Cooking Energy Access Survey 2020

Design and data quality

Shaily Jha, Sasmita Patnaik, Carlos F Gould, Sunil Mani, Shalu Agrawal, Johannes Urpelainen and Abhishek Jain

Technical Document | March 2021

Refer to the detailed study based on the dataset:

Are India's Urban Poor Using Clean Cooking Fuels? Insights from Urban Slums in Six States

Shaily Jha, Sasmita Patnaik and Rithima Warrier

Suggested citation: Jha, Shaily*, Sasmita Patnaik*, Carlos F Gould*, Sunil Mani, Shalu Agrawal, Johannes Urpelainen and Abhishek Jain. 2020. Cooking Energy Access Survey 2020: Design and data quality. New Delhi: Council on Energy, Environment and Water.
** Equal first authorship*

Publication team: Alina Sen (CEEW) and Milan Jacob (CEEW).

Organisations: The Council on Energy, Environment and Water (ceew.in) is one of South Asia's leading not-for-profit policy research institutions. The Council uses data, integrated analysis, and strategic outreach to explain and change the use, reuse, and misuse of resources. It prides itself on the independence of its high-quality research, develops partnerships with public and private institutions and engages with the wider public. In 2020, CEEW has once again been featured across nine categories in the 2019 Global Go To Think Tank Index Report. It has also been consistently ranked among the world's top climate change think tanks. Follow us on Twitter @CEEWIndia for the latest updates.

Initiative for Sustainable Energy Policy (ISEP) is an interdisciplinary research program that uses cutting-edge social and behavioral science to design, test, and implement better energy policies in emerging economies. Hosted at the Johns Hopkins School of Advanced International Studies (SAIS), ISEP identifies and pursues opportunities for policy reforms that allow emerging economies to achieve human development at minimal economic and environmental costs. The initiative is based on the premise that the obstacle to energy policy reform is rarely the lack of better alternatives to the current situation, but rather the vexing difficulty of enacting, implementing, and sustaining these alternatives.

Council on Energy, Environment and Water
Sanskrit Bhawan, A-10, Qutab Institutional Area
Aruna Asaf Ali Marg, New Delhi—110067, India

The authors

Shaily Jha | shaily.jha@ceew.in

A Research Analyst in the Energy Access team at The Council, Shaily works on access to clean cooking energy and livelihoods with a focus on gender and social inclusion. Shaily holds a postgraduate degree in Disaster Management from the Tata Institute of Social Sciences (TISS), Mumbai, and an undergraduate degree in Geography from Miranda House, University of Delhi.

Sasmita Patnaik | sasmitaa.patnaik@gmail.com

A former Programme Lead of the Energy Access team at The Council, Sasmita led the work on access to clean cooking and livelihoods, with a focus on gender and social inclusion. She holds an MSc in Development Studies from the School of Oriental and African Studies (SOAS) and a postgraduate diploma in Rural Management from Xavier Institute of Management, Bhubaneswar (XIMB).

Carlos F Gould | gould.cf@gmail.com

As an ISEP student fellow, Carlos is working to understand LPG adoption and usage patterns in India. Carlos Gould studies household energy patterns and energy transitions. His research focuses on the health impacts from traditional solid fuel cooking and the promise of clean fuel alternatives to achieve massive health, economic, and social benefits. He is currently pursuing his PhD in Environmental Health Science Columbia University Mailman School of Public Health.

Sunil Mani | sunil.mani@ceew.in

A Programme Associate at The Council, Sunil's work, for the last four years, has centred around collecting and using primary data to understand and improve the energy access situation at the household and community level in India. He holds a master's degree in economics from Shiv Nadar University, India.

Shalu Agrawal | shalu.agrawal@ceew.in

A Programme Lead at CEEW, Shalu's research focuses on the changing residential energy landscape and exploring the role of institutional, technology, and policy reforms to enable access to affordable, reliable, and sustainable energy for all. She also supports The Council's work on power sector reforms and solar-powered irrigation. Shalu holds an MSc from University College London and a BTech in Electrical Engineering from IIT Roorkee.

