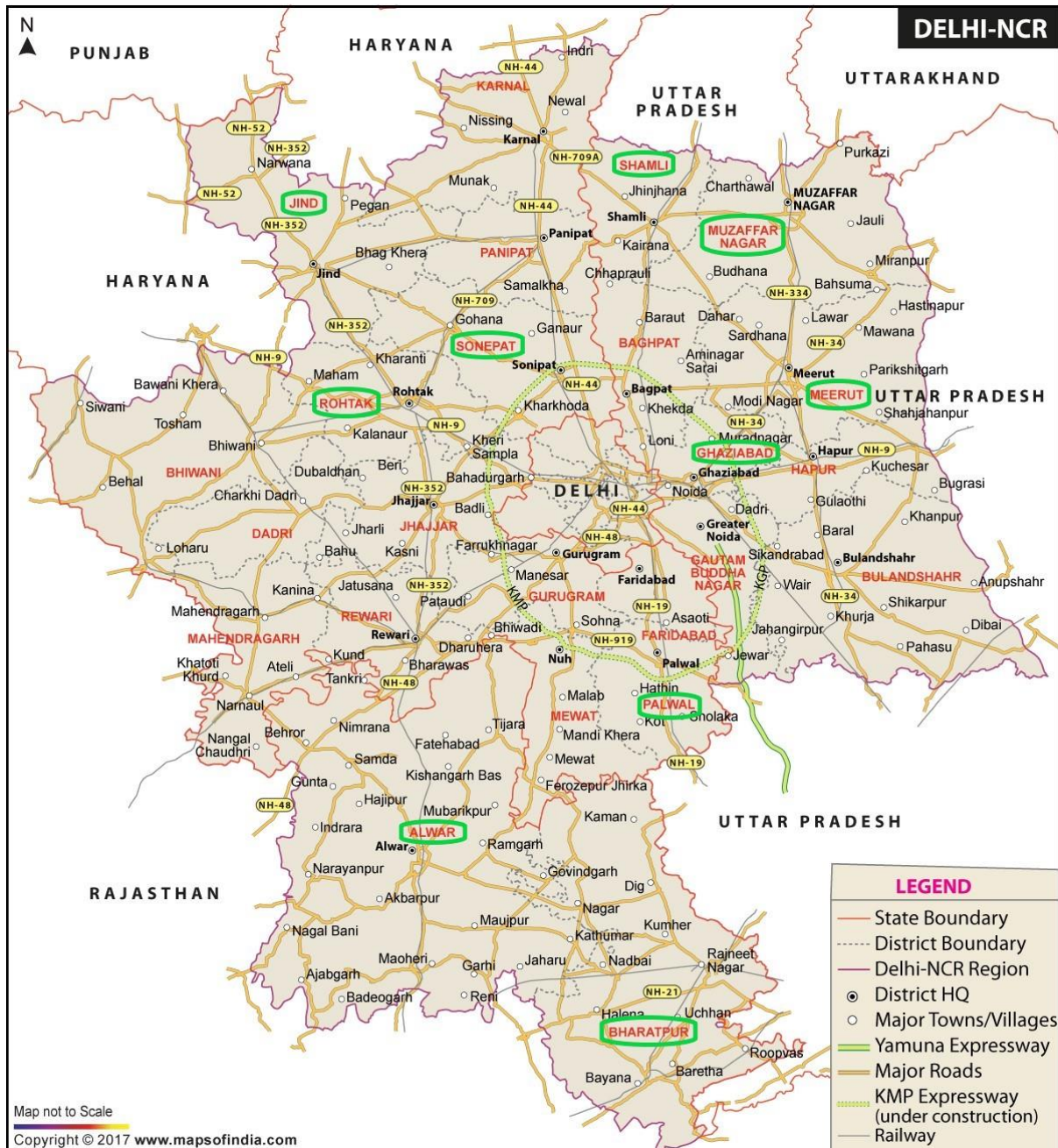# Delhi Metropolitan Area Study (DMAS):

