Online Appendix

Temperature, Worker Productivity, and Adaptation: Evidence from Survey Data Production

Melissa LoPalo

A Tables and Figures

Figure A1: Sheet for Evaluating Interviewer Performance

DHS INTERVIEWER PROGRESS SHEET
(USE A SEPARATE SHEET FOR EACH INTERVIEWER)

INTERVIEWER NAME ______________________________

	HOUSEHOLD QUESTIONNAIRES				WOMAN'S QUESTIONNAIRES				MAN'S QUESTIONNAIRES			
	NUMBER COMPLETED		NUMBER NOT COMPLETED		NUMBER COMPLETED		NUMBER NOT COMPLETED		NUMBER COMPLETED		NUMBER NOT COMPLETED	
CLUSTER NUMBER (1)	PER CLUSTER (2)	CUMULATIVE (3)	PER CLUSTER (4)	CUMULATIVE (5)	PER CLUSTER (6)	CUMULATIVE (7)	PER CLUSTER (8)	CUMULATIVE (9)	PER CLUSTER (10)	CUMULATIVE (11)	PER CLUSTER (12)	CUMULATIVE (13)

Note: This figure contains the sheet used in the DHS by supervisors to evaluate interviewer performance throughout the survey round. The sheet is intended to help the supervisor ensure that each member of his or her team is keeping up in terms of interviewing workload.
Source: DHS Program Supervisor's and Editor's Manual, July 2015.

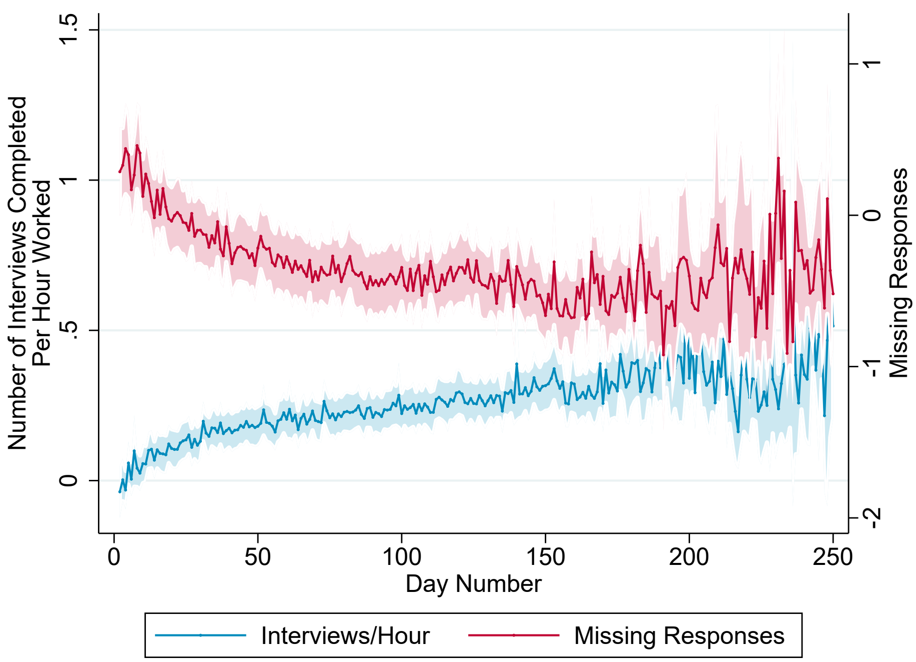
Figure A2: Questionnaire Example: Valid vs. Invalid Skips

SECTION 3. CONTRACEPTION

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
302	CHECK 226: NOT PREGNANT OR UNSURE	PREGNANT	312
303	Are you or your partner currently doing something or using any method to delay or avoid getting pregnant?	YES . . . . 1 NO 2	 312
304 (4)	Which method are you using? RECORD ALL MENTIONED. IF MORE THAN ONE METHOD MENTIONED, FOLLOW SKIP INSTRUCTION FOR HIGHEST METHOD IN LIST.	FEMALE STERILIZATION . . . . . . A MALE STERILIZATION . . . . . . B IUD . . . . C INJECTABLES . . . . D IMPLANTS . . . . E PILL . . . . F CONDOM . . . . G FEMALE CONDOM . . . . H EMERGENCY CONTRACEPTION . . . . I STANDARD DAYS METHOD . . . . . J LACTATIONAL AMENORRHEA METHOD . . . . K RHYTHM METHOD . . . . L WITHDRAWAL . . . . M OTHER MODERN METHOD . . . . . X OTHER TRADITIONAL METHOD . . . . Y	307 (A–B) 309 (C–E) 306 (G) 309 (H–Y)
305	What is the brand name of the pills you are using? IF DON'T KNOW THE BRAND, ASK TO SEE THE PACKAGE.	BRAND A . . . . 01 BRAND B . . . . 02 BRAND C . . . . 03 OTHER ________ 96 (SPECIFY) DON'T KNOW 98	309

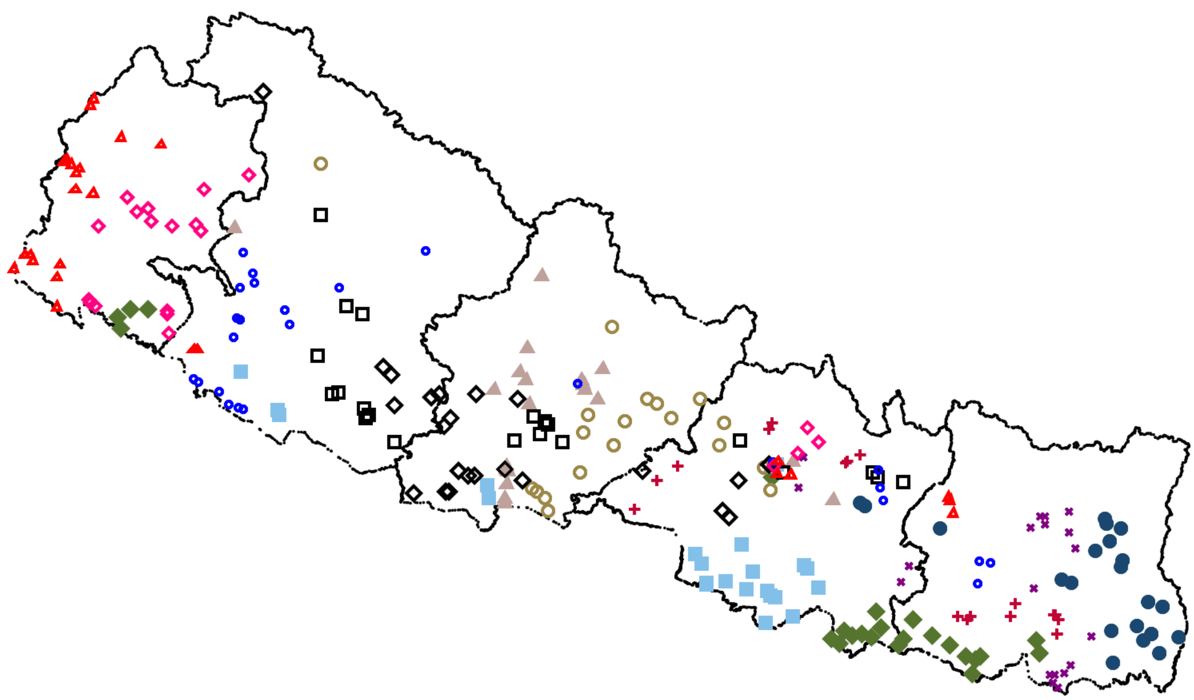
Note: This figure displays an illustrative example of the skip patterns in the DHS Wave 7 Individual Questionnaire. The section begins with instructions to check whether the respondent is pregnant from a previous question. If the respondent is pregnant, then questions 303-311 will be marked as valid skips. If the respondent is not pregnant, then the interviewer should move onto question 303. If the respondent is marked not pregnant or unsure but question 303 is left blank, then question 303 is a missing (or invalid missing) response.
Source: DHS Model Questionnaire - Phase 7, October 2015.