Johannes Urpelainen | JohannesU@jhu.edu

Johannes Urpelainen is the Prince Sultan bin Abdulaziz Professor of Energy, Resources and Environment at Johns Hopkins SAIS and the Founding Director of ISEP. He received his Ph.D. in Political Science from the University of Michigan in 2009 and spent the next eight years at Columbia University. He is the award-winning author of four books and over a hundred refereed articles on environmental politics, energy policy, and global governance.

Abhishek Jain | abhsihek.jain@ceew.in
As a Fellow, Abhishek built and leads The Council's practices on energy access, rural livelihoods, and sustainable food systems. He is directing 'Powering Livelihoods', a $3 million initiative. He co-conceptualised and leads CEEW's flagship research on ACCESS (Access to Clean Cooking energy and Electricity—Survey of States). He is an alumnus of the University of Cambridge and IIT Roorkee.

Authors' contribution

Shaily Jha*: Contributed to design, field-testing, and revision of the survey instrument, training of the enumerators, field supervision, led the data monitoring, cleaning, and analysis.

Sasmita Patnaik*: Led the execution of the entire project including survey and questionnaire design, enumerator trainings, and supervised the data monitoring and cleaning.

Carlos F Gould*: Co-conceptualised the project, led the questionnaire design, contributed to data monitoring and cleaning.

Shalu Agrawal: Designed the sampling strategy and contributed to survey design.

Sunil Mani: Contributed to data monitoring and cleaning.

Johannes Urpelainen: Co-conceptualised the project and gave inputs at all stages.

Abhishek Jain: Co-conceptualised the project, contributed to survey design.

** Equal first authorship*

Contents

1. Introduction

The Council on Energy, Environment and Water (CEEW), in collaboration with the Initiative for Sustainable Energy Policy (ISEP), Johns Hopkins University, conducted the Cooking Energy Access Survey in 2019–20. This survey captures detailed information about cooking energy access in households across six states in India: Bihar, Jharkhand, Uttar Pradesh, Rajasthan, Madhya Pradesh, and Chhattisgarh. Cooking Energy Access Survey 2020 is a survey of 2,765 urban slum and rural households across the six states. The survey was conducted with the primary cook of the household.

This survey was conducted along with the India Residential Energy Survey (IRES) 2020[1]. In the six states where the Cooking Energy Access Survey 2020 was conducted, the sampling strategy was similar to IRES 2020. While the rural sample for this survey is the same as that of IRES 2020, the urban sample was selected based on the list of urban slums in the district. We discuss the summary of the survey design and its implementation below:

1.1. Selection of states

The states covered in the survey account for six out of the eight lowest socio-demographic index (SDI) states and exhibit the highest health impacts due to household air pollution. These states also have the highest percentage of households covered under *Pradhan Mantri Ujjwala Yojana* (PMUY) – the flagship clean cooking energy programme of the Government of India. The states of Odisha and Assam have not been considered in the sample because of language barriers. The focus on six Hindi-speaking states helped us focus on one language (and translation) and ensured comparability across the states surveyed.

Table 1: Six out of eight lowest SDI states considered for the Cooking Energy Access Survey 2020

Lowest SDI State	DALY rate per 100 000 population attributable to household air pollution*	Percentage of households covered under PMUY (as of November 2019) ^
Uttar Pradesh	1206	34%
Bihar	1237	35%
Madhya Pradesh	1591	37%
Rajasthan	1752	39%
Jharkhand	1237	40%

[1] IRES 2020, conducted by the Council on Energy, Environment and Water, provides detailed information about the state of energy access and energy-usage patterns in a nationally representative sample of Indian homes. IRES 2020 is a survey of 14,881 urban and rural households spread across 152 districts in 21 large states (by population) of India (Agrawal et al. 2020). For more details, refer to IRES technical documentation.

Chhattisgarh	1926	40%
Odisha	1340	39%
Assam	1495	39%

Note: DALY is disability-adjusted life years

Source: * *India State-Level Disease Burden Initiative Air Pollution Collaborators (Balakrishnan et al. 2019)*

^ *Petroleum Planning and Analysis Cell (PPAC 2019)*

2. Sampling strategy

The survey used a stratified, multi-stage sampling design. While the sampling frame constitutes the list of urban slums in a district and a list of villages in the rural sector as per the 2011 census, the districts were selected based on the IRES 2020 sample frame.