Study Design and Survey Methodology

In a dynamic research and policy environment with a growing demand for data, it is crucial to invest in methods of data collection leading to timely, high-quality, and policy-relevant data. Changing socio-economic conditions and technological innovations necessitate rethinking about the kind of data that are collected and how they are collected, and also about ways in which they are collated and made accessible to users. In this context, the NCAER National Data Innovation Centre has been visualized as a center of innovation and excellence in data collection to strengthen the data ecosystem in India. The Delhi Metropolitan Area Study (DMAS), a flagship study of NCAER, serves as an incubator for experimenting with innovations in data collection on various domains, including household income and expenditure, labor force participation, financial inclusion, health insurance, and healthcare expenditure, gender equality, and empowerment, among others. Two key objectives of DMAS are: 1) Conducting methodological experiments in data collection involving technological innovations and innovations in questionnaire designing, and 2) Demonstrating the feasibility and usefulness of remote monitoring of data collection activities to improve data quality.

The target geographical area for DMAS is the Delhi National Capital Region (NCR), which comprises 31 districts spread over four States, viz., Haryana (13 districts), Delhi (9 districts), Rajasthan (2 districts), and Uttar Pradesh (7 districts) (see Figure 1 for the districts).[1] Although it may not be apparent from the name, Delhi NCR is a highly diverse region, spanning the metropolitan areas of Delhi as well as the rural areas of districts in Haryana, Rajasthan, and Uttar Pradesh. Within a State, we considered a multi-stage stratified cluster sampling design. Districts, clusters, and households were selected in the first, second, and third stages of sampling. Clusters or the secondary sampling units (SSUs) were defined as census villages in rural areas and NSS Urban Frame Survey (UFS) blocks in urban areas (see Table 1 for details of the SSUs and households selected in the sample). The goal of the sampling design was to select a representative random sample during each stage of selection.

[1] NATIONAL CAPITAL REGION PLANNING BOARD 2017. Annual Report, 2016-17. New Delhi: Ministry of Housing and Urban Affairs, Government of India.

Online Appendix Figure 1: Map of Delhi NCR: Selected DMAS districts are circled in green.

Three Delhi districts (North West, East, South) are not visible because of the scaling of the map.

The total number of households that completed the DMAS baseline survey was 5,253. We considered equal allocation of the sample across districts, resulting in approximately 22 SSUs per district, with a few exceptions. Within a district, SSUs were allocated to urban (UFS blocks) and rural areas (villages) in proportion to the percentage of urban and rural households in the district. The median number of

households surveyed from an SSU was 20. An equal number of households were considered from each SSU to manage the interviewer workload efficiently.

Of this baseline sample, 1098 men and 1128 women ages 20-60 were randomly selected for a monthly telephone interview to obtain employment information in Delhi Metropolitan Area Study – Telephone Survey (DMAS-T).

Online Appendix Table 1: Sample Size across Selected Districts in the Delhi NCR: Delhi Metropolitan Area Study (DMAS) Baseline Survey

		Second Stage Units		Households		
State	District	Villages	Urban Blocks	Rural	Urban	Total
Haryana	Palwal	17	5	340	100	440
Haryana	Rohtak	12	10	240	180	420
Haryana	Jind	17	5	340	97	437
Haryana	Sonipat	15	7	300	139	439
Delhi	North West	0	22	0	400	400
Delhi	East	0	22	0	406	406
Delhi	South	0	22	0	427	427
Rajasthan	Alwar	18	5	360	100	460
Rajasthan	Bharatpur	18	4	360	80	440
Uttar Pradesh	Muzaffarnagar	17	7	339	140	479
Uttar Pradesh	Meerut	11	12	221	239	460
Uttar Pradesh	Ghaziabad	7	17	139	306	445
	Total	132	138	2,639	2,614	5,253

The DMAS-T ample has been randomly drawn from the Delhi National Capital Region (Delhi NCR), which comprises 31 districts spread across four states—Delhi, Haryana, Uttar Pradesh, and Rajasthan (see Online Appendix Figure 1). This is a highly diverse region spanning the metropolitan areas of Delhi as well as the rural areas of far-flung districts such as Alwar and Bharatpur in Rajasthan. About half the sample for DMAS-T is located in rural areas, and over 90 percent of the respondents are married. Descriptive statistics are presented in Appendix Table 1. The sample is designed to be representative of the Delhi NCR. The implementation of the lockdown had a severe effect on our study population since all our sample districts were located in either red or orange zones.