Figure A3: Outcome Variable Validation: Productivity Measures Show Returns to Experience



Note: This figure shows summary statistics on the average of number of interviews completed per hour and counts of missing responses as the survey rounds progress and interviewers gain experience, relative to the first day of the survey round. The horizontal axis is the number of days completed in the survey round (days between the day of interview and the first day in the survey wave). The figure shows the results of regressions that additionally control for region of country by survey round and interviewer fixed effects. The shaded area represents 95% confidence intervals.

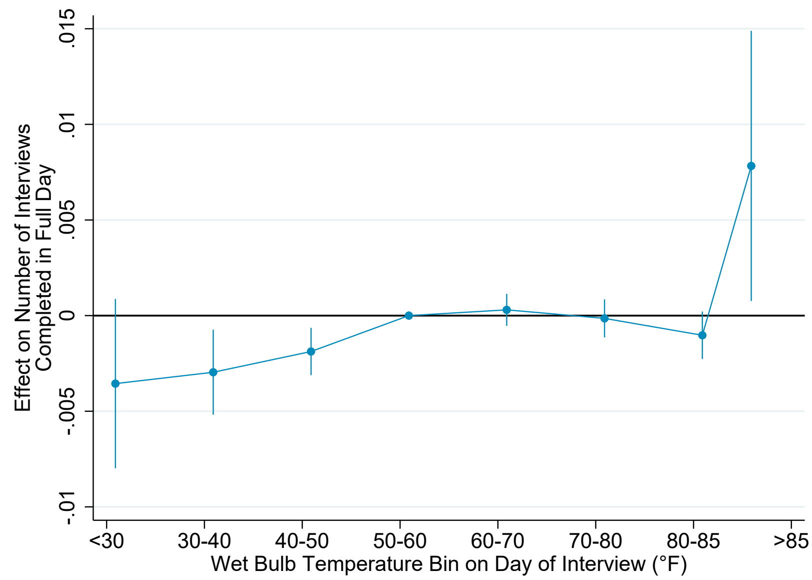
Figure A4: Summary Statistics: Visualization of the 2006 Nepal Survey Round



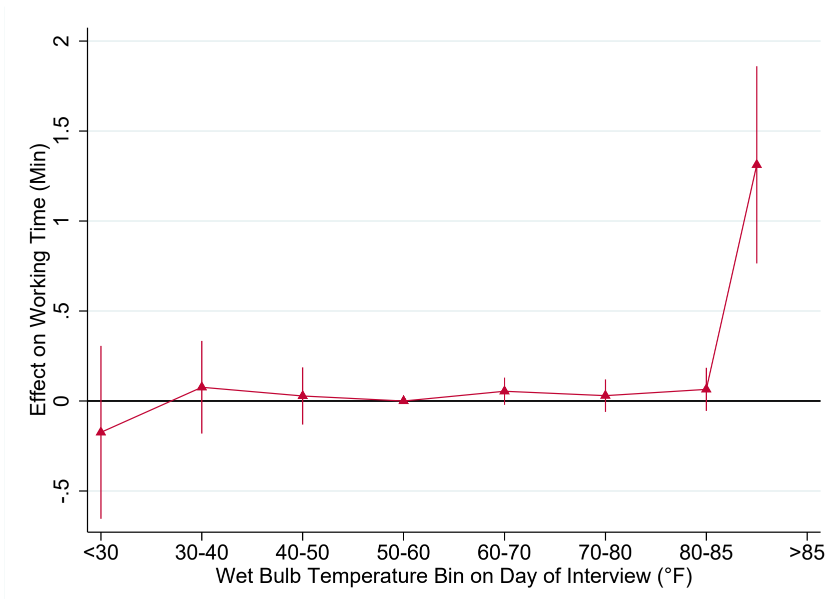
Note: This map gives a visual depiction of the 2006 Nepal survey round. The 5 regions depicted in the map are the regions used as place fixed effects in the analysis. Each color/shape combination indicates a different interviewing team, and the points on the map show where in Nepal the interviews took place.

Figure A5: More Experienced Interviewers Work Longer on Hot Days, Conduct Fewer Interviews on Cold Days (Interaction Effects)

(A) Number of Interviews

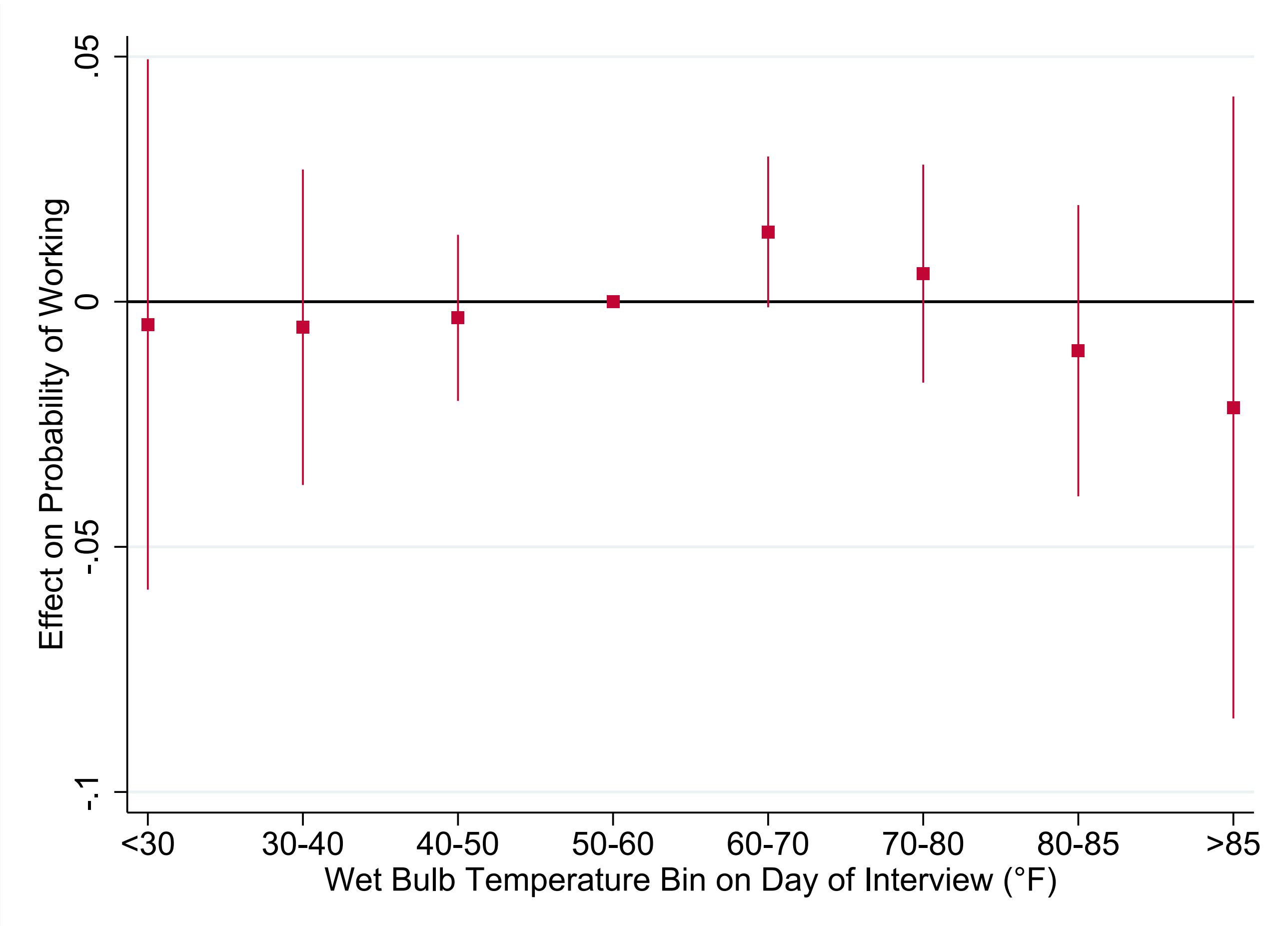


(B) Working Time



Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed in a day (Panel A) and working time (Panel B) as the outcome variables of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. Each wet bulb bin is interacted with a measure of how many days the interviewer has worked on that survey round. The lines displayed show the estimated interaction effect: each coefficient is interpretable as the impact of a day of experience on the effect of wet bulb temperature on the outcome variable. The regressions also include interviewer fixed effects and region of country by survey round fixed effects as well as controls for the characteristics of the set of respondents, daily precipitation in mm, the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview. Point estimates and 95% confidence intervals are shown.