2.1 Allocation of the total sample to states and sampling of districts

In IRES 2020, districts have been assigned to the states in proportion to the state population while ensuring a minimum of 4 and a maximum of 16 districts in any state. This was done to avoid oversampling in large states (such as Uttar Pradesh) and under-sampling in small states (such as Jharkhand). The number of districts (d) allocated to every state was adjusted to a multiple of two to facilitate creating at least two strata from which two districts each could be sampled randomly (Agrawal et al. 2020).

For the creation of strata, all districts within a state were first arranged in descending order of the household population. From this frame, d/2 strata were formed in a way that each stratum had a more or less equal number of households. From each stratum thus created, two districts each were sampled by population proportional to size without replacement (PPSWOR) (Agrawal et al. 2020).

For the Cooking Energy Access Survey 2020, slums and villages have been picked from the same districts sampled under IRES 2020 across the six states. This was done to ensure logistical ease while conducting both surveys, and to share the resources of the national-level survey. While the rural sample for this survey is the same as that of IRES 2020, slums within the selected districts were sampled in the urban areas.

2.2 Allocation of the sample within districts

Within each district, two strata were formed: i) a rural stratum comprising all rural areas of the district and (ii) an urban stratum containing all the urban slums in the district. We allocated a sample of 48 households between urban and rural strata in proportion to the urban and rural household population

in the district, as per the 2011 census. Overall, 1–5 urban slums (u) and 6–10 villages (r) were sampled from each district; from each slum and village, eight and four households were sampled, respectively.

Sampling villages: The list of 2011 census villages in a district constitutes the rural sampling frame. In each sampled district, villages were first arranged in ascending order of the total number of households. From this frame, two sub-strata were formed so that each sub-stratum had a more or less equal population. The required number of sample villages (r/2) was selected with probability proportional to size with replacement (PPSWR) from each substratum. Here, size is the total number of rural households in the village as per the 2011 census, and *r* is the number of villages to be sampled from the district. From each village, four households were randomly sampled. Villages with less than 50 households were excluded from the sampling frame.

Sampling urban slums: The list of urban slums per the 2011 census in the sampled districts under IRES 2020 constitutes the urban sampling frame. In each sampled district under IRES 2020, for the Cooking Energy Access Survey 2020, the required number of slums (u) was selected with PPSWR. Here, size is the total number of households in the slum as per the 2011 census. In some instances, when all the slums within a district were aggregated as per the 2011 census, the survey team selected the slums geographically. From each slum, eight households were randomly sampled. Slums with less than 50 households were excluded from the sampling frame. Table 2 summarises the sample distribution across the six states for the survey. In our sample, approximately 82 per cent of the households are from notified slums.

Table 2: Sample allocation across states and urban/rural sector

State	Districts sampled	Villages sampled	Urban slums sampled	Rural households surveyed	Urban slum households surveyed	Grand Total
Uttar Pradesh	16	130	31	526	249	775
Bihar	14	136	15	540	124	664
Madhya Pradesh	10	89	16	355	128	483
Rajasthan	10	90	12	368	94	462
Jharkhand	4	38	5	152	40	192
Chhattisgarh	4	40	4	157	32	189
	58	**523**	**83**	**2,098**	**667**	**2,765**

Source: Authors' analysis

2.3 Selection of households

From each village and slum sampled for this study, 4 and 8 households were selected for the survey, respectively, using the *random walk procedure*. Here, the survey team was advised to choose a random geographic location (polling booth, municipal school, etc.) in each village/slum and sample every i_{th}

