Initiatives (CDI)

For official use only

Registration Number:

FDWA/20...../...../.....

Date..... State.....

City.....

I. PERSONAL PROFILE

1. Name..... FIRST NAME..... MIDDLE NAME..... LAST NAME.....
2. Age..... 3. Sex: Male/ Female 4. Phone number: +91-.....
5. Marital Status: Single/Married/Divorced/Widow 6. Caste: SC/ST/OBC/General/Other.....
7. Religion: Hinduism/Islam/ Christianity/Any other.....
8. Present address..... Pin [] [] [] [] [] []
Dist..... State.....
9. Permanent address..... Pin [] [] [] [] [] []
Dist..... State.....

II. FAMILY STATUS

Sl. No	Names	Age	Relationship	Education	Occupation	Income
1.			Father			
2.			Mother			
3.			Husband			
4.						
5.						
6.						
7.						
8.						
9.						
10.						

III. EDUCATIONAL STATUS

10. Have you ever been to school? Yes/No If yes, till which Standard.....
11. Nature of the school: Private/ Govt
12. Who meets/met educational expense? Employers/Self/Guardian
13. Reasons for discontinuing the studies: Poverty/Not interested/Parents death/Any Other.....

IV. OCCUPATIONAL DATA

14. At what age did you start working?..... No. of years in this profession.....
15. In how many places have you worked before?..... Name of the Places.....
16. Reason for leaving/ changing the job..... As Part time/ full time? If part time
17. Since when are you employed in the present job?..... then how many houses.....

V. EMPLOYER

18. Name..... Phone Number.....
19. Address.....
Dist..... State..... Pin [] [] [] [] [] []

VI. TERMS OF EMPLOYMENT

20. Any contract of work made? Yes/ No Oral/ Written.
21. Are you working to: Support parents/working for education/ working livelihood
22. How much salary do you receive?.....
23. When do you get salary? Daily/ Weekly/Monthly/ Quarterly/Yearly
24. Are you being Paid: Regularly/Irregular
25. Is there any deduction from salary? Yes/No. If yes, for leave taken/things broken/others
26. Is there any extra allowance given? Yes/ No. If yes, then how much?.....
27. Are you allowed weekly off: Yes/No.
28. Do you get annually leave permit: Yes/No. If yes then: With pay/ Without pay.

VII. WORKING CONDITION

29. Task done in the houses: Cooking and other kitchen related work\ Washing cleaning house\ child rearing\caring for the elderly\health attendant\ shopping\attending telephones\ gardening\any other (please specify).....
30. Describe your routine:
i) From to.....
ii) From to.....
iii) From to.....
31. Total hours of work per day.....
32. Are you given clothes Yes/ No. If yes, Old/ New/ Both.
33. Medical facilities: Yes/ No
34. Attitude of the employer towards you? Affectionate/ Indifferent/Suspicious/Cruel
35. Any social security scheme applied?..... Year applied.....
36. Any social security scheme received?..... Year received.....

List of Common Social Security Schemes:

- A. BPL
B. RSBY
C. Disability Scheme
D. Educational Assistance
E. Educational Assistance
F. Food Security Scheme
G. Health Assistance
H. Housing Assistance
I. Life Insurance
J. Accident Insurance
K. Pension Scheme
L. Sukanya Samridhi
M. Vocational Training Assistance
N. Widow Pension

Survey done by.....
Signature.....



University of
Nottingham
Rights Lab

Discover more about our world-class research

nottingham.ac.uk/rights-lab



rightslab@nottingham.ac.uk



[@rightsbeacon](https://twitter.com/rightsbeacon)

This publication is available
in alternative formats.
+44 (0)115 951 5559