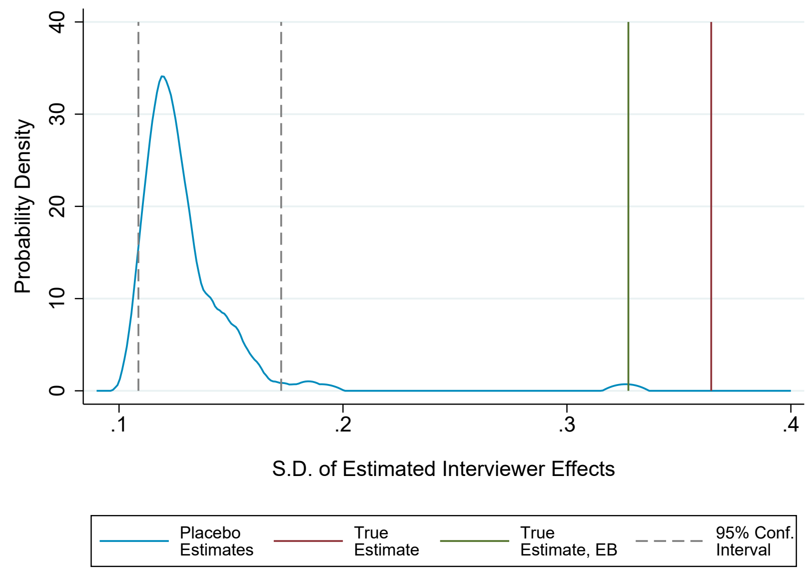
Figure A6: Effect on Labor Supply: No Significant Effect on Probability of Conducting Interviews



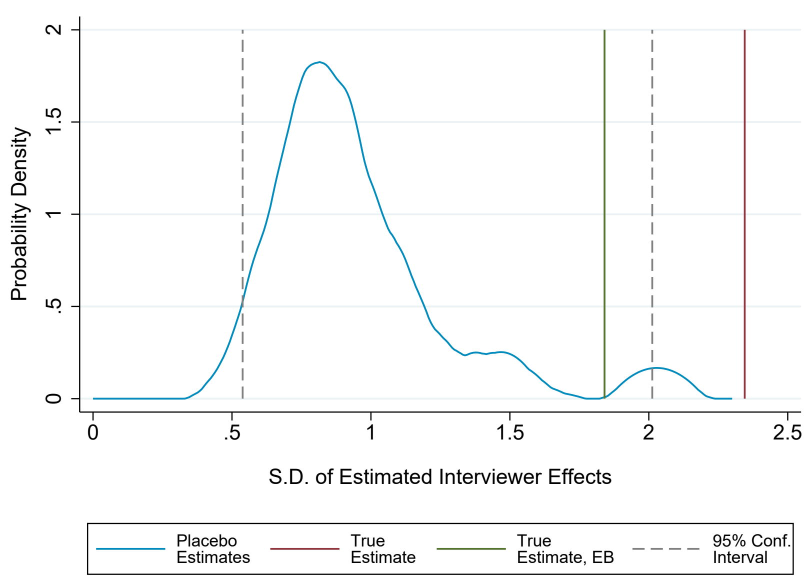
Note: This figure shows the results of an interviewer-day level regression using a dummy variable for whether the interviewer was observed conducting interviews that day as the outcome of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the most recent survey cluster visited by the interviewer on the day of interview fell into the given bin. The regression also includes interviewer fixed effects and fixed effects for the region of country by survey round, as well as controls for daily precipitation in mm, the 10-year average of wet bulb temperature in the survey cluster in the calendar day of interview, the number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview. Point estimates and 95% confidence intervals are shown.

Figure A7: Placebo Tests: Interviewer Fixed Effects Explain Large Portion of Overall Variation in Outcomes

(A) Number of Interviews Completed Per Hour Worked



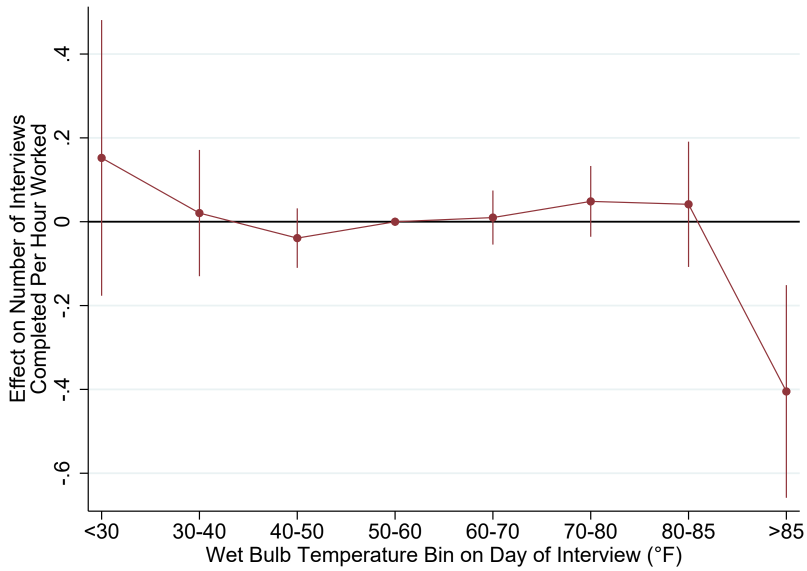
(B) Interaction Effect with Avg. Data Quality



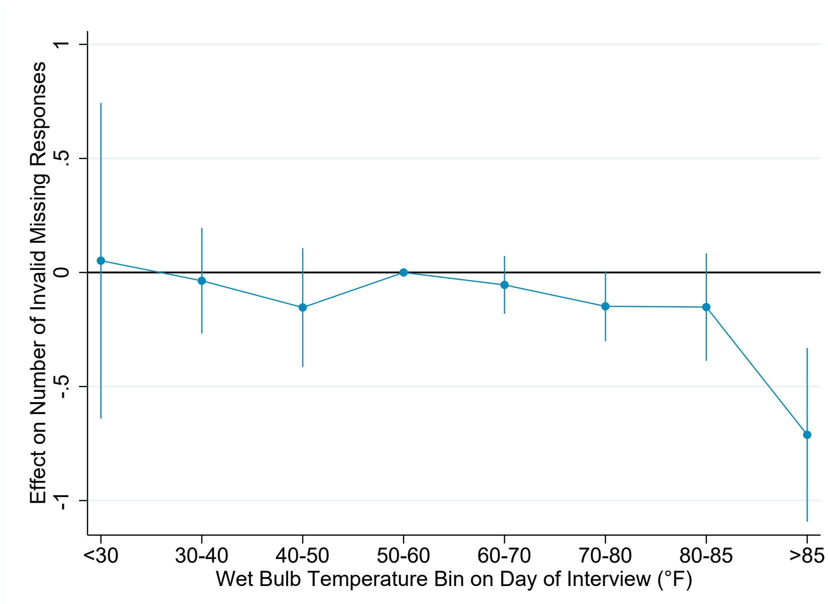
Note: This figure shows the results of a robustness check for Figure 7 Panels A and B, which examines the role of individual interviewer productivity in explaining the dispersion in overall data production outcomes. In this figure, I conduct a placebo exercise, where I randomly re-assign interviewing days (for Panel A) or interviews (for Panel B) to interviewers. The figure shows the results of 100 replications of this placebo exercise. In each replication, I estimate fixed effects for each interviewer with no Empirical Bayes adjustment to be conservative. The distribution of these standard deviations is plotted in blue. The gray lines display a 95% confidence interval. The true value, with no Empirical Bayes adjustment, is shown in the red line, while the true value after the adjustment is shown in green. The outcome variable in Panel A is number of interviews completed per hour worked, and in Panel B it is a count of invalid missing responses.