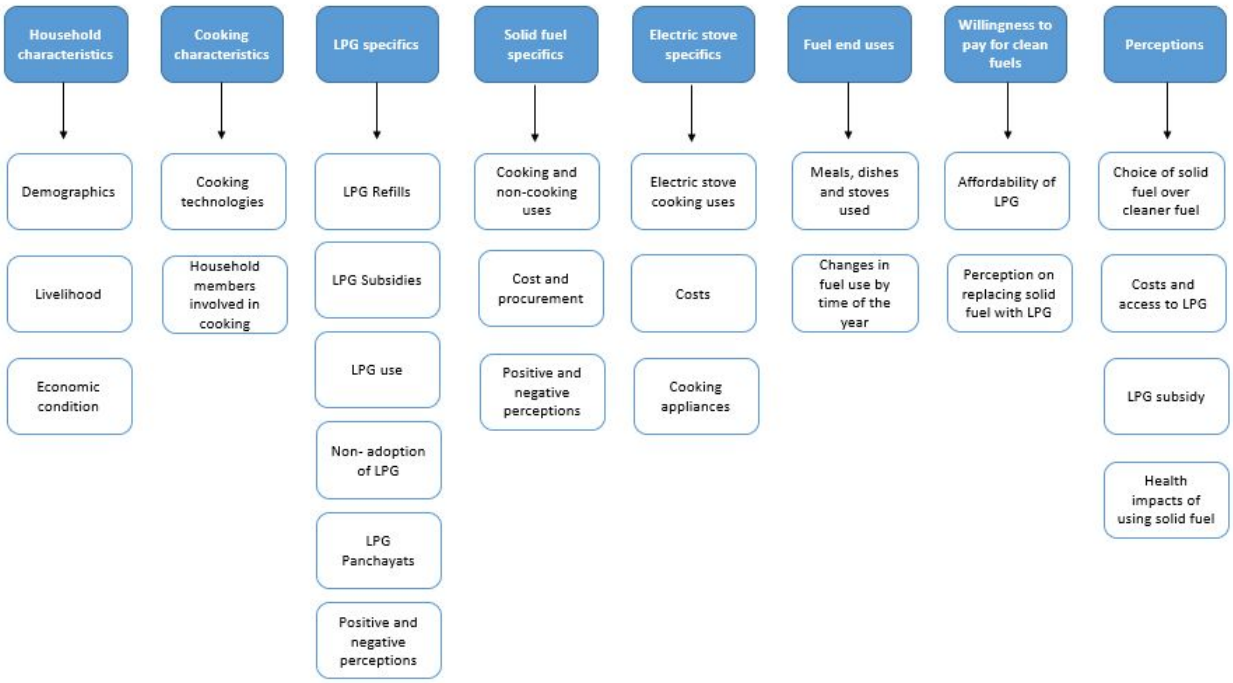
household, following the right-hand rule. The skip pattern, the value of parameter *i*, was five in urban slums and three in villages. We determined the skip parameter based on two main factors: non-response rates and household density in urban and rural areas.

We instructed the enumerators to interview the primary cook of the household. If the primary cook was not available or unwilling to participate, another household was selected following the prescribed skip pattern.

3. Questionnaire design

We designed the survey questionnaire to capture the households' socioeconomic information, livelihood conditions of the primary cook (in most cases women), fuel use pattern, use of LPG and solid fuels, fuel-stacking behaviours, fuel end-uses, and perception of various cooking fuels and their health impacts. The survey questionnaire was designed to be completed in 30–45 minutes (depending on the household's context). The median time taken per interview was 35 minutes. Figure 1 depicts the various parameters considered in the survey.

Figure 1: Cooking Energy Access Survey 2020 questionnaire framework



Source: Authors' compilation

The first draft of the questionnaire was developed after reviewing existing survey instruments, borrowing elements from ACCESS 2018 by ISEP and CEEW. CEEW researchers piloted the second draft of the questionnaire in two areas: villages in Dankaur, Gautam Buddh Nagar in Uttar Pradesh, and Lajpat Nagar of the South East Delhi district, Delhi. The final questionnaire incorporated the inputs from the pilot studies and was translated into Hindi.

The survey questionnaires can be downloaded from <here>.

4. Data collection

The Cooking Energy Access Survey 2020 was conducted between November 2019 and March 2020, with most data collection concentrated between December, January, and February.

4.1 Process

All the surveys were conducted through in-person interviews by trained professional interviewers from Market Xcel Data Matrix Private Limited. All interviewers used handheld tablets for data recording. We used the SurveyToGo application to collect data and ensure better monitoring and control in real-time. It allowed enumerators to conduct interviews in Hindi or English. The questionnaires' digital versions were thoroughly tested through mock interviews during the training sessions and corrected before the survey roll-out.