The initial sample of 1,098 men and 1,129 women yields 21,917 monthly interviews. Not all individuals were interviewed in each round—sometimes phone contact was not feasible; at times, the respondents were not free to answer survey questions; and at other times they refused to participate in the survey at all.

Attrition, however, does not seem to have had a great impact on sample selection. Appendix Table 1 presents the characteristics of individuals interviewed at three points in time. The first column reflects the background characteristics for individuals interviewed in face-to-face interviews at the baseline, the second column is based on individuals who responded to one or more telephonic surveys, and the third column reveals the characteristics of the individuals interviewed during the lockdown. Since our original goal was to conduct monthly interviews over the course of 12 months and stop thereafter, the respondents who would have rotated out of the sample in April were not interviewed that month. But to capture the effects of the COVID-19 pandemic, we restarted the interviews in May and included the latter respondents in the interviews for May and June. Thus, the sample interviewed during the lockdown period is smaller than the sample interviewed in regular rounds, comprising about 70 percent of the original sample. However, as Online Appendix Table 2 shows, the lockdown sample is similar to the overall sample with respect to most of the socioeconomic background variables.

Online Appendix Table 2: Comparison of Background Characteristics for baseline sample, telephone interview sample and lockdown sample to Estimate Potential for Selection Bias

	Men			Women		
	All Baseline	Any Tel Interview	Lockdown Interview	All Baseline	Any Tel Interview	Lockdown Interview
Age	39.90	40.01	39.87	38.98	38.97	38.94
Unmarried	0.10	0.10	0.09	0.07	0.07	0.08
Completed Education						
No Education	0.12	0.12	0.11	0.39	0.40	0.40
Grades 1-5	0.11	0.11	0.12	0.13	0.13	0.14
Grades 6-8	0.17	0.17	0.18	0.15	0.15	0.15
Grades 9-10	0.25	0.25	0.26	0.11	0.11	0.12
Grades 11-12	0.15	0.15	0.14	0.10	0.10	0.10
Any College	0.20	0.20	0.19	0.12	0.11	0.09
Household Asset Quintile						
Poorest	0.18	0.18	0.18	0.18	0.18	0.18
2nd Quintile	0.20	0.20	0.22	0.20	0.21	0.21
Middle Quintile	0.20	0.20	0.22	0.20	0.20	0.22
4th Quintile	0.21	0.21	0.21	0.21	0.21	0.21
Richest	0.21	0.20	0.18	0.21	0.20	0.18
Caste						
Forward/General/No caste	0.38	0.37	0.39	0.37	0.36	0.37
Other Backward Classes	0.34	0.35	0.34	0.36	0.36	0.35
Schedule Caste/Tribe	0.27	0.28	0.27	0.27	0.27	0.27
Religion						
Hindu, Christian, Sikh etc.	0.89	0.89	0.88	0.89	0.89	0.88
Muslim	0.11	0.11	0.12	0.11	0.11	0.12
Urban residence	1.51	1.50	1.52	1.51	1.49	1.52
State of Residence						
Haryana	0.34	0.34	0.32	0.33	0.33	0.31
Delhi	0.25	0.23	0.26	0.24	0.23	0.26
Rajasthan	0.16	0.16	0.11	0.16	0.17	0.11
Uttar Pradesh	0.26	0.27	0.31	0.27	0.27	0.32
Sample Size (Respondents)	1174	1098	765	1209	1128	786
Sample Size (Monthly records)		10765	951		11152	992