Figure A8: Split Sample Approach: Effects of Heat Vary by Usual Productivity

(A) Interaction Effect with Avg. Productivity

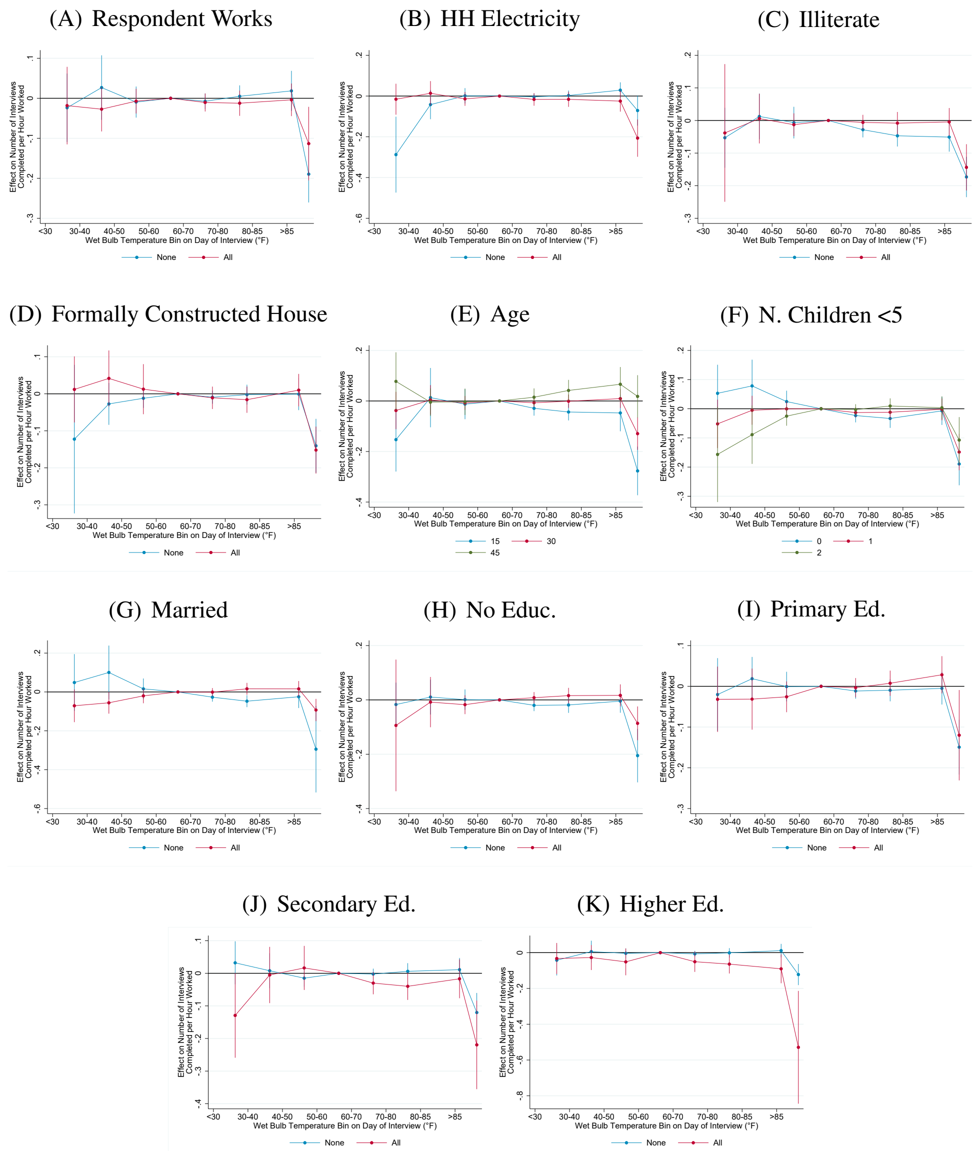


(B) Interaction Effect with Avg. Data Quality



Note: This figure shows the results of a robustness check for Figure 7 Panels C and D. The two figures display estimates from regressions using the number of interviews completed per hour worked (Panel A) and a count of invalid missing responses (Panel B) as the outcome variables of interest. Each panel plots interaction effects between interviewers' average productivity and temperature on the outcome. In this figure, the interviewers' average productivity is estimated using one third of each interviewers' days (Panel A) or interviews (Panel B), chosen at random. The average productivity estimates are then used to estimate the interaction effects in the regression using the other two thirds of the data. The regressions include interviewer fixed effects and fixed effects for the survey round by region of country as well as controls for other characteristics of the set of respondents, daily precipitation in mm, the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview. Point estimates and 95% confidence intervals are shown.

Figure A9: Role of the Respondent: Little Heterogeneity by Respondent Characteristics



Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. Each panel shows interaction effects with the mix of respondents along one observable characteristic. The red line gives the effect of wet bulb temperature if all of the respondents in the interviewer-day have the characteristic, and the blue line gives the effect if none of the respondents have the characteristic, with the exception of age and number of children under the age of 5. These graphs show fitted values for several points in the distribution of the characteristic. The regressions include interviewer fixed effects and fixed effects for the survey round by region of country as well as controls for other characteristics of the set of respondents, daily precipitation in mm, the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview. Point estimates and 95% confidence intervals are shown.

Table A1: Interviewer Productivity Correlated with Probability of Leaving Survey Early

	Left Before Team				Days Between Last Day and Supervisor's Last Day			
Productivity Measure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Invalid Missings	0.002			0.002	0			-0.016
	(0.001)			(0.001)	(0.083)			(0.083)
Number of Interviews		-0.031		0		-3.141		-1.959
		(0.003)		(0.005)		(0.162)		(0.288)
Working Time (Min)			0	-0.001			-0.035	-0.019
			0.000	0.000			(0.002)	(0.004)
Observations	11,088	13,630	13,630	11,088	11,088	13,630	13,630	11,088

Clustered standard errors in parentheses

Note: This table shows the results of regressions showing the impact of average interviewer productivity on the probability of early separation from the survey team. Columns 1-4 examine the impact on the probability that the interviewer is last observed at least one day before the last member of their survey team. Columns 5-8 examine the impact on the number of days between the day that the interviewer is last observed and the day that the last member of their team is last observed. The regressions do not include other controls.