A team of 85 enumerators (40 per cent of whom were female) were employed to collect the data. The survey training involved a training session for trainers in New Delhi, followed by training sessions for enumerators across three locations: Delhi, Patna, and Bhopal. Each training session lasted for three days and involved classroom training, role-play exercises (on paper and digital questionnaires), and dry runs (mock surveys in the field). The enumerators who did not meet expectations were either re-trained or dropped from the survey team. Each enumerator was given a survey kit, which included a questionnaire with detailed instructions, show cards to be used during select questions, and the authorisation letter from CEEW. Enumerators contacted the sampled household at least three times to maximise the response rate.

4.2 Data quality and limitations

Survey data are vulnerable to multiple errors arising from recall bias, enumerator bias, or measurement error. We took the utmost care to minimise these errors and ensure data quality through various strategies. Adequate checks, skips, and value limits (upper and lower bounds) were built into the data collection software to reduce incorrect, missing, or invalid responses. To select questions, show cards with pictures and response options were used to assist the respondents. Enumerators were trained to code responses framed diversely and avoid leading responses.

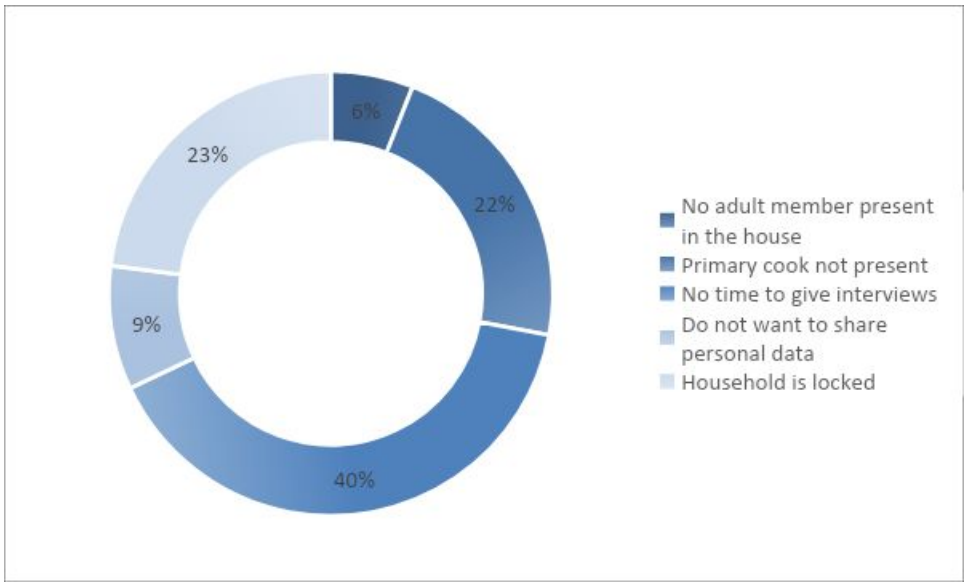
Throughout the data collection process, data quality checks were carried out on small datasets to identify various gaps, such as missing values, incorrect or inconsistent values, and deviation from expected trends or outliers. All cases of incorrect responses were reported to the survey company for cross-verification or re-survey. Many observations were dropped, and re-surveys were conducted where the quality of the data was doubtful. CEEW researchers also visited multiple survey sites for observation purposes while the surveys were in progress. This helped us prescribe timely, corrective measures regarding the interview process and better understand the household responses' context.

Despite these efforts, the possibility of errors in the survey data cannot be rejected. Potential users of the datasets must note the following sources of errors:

1. *Recall bias* - Questions such as monthly household expenditure or monthly household income are particularly vulnerable to recall bias and are sensitive in nature; these were difficult to cross-verify. We leave the decision of data-cleaning/treatment for such questions to the potential users of this dataset.
2. *Language related errors:* The survey was conducted in Hindi. Adequate efforts were made to minimise translation and interpretation errors through questionnaire review as well as pilots. However, given that multiple dialects are used in every state, some questions may not have been adequately administered to some households.
3. *Non-responses* - We observed a non-response rate of 22 per cent, with higher non-response in urban areas (29 per cent) than in rural ones (20 per cent).[2] Figure 2 shows the key reasons behind these non-responses. While most non-responses were due to the sampled house being locked or the required adult member not at home, nearly 40 per cent of non-responses were due to household members unwilling to spend time or share their personal data. The reasons for refusal were similar for both urban and rural households.