Table A2: Wet Bulb Temperature Predicts Day of Survey Round

	Day of Round			
Wet Bulb Bin (°F)	(1)	(2)	(3)	(4)
>85 Degrees	93.4	74.36	61.76	47.21
	(1.794)	(6.875)	(8.646)	(7.140)
80-85 Degrees	50	24.780	20.180	10.230
	(0.287)	(7.870)	(7.833)	(4.383)
70-80 Degrees	27.82	26.41	23.93	16.49
	(0.156)	(5.379)	(5.251)	(2.979)
60-70 Degrees	16.32	13.350	12.460	8.722
	(0.170)	(3.132)	(2.822)	(2.126)
40-50 Degrees	-7.947	-2.186	-2.484	0.565
	(0.273)	(3.167)	(3.063)	(2.911)
30-40 Degrees	9.66	12.710	10.960	18.120
	(0.545)	(5.913)	(5.347)	(5.699)
<30 Degrees	14.16	36.19	33.46	45.54
	(0.961)	(6.042)	(6.674)	(8.617)
Expected Temp				0.646
				(0.381)
Interviewer FE			X	X
Place FE		X	X	X
Observations	1,104,677	1,104,677	1,104,099	1,104,099

Note: This table shows the results of interview-level regressions using the day of the survey round (the number of days between the interview date and the date of the first interview in the survey round) as the outcome variable. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The first column just gives raw correlations, while the second column adds fixed effects for the survey round by region of country. The third column additionally includes interviewer fixed effects, and fourth column adds a control for the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview. Standard errors are clustered by region-of-country and date of interview in columns 2-4.

Table A3: Mechanisms: More Data Quality Issues Arise on Hot Days

Dependent variable:	Quality Flags	Valid Skips	Invalid Missing	Don't Know/Inconsistent
	(1)	(2)	(3)	(4)
Wet Bulb Bin (°F)				
>85	0.092	8.738	0.221	0.119
	(0.048)	(5.776)	(0.077)	(0.163)
80-85	-0.01	-0.945	0.064	-0.094
	(0.024)	(1.942)	(0.066)	(0.099)
70-80	-0.008	-0.541	-0.035	0.098
	(0.019)	(1.455)	(0.047)	(0.072)
60-70	-0.007	0.517	0.034	0.049
	(0.012)	(0.948)	(0.032)	(0.047)
40-50	0.045	-0.245	-0.046	0.057
	(0.022)	(1.201)	(0.027)	(0.069)
30-40	0.067	1.348	-0.039	0.142
	(0.038)	(2.189)	(0.072)	(0.130)
<30	0.051	1.581	0.066	-0.061
	(0.040)	(3.670)	(0.204)	(0.164)
10-Yr Avg Wet Bulb	-0.001	-0.417	-0.001	0.006
	(0.001)	(0.122)	(0.003)	(0.005)
Day of Survey Round	-0.001	0.049	-0.003	-0.001
	0.000	(0.008)	(0.001)	(0.001)
Daylight Hours	0	0.036	0	-0.002
	0.000	(0.017)	(0.001)	(0.001)
Region of Country FE	X	X	X	X
Interviewer FE	X	X	X	X
Observations	845,868	845,868	845,868	845,868

Standard errors in parentheses

Note: This table shows the results of interview-level regressions using counts of data quality problems as the outcome variables of interest. Column 1 examines total counts of data quality flags as the outcome variable, column 2 uses counts of valid skips (not applicable or not in universe), and column 3 uses counts of invalid missing responses. Column 4 uses a count of responses where the respondent did not know the answer to the question or where the response given was inconsistent with another response. More details on the construction of the outcome variables are available in Appendix B. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. All regressions also include interviewer fixed effects, fixed effects for the survey round by region of country, controls for characteristics of the set of respondents, daily precipitation in mm, the 10-year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number completed days in the survey round. Standard errors are clustered by region-of-country and date of interview.

Table A4: Heterogeneity by World Region

Dependent variable:	Number of Interviews Completed Per Hour Worked					
	Africa	East Asia Pacific	Europe & C. Asia	Latin America	Middle East & N. Africa	South Asia
Wet Bulb Bin (°F)	(1)	(2)	(3)	(4)	(5)	(6)
>85	0.058	-	-	-0.361	-	0.001
	(0.024)			(0.134)		(0.031)
80-85	-0.023	-0.014	-	0.091	-	0.001
	(0.026)	(0.028)		(0.033)		(0.031)
70-80	-0.016	-0.025	0.071	0.031	-0.094	0.046
	(0.019)	(0.023)	(0.100)	(0.031)	(0.037)	(0.013)
60-70	-0.022	0	0.075	-0.012	-0.06	0.016
	(0.011)	0.000	(0.060)	(0.023)	(0.026)	(0.012)
40-50	-0.01	-	-0.012	0.043	-0.027	-0.043
	(0.023)		(0.089)	(0.032)	(0.023)	(0.011)
30-40	-0.069	-	0.035	0.041	-0.087	-0.066
	(0.037)		(0.079)	(0.042)	(0.079)	(0.036)
<30	-	-	-0.045	-0.044	-	-0.127
			(0.096)	(0.064)		(0.080)
10-Yr Avg Wet Bulb	0	0	-0.005	-0.001	0	-0.004
	(0.001)	(0.003)	(0.003)	(0.002)	(0.003)	(0.001)
Day of Survey Round	0.002	0	0.002	0.001	0.006	0.001
	0.000	0.000	(0.001)	0.000	(0.001)	0.000
Daylight Hours	0.000	0.000	0.000	0.001	0.001	0.000
	0.000	0.000	0.000	0.000	(0.001)	0.000
Observations	158,575	37,041	7,395	40,763	19,272	25,616

Standard errors in parentheses

Note: This table shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, run separately for six world regions. Hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include interviewer fixed effects and fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, daily precipitation in mm, the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview. Table B1 shows the list of countries in each region.

Table A5: Interviewer-Specific Outcomes

Dependent variable:	Women in HH Eligible for Interview	Children in HH Eligible for Child Modules	Anthropometry Data Quality Flags
	(1)	(2)	(3)
Wet Bulb Bin (°F)			
>85	0.013	-0.027	-0.005
	(0.049)	(0.089)	(0.005)
80-85	-0.002	0.006	-0.001
	(0.009)	(0.011)	(0.003)
70-80	0.001	0.005	-0.001
	(0.007)	(0.008)	(0.002)
60-70	0	-0.001	-0.003
	(0.004)	(0.005)	(0.001)
40-50	-0.013	-0.005	0.001
	(0.006)	(0.008)	(0.002)
30-40	-0.036	0.001	0.001
	(0.016)	(0.014)	(0.003)
<30	-0.01	-0.019	-0.004
	(0.025)	(0.017)	(0.006)
10-Yr Avg Wet Bulb	0.002	0	0.000
	0.000	(0.001)	(0.000)
Day of Survey Round	0	0	0.000
	0.000	0.000	(0.000)
Daylight Hours	0	0	0.000
	(0.000)	(0.000)	(0.000)
F-Stat: wet bulb bins	1.498	0.431	1.220
Region of Country FE	X	X	X
Interviewer FE	X	X	X
Unit of Observation	Household	Household	Woman
Observations	927,482	927,482	845,868

Standard errors in parentheses

Note: This table shows the results of regressions using outcome variables of interest that are plausibly much more attributable to the interviewer than the respondent. The first two columns show the results of household-level regressions, where the outcome variables of interest are the number of women in the household eligible for individual interviews (col 1), and the number of children in the household eligible for the child health and anthropometry modules (col 2). Both of these variables correspond to interviewing workloads in the household. The third column examines data quality flags in the child and maternal anthropometry modules as the outcome variables. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include interviewer fixed effects and fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, daily precipitation in mm, the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview.