[2] If for a targeted sample of s households, r is the number of houses that could not be interviewed, we estimate the non-response rate using the formula r/(r+s). Thus, a non-response rate of 25 per cent implies that a sample of 75 was achieved after non-response from 25 households.

Figure 2: Reasons behind non-responses recorded in Cooking Energy Access Survey



Source: Authors' analysis

5. Survey weights

Cooking Energy Access 2020 used a stratified multi-stage probability sample design. To produce population estimates, we provided design (base) weights for sample households at two levels: district (rural households and urban slum households separately) and state (national). The design weight for each responding household is the number of households in the population that the household represents, estimated as the reciprocal of the probability of selection of that household for the Cooking Energy Access Survey 2020 sample.

However, we did not conduct non-response and post-stratification adjustments to the survey weights because of a lack of adequate information. Box 1 illustrates the detailed procedure for calculating the survey weights.

Box 1: Estimating design weights

We have estimated design (base) weights for each surveyed household to reflect the unequal probabilities of selection in line with the multi-stage sampling strategy discussed in section 1. We first estimate the probability of sampling a given household. We use these probabilities to estimate the design weights, which are reciprocal probability values.

Probability (P) of sampling k^{th} rural HH (p_r) =

P of sampling i^{th} district from m^{th} strata * P of sampling j^{th} village from i^{th} district * P of sampling k^{th} HH from j^{th} village

$= p_d * p_v * p_h$

where

- p_d = No. of districts sampled from m^{th} strata * Total HHs in i^{th} district / Total HHs in m^{th} strata
- p_v = No. of villages sampled from i^{th} district * Total HHs in j^{th} village / Total rural HHs in i^{th} district
- p_h = No. of HHs sampled from j^{th} village (or slum)/ Total HHs in j^{th} village (or slum)

Similarly, we calculate probability of sampling k^{th} urban slum HH (p_u) =

P of sampling i^{th} district from m^{th} strata * P of sampling j^{th} slum from i^{th} district * P of sampling k^{th} HH from j^{th} slum

We also corrected the design weights for under/oversampling of households by multiplying with the ratio of planned versus actual surveys at village (or slum) level.

6. Research ethics and confidentiality of information

The surveys were undertaken after exercising due diligence and obtaining Institutional Review Board approvals. In line with research ethics, enumerators communicated the survey objectives to every respondent, along with the approximate time required and the nature of the questions. In each case, enumerators also took written or verbal consent, depending on the respondent's comfort level.

Any information collected during the survey that might permit the identification of respondents or their households, such as respondent details, address, and phone numbers, is held by the survey company for verification purposes and kept confidential. The data that CEEW and ISEP have from this survey, and would place on public platforms, do not have any information that can be used to identify the household or individual respondents.

References

Agrawal, Shalu, Sunil Mani, Abhishek Jain, Karthik Ganesan and Johannes Urpelainen. 2020. India Residential Energy Survey (IRES) 2020: Design and data quality. New Delhi: Council on Energy, Environment and Water https://www.ceew.in/sites/default/files/CEEW%20-%20IRES%202020%20-%20Design%20and%20data%20quality%20-%204Oct20.pdf#overlay-context=our-story

Balakrishnan, Kalpana, Sagnik Dey, Tarun Gupta, R. S. Dhaliwal, Michael Brauer, Aaron J. Cohen, Jeffrey D. Stanaway, et al. 2019. "The Impact of Air Pollution on Deaths, Disease Burden, and Life Expectancy across the States of India: The Global Burden of Disease Study 2017." *The Lancet Planetary Health* 3 (1): e26–39. https://doi.org/10.1016/S2542-5196(18)30261-4.

Department of Social Justice and Empowerment. 2018. "Social Welfare Statistics."

PPAC. 2019. "Oil Industry Information at a Glance." https://www.ppac.gov.in/WriteReadData/Reports/201912060336500305462ReadyReckonerNov2019WebVersion.pdf. Accessed: 25 July 2020.

COUNCIL ON ENERGY, ENVIRONMENT AND WATER (CEEW)

Sanskrit Bhawan, A-10, Aruna Asaf Ali Marg

Qutab Institutional Area

New Delhi 110 067, India

T: +91 11 4073 3373