Table A6: Alternative Binning

Binning:	5 degree bins	Daytime bins
	(1)	(2)
>85	-0.135	-0.053
	(0.036)	(0.026)
80-85	0.008	-0.027
	(0.024)	(0.018)
75-80	0	-0.017
	(0.019)	(0.016)
70-75	-0.002	-0.037
	(0.016)	(0.013)
Region of Country FE	X	X
Interviewer FE	X	X
Observations	288,662	288,662

Standard errors in parentheses

Note: This table shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The two columns show regressions that use alternative binning choices for wet bulb temperature relative to Table 2. Column 1 shows the results where wet bulb temperatures are broken into 5-degree bins consistently throughout the distribution. Column 2 also uses 5-degree bins, but rather than using daily average wet bulb temperature to identify exposure to weather, it uses average daytime wet bulb temperature, calculated as the average of wet bulb temperature readings from 9AM through 6PM. The regressions include interviewer fixed effects and fixed effects for the survey round by region of country as well as controls for characteristics of the set of respondents, daily precipitation in mm, the 10 year average of wet bulb temperature in the survey cluster in the calendar day of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered by region-of-country and date of interview.

B Data Appendix

This section details the construction of the weather dataset, as well as several key outcome variables in the data quality analysis.

B.1 Princeton Data

The weather data throughout the analysis come from the Princeton Meteorological Forcing Dataset, which is a reanalysis dataset that has been bias-corrected using observational data. Reanalysis datasets combine observational data from multiple sources (such as satellites, weather balloons, ground stations, etc.) with physics-based weather models that extend the data to observationally sparse geographies. The Princeton Data incorporates reanalysis data from the National Centers for Environmental Prediction and National Center for Atmospheric Research (the NCEP-NCAR Reanalysis Data) with observational data from the Climactic Research Unit (CRU) and the Global Precipitation Climatology Project (GPCP). For more details on the Princeton Data, see Sheffield, Goteti and Wood (2006). I use data on dry bulb temperature, specific humidity, and pressure for 1990-2010 and calculate relative humidity and wet bulb temperature using the following calculations:

1. Relative humidity is calculated as follows, combining standard formulas for the mixing ratio, saturation mixing ratio, and specific humidity from the World Meteorological Organization:

$$rh = 0.263 * p * sh * \left[exp\left(\frac{17.67(t - 273.16)}{t - 29.65} \right) \right]^{-1} \tag{1}$$

   where rh is relative humidity (%), p is pressure (Pa), sh is specific humidity, and t is temperature (K).

2. Then, wet bulb temperature is calculated using the Stull Calculation, which is standard for sea-level pressure and uses temperature in degrees Celsius rather than Kelvin:

$$\begin{aligned} wb = t * [atan(0.151977 * (rh + 8.313658)^{\frac{1}{2}}] + atan(t + rh) - atan(rh - 1.676331) + \\ 0.00391838(rh)^{\frac{3}{2}} * atan(0.023101rh) - 4.686035 \end{aligned} \tag{2}$$

Once wet bulb temperature is calculated, the weather variables are merged with the DHS variables in the manner described in the text: the four surrounding grid points for each DHS survey cluster are located and merged with the weather data, and then the weather variables are calculated as the averages of the four surrounding grid points, weighted by inverse distance between each grid point and the survey cluster.

B.2 Counts of Data Quality Flags

There are three major types of data quality flags used in the analysis, as follows:

1. Imputed Dates: these are important dates where the information has been imputed, either because the full date was not recorded or because the date given was inconsistent with another date (for example, births that are less than 7 months apart).

    - Date of birth of the respondent
    - Date of birth of each child of the respondent and date of death of any children who have passed
    - Date of conception of the current pregnancy
    - Date of start of use of current method of contraception
    - Date of last terminated pregnancy

2. Flagged body measurements: in the DHS, several body measurements are recorded. When they are out of the "acceptable" range of that measurement, they are flagged as such.

    - Child height or weight measurement
    - Women's weight or height

3. Duration variables: the DHS asks for the duration of key reproduction-related activities, such as breastfeeding, and it includes flag variables for inconsistencies in the data processing phase.

    - Duration of breastfeeding
    - Duration of postpartum amenorrhea
    - Duration of postpartum abstinence
    - Time since last menstrual period
    - Time since last sexual intercourse
    - Time since first sexual intercourse

Lastly, I create a flag for cases where the total number of ideal children (v613) is not equal to the ideal number of boys (v627) + the ideal number of girls (v628) + the ideal number of either sex (v629). The outcome variable for count of data quality flags gives the total quantity of these variables that are flagged. Since many respondents have more than one child, such that there is the potential for flags on important dates or measurements for multiple children, this variable ranges in practice from 0 to 23 for women.

B.3 Missing Data

Missing data in the DHS is coded consistently across variables, and variable names are consistent within modules of the survey, allowing me to construct a count of total missing variables in each interview. There are four categories of codes, which are described below. The descriptions include the relevant information from the DHS Recode Manual.

1. Missing data: “This question should have been answered by the respondent, but the questionnaire contained no information for this variable." Depending on the range of codes for the variable, this is coded as 9, 99, 999, or 9999.

2. Respondent did not know the answer: “The respondent replied ‘Don’t know’ to this question." Depending on the range of codes for the variable, this is coded as 8, 98, 998, or 9998. There are exceptions to this rule, which I accommodate to the best of my ability. These include the following:

   - In the contraception module, the questionnaire asks about knowledge and use of a range of methods of contraception. Some of these are country specific. If a method is not mentioned in a certain country’s survey, it is coded as “8." I remove these from the count of “I don’t know" answers and add them to the valid skips.
   - For several body measurement variables, measurements flagged as outside the usual range are coded as “9998" or “99998." I remove these from the count of “I don’t know" answers and include them as data quality flags.

3. Inconsistent answer: “The answer to this question was inconsistent with other responses in the questionnaire and it was thought that this response was probably in error. The response was changed to this code to avoid further problems due to inconsistency of information. This usually takes place during the secondary editing stage of data processing." Depending on the range of codes for the variable, this is coded as 7, 97, 997, 9997. There are exceptions to this rule, including:

   - In the module on basic respondent data, the code 7 often means that the respondent is not a de jure resident of the household. In these cases, I remove the response from the count of inconsistent responses and add it to the count of valid skips.

4. Valid skips: “Variable is not applicable for this respondent either because the question was not asked in a particular country or because the question was not asked of the respondent due to the flow or skip pattern of the questionnaire." This is coded to be missing.

Most survey modules in the DHS are either used by every survey implementing country or are optional ones implemented by a substantial fraction of them. However, many countries also include a multitude of country-specific variables with a variety of names. The count of missing data instances in this paper encompass only standard variables, as defined by the names. The included modules are listed below with the relevant variable names. The variable names are usually built in 2 to 3 components. For three-component variables, the first component indicates the type of interview (v for women, hv or hc for household, mv for men), the second component indicates the module, and the third gives the variable number. For two-component variables, the first component indicates the module and the second the variable number. The two-component variables usually are for questions about children or special modules.

1. Respondent's basic data (v0's and v1's)
2. Reproduction (v2's and b's)
3. Contraceptive use (v3's)
4. Maternity (m1-m73)
5. Maternity and feeding (v4's)
6. Health history (of children under five) (h1-h22)
7. Height and weight (of children under five) (hw's)
8. Marriage (v5's)
9. Fertility preference (v6's)
10. Partner's characteristics and women's work (v71's through v74's)
11. AIDS, STI's, and condom use (v75's through v77's; v82's through v85's)
12. Interview characteristics (v80's and v81's)
13. Maternal mortality (mm's)
14. Malaria (ml's)
15. Domestic violence (d's)
16. Female genital cutting (g's)

Table B1: DHS Survey Rounds in Sample

Country Name	Dates of Fieldwork	Number of Interviewers	Number of Regions	Number of Clusters	Region
Albania	10/2008-4/2009	61	4	450	Europe & Central Asia
Armenia	10/2010-12/2010	73	11	289	Europe & Central Asia
Bangladesh	11/1999-4/2000	94	6	341	South Asia
Bangladesh	1/2004-5/2004	115	6	359	South Asia
Bangladesh	3/2007-8/2007	106	6	361	South Asia
Burkina Faso	12/1992-4/1993	48	5	230	Africa
Burkina Faso	11/1998-3/1999	50	5	208	Africa
Burkina Faso	6/2003-12/2003	77	14	397	Africa
Burkina Faso	5/2010-12/2010	113	13	541	Africa
Benin	6/1996-8/1996	54	6	81	Africa
Benin	8/2001-11/2001	65	6	247	Africa
Bolivia	2/2008-6/2008	260	9	997	Latin America & Caribbean
Burundi	8/2010-12/2010	72	5	338	Africa
Central African Republic	9/1994-3/1995	43	6	231	Africa
Cote d'Ivoire	6/1994-11/1994	48	10	246	Africa
Cote d'Ivoire	9/1998-3/1999	41	3	140	Africa
Cameroon	4/1991-10/1991	37	5	148	Africa
Cameroon	2/2004-9/2004	90	12	464	Africa
Colombia	11/2009-12/2010	91	6	4848	Latin America & Caribbean
Dominican Republic	3/2007-8/2007	163	32	1425	Latin America & Caribbean
Egypt	11/1992-2/1993	60	5	546	Middle East & North Africa
Egypt	11/1995-2/1996	69	6	927	Middle East & North Africa
Egypt	2/2000-5/2000	68	6	976	Middle East & North Africa
Egypt	3/2008-6/2008	57	6	1231	Middle East & North Africa
Egypt	4/2005-7/2005	64	6	1281	Middle East & North Africa
Ethiopia	2/2000-6/2000	229	11	535	Africa
Ethiopia	4/2005-9/2005	177	11	528	Africa
Ethiopia	12/2010-12/2010	174	11	31	Africa
Ghana	10/1993-2/1994	58	10	400	Africa
Ghana	7/2003-11/2003	73	10	410	Africa
Ghana	9/2008-12/2008	110	10	404	Africa
Ghana	11/1998-2/1999	69	10	400	Africa
Guinea	4/1999-8/1999	54	5	293	Africa
Guinea	2/2005-6/2005	51	8	291	Africa
Guyana	3/2009-8/2009	99	10	312	Latin America & Caribbean
Haiti	2/2000-7/2000	65	10	316	Latin America & Caribbean
Haiti	10/2005-5/2006	65	10	332	Latin America & Caribbean
Indonesia	10/2002-4/2003	375	26	1296	East Asia Pacific
Jordan	7/2002-10/2002	70	3	495	Middle East & North Africa
Jordan	6/2007-10/2007	56	3	923	Middle East & North Africa
Kenya	4/2003-9/2003	97	8	399	Africa
Kenya	11/2008-3/2009	123	8	397	Africa

Cambodia	1/2000-7/2000	94	23	470	East Asia Pacific
Cambodia	9/2005-3/2006	92	19	548	East Asia Pacific
Cambodia	7/2010-12/2010	79	19	589	East Asia Pacific
Liberia	12/2006-4/2007	99	6	291	Africa
Lesotho	9/2004-2/2005	58	10	381	Africa
Lesotho	10/2009-2/2010	77	10	395	Africa
Morocco	10/2003-2/2004	58	15	480	Middle East & North Africa
Moldova	6/2005-8/2005	76	4	399	Europe & Central Asia
Madagascar	8/1997-12/1997	45	6	268	Africa
Madagascar	11/2008-7/2009	123	22	585	Africa
Mali	11/1995-5/1996	59	8	299	Africa
Mali	1/2001-6/2001	146	9	399	Africa
Mali	4/2006-12/2006	198	9	405	Africa
Malawi	7/2000-11/2000	167	3	559	Africa
Malawi	1/2004-2/2005	130	3	520	Africa
Malawi	6/2010-10/2010	301	3	827	Africa
Nigeria	4/1990-12/1990	132	4	297	Africa
Nigeria	3/2003-8/2003	77	6	360	Africa
Nigeria	6/2008-11/2008	253	6	886	Africa
Niger	3/1992-6/1992	41	8	235	Africa
Niger	1/1998-7/1998	64	6	268	Africa
Namibia	9/2000-12/2000	109	13	259	Africa
Namibia	11/2006-4/2007	193	13	491	Africa
Nepal	1/2001-7/2001	74	5	251	South Asia
Nepal	2/2006-8/2006	87	5	260	South Asia
Peru	7/2000-11/2000	210	24	1407	Latin America & Caribbean
Philippines	6/2003-9/2003	303	17	815	East Asia Pacific
Philippines	8/2008-9/2008	276	17	786	East Asia Pacific
Pakistan	9/2006-3/2007	130	4	955	South Asia
Rwanda	2/2005-8/2005	89	12	456	Africa
Rwanda	9/2010-12/2010	111	5	279	Africa
Sierra Leone	4/2008-6/2008	131	4	349	Africa
Senegal	11/1992-8/1993	42	4	258	Africa
Senegal	1/1997-5/1997	72	4	319	Africa
Senegal	1/2005-6/2005	77	11	366	Africa
Senegal	10/2010-12/2010	64	11	161	Africa
Swaziland	7/2006-3/2007	80	4	270	Africa
Togo	2/1998-5/1998	64	6	287	Africa
Timor-Leste	8/2009-2/2010	72	13	454	East Asia Pacific
Tanzania	9/1999-12/1999	78	22	173	Africa
Tanzania	12/2009-5/2010	70	26	458	Africa
Uganda	9/2000-3/2001	72	4	266	Africa
Uganda	5/2006-10/2006	87	9	336	Africa
Zambia	4/2007-10/2007	105	9	319	Africa
Zimbabwe	8/1999-12/1999	83	10	221	Africa
Zimbabwe	8/2005-4/2006	134	10	396	Africa
Zimbabwe	9/2010-12/2010	72	10	219	Africa

Note: This table gives descriptions of each of the 90 survey rounds used in the main analysis of the paper. Each row of the table represents one survey round; some countries have multiple. The second column gives the start and end date of the interviews. Column 3 gives the number of unique interviewers observed in the survey round, column 4 gives the number of regions of that country used in the round (used as fixed effects in the main analysis), and column 5 gives the number of clusters where interviews were conducted.

C Theoretical Framework

In this section I present a simple framework for the interviewer's endogenous choice of effort allocation on days with varying temperatures. I closely follow the model set forth by Graff Zivin and Neidell (2012) to describe worker responses to air pollution in assuming that output for an interviewer is a function of effort e and temperature Ω. Here, a large value of Ω can be thought of a more extreme temperature. The interviewer's output has two components: data quality (q) and data quantity (y), and the interviewer chooses her effort allocation for both components (e_q and e_y, respectively). The interviewer earns a fixed wage $\bar{w}$ for as long as she works on the survey round and has a continuation value of K associated with keeping her job (not being fired for poor performance). The worker's probability of job retention depends linearly on output, for simplicity. The probability of job retention depends on both the quality and quantity of the data produced; each dimension is weighted by α and $1 - \alpha$ in the job retention decision, respectively.[1]

Interviewers choose effort levels e_y and e_q in order to maximize the following:

$$\max_{e_y, e_q} \bar{w} + (\alpha y(e_y) + (1-\alpha)q(e_q))K - c(e_y, e_q, \Omega) \tag{3}$$

The first order conditions are the following:

$$\alpha \frac{\partial y}{\partial e_y} K = \frac{\partial c}{\partial e_y} \tag{4}$$

$$(1-\alpha)\frac{\partial q}{\partial e_q} K = \frac{\partial c}{\partial e_q} \tag{5}$$

Taking a total derivative with respect to Ω yields:

$$\alpha K (\frac{\partial^2 y}{\partial {e_y}^2}) \frac{\partial e_y}{\partial \Omega} = \frac{\partial^2 c}{\partial {e_y}^2} \frac{\partial e_y}{\partial \Omega} + \frac{\partial c}{\partial e_y \partial \Omega} \tag{6}$$

$$(1-\alpha) K (\frac{\partial^2 q}{\partial {e_q}^2}) \frac{\partial e_q}{\partial \Omega} = \frac{\partial^c}{\partial {e_q}^2} \frac{\partial e_q}{\partial \Omega} + \frac{\partial c}{\partial e_q \partial \Omega} \tag{7}$$

Solving for $\frac{\partial e_q}{\partial \Omega}$ and $\frac{\partial e_y}{\partial \Omega}$, respectively, yields the following relationships:

[1] These probabilities are assumed to differ because of differences in supervisor monitoring of interviewer performance on data quality vs. quantity. In practice, this may in turn reflect the strength of expectations for the supervisor to keep the team on schedule. If the DHS expects some risk of weather causing teams to slow down, for example, this may cause α to be lower than otherwise, or even to vary with temperature.

$$\frac{\partial e_y}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_y \partial \Omega}}{\alpha K \frac{\partial^2 y}{\partial {e_y}^2} - \frac{\partial^2 c}{\partial {e_y}^2}} \tag{8}$$

$$\frac{\partial e_q}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_q \partial \Omega}}{(1-\alpha) K \frac{\partial^2 q}{\partial {e_q}^2} - \frac{\partial^2 c}{\partial {e_q}^2}} \tag{9}$$

Assuming that the cost of effort function is convex, production is concave in effort, and that extreme temperature increases the marginal cost of effort, the effect of temperature on effort allocation is negative for both tasks. This model also predicts that the effect of temperature on effort allocation is decreasing in the weight put on the dimension of productivity in the probability of job retention. Similarly, it is decreasing in the continuation value of the job, K. The model therefore predicts that the negative effect of temperature on effort will be stronger on the less observable dimension of productivity, quality.

C.1 Alternative Assumptions

In the previous section, the only way in which temperature (Ω) entered the interviewer's decision was through disutility of effort. It may also be reasonable to assume that temperature also has a direct effect on productivity, through some unavoidable physiological effect. The maximization problem could be changed to reflect this possibility in the following way:

$$\max_{e_y, e_q} \bar{w} + (\alpha y(e_y, \Omega) + (1-\alpha) q(e_q, \Omega) K - c(e_y, e_q, \Omega) \tag{10}$$

The first order conditions remain the same, but taking the derivative with respect to temperature now yields:

$$\alpha K (\frac{\partial^2 y}{\partial {e_y}^2}) \frac{\partial e_y}{\partial \Omega} + \alpha \frac{\partial y}{\partial e_y \partial \Omega} K = \frac{\partial^2 c}{\partial {e_y}^2} \frac{\partial e_y}{\partial \Omega} + \frac{\partial c}{\partial e_y \partial \Omega}$$

$$(1-\alpha) K (\frac{\partial^2 q}{\partial {e_q}^2}) \frac{\partial e_q}{\partial \Omega} + \alpha \frac{\partial q}{\partial e_q \partial \Omega} K = \frac{\partial^c}{\partial {e_q}^2} \frac{\partial e_q}{\partial \Omega} + \frac{\partial c}{\partial e_q \partial \Omega}$$

The resulting expression for the effect of temperature on effort allocation is as follows:

$$\frac{\partial e_y}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_y \partial \Omega} - \alpha K \frac{\partial y}{\partial e_y \partial \Omega}}{\alpha K \frac{\partial^2 y}{\partial {e_y}^2} - \frac{\partial^2 c}{\partial {e_y}^2}}$$

$$\frac{\partial e_q}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_q \partial \Omega} - (1-\alpha) K \frac{\partial q}{\partial e_q \partial \Omega}}{(1-\alpha) K \frac{\partial^2 q}{\partial {e_q}^2} - \frac{\partial^2 c}{\partial {e_q}^2}}$$

The overall effects of temperature on effort allocation on both tasks remains negative, as long as temperature is assumed to decrease the marginal productivity of effort. The weight put on each task in the job retention probability function now appears both in the numerator and the denominator, however. The effect of an increase in the weight is to increase the size of both the numerator and the denominator in absolute value. Intuitively, temperature now decreases the returns to effort, and if this effect is large enough the differential effect on separate tasks may be swamped. Therefore, it is ultimately an empirical question whether temperature will have larger effects on the less observable dimension.

D Empirical Bayes Adjustment

This appendix details the implementation of Empirical Bayes techniques in adjusting the estimates of interviewer fixed effects in Panels A and B of Figure 7. In the implementation of the procedure, I use the STATA program provided by Adam Sacarny.[2] Therefore, my implementation very closely follows the procedure in Chandra et al. (2016). The implementation proceeds as follows. Bold letters denote vectors, while non-bold letters denote scalars.:

The first step is to estimate the first-stage regression, which corresponds to Equation 2:

$$y_{icprd} = \lambda_i + \theta_p + \nu X_{rcpd} + \rho Daylight_{cpd} + \gamma DayinRound_{pd} + \epsilon_{icprd}$$

This equation yields an estimate of $\hat{\lambda}_i$ for each interviewer, netting out variation in the outcome variables due to place-specific and observable respondent-specific factors. $\hat{\lambda}_i$ is equal to the interviewer's true average productivity plus the error term, ϵ_i:

$$\hat{\lambda}_i = \lambda_i + \epsilon_i$$

If we assume that each $\hat{\lambda}_i$ is independently and normally distributed across interviewers with variance $\sigma^2_{\epsilon,i}$, then the distribution of of sample estimates is:

$$\hat{\lambda}_i | \lambda_i, \sigma^2_{\epsilon,i} \sim N(\lambda_i, \sigma^2_{\epsilon,i})$$

If we also assume that λ_i is normally distributed with a known average, $\mathbf{x_i}'\boldsymbol{\pi}$, a linear function of the vector of covariates included in Equation 2 (x_i), and variance σ^2_a, then given a prior distribution for λ_i, the posterior

[2] The program can be found at http://sacarny.com/programs/.

distribution distribution is as follows:

$$\lambda_i|\hat{\lambda}_i, \mathbf{x_i}, \boldsymbol{\pi}, \sigma_a^2, \sigma_{\epsilon,i}^2 \sim N(\lambda_i^{EB}, \sigma_{\epsilon,i}^2(1-b_i))$$

In this equation, λ_i^{EB} is the interviewer average productivity after the Empirical Bayes adjustment, which incorporates a shrinkage factor b_i:

$$\lambda_i^{EB} = (1-b_i)\hat{\lambda}_i + b_i\mathbf{x_i}'\boldsymbol{\pi}$$

where

$$b_i = \sigma_{\epsilon,i}^2/(\sigma_{\epsilon,i}^2 + \sigma_a^2)$$

To implement the procedure, sample estimates of $\sigma_{\epsilon,i}^2$, σ_a^2, and $\boldsymbol{\pi}$ are needed. Following Chandra et al. (2016), I estimate $\sigma_{\epsilon,i}^2$ by squaring the estimated standard error for the interviewer fixed effect, assuming homescedastic errors in Equation 2. I estimate $\boldsymbol{\pi}$ using a weighted least squares regression of the interviewer fixed effect estimates on the regressors from the first stage regression. The weights in the regression for each interviewer are $\frac{1}{\hat{\sigma}_{\epsilon,i}^2+\hat{\sigma}_a^2}$. Finally, the sample estimate for σ_a^2 is as follows:

$$\hat{\sigma}_a^2 = \frac{\sum_i w_i(\frac{n_i}{n_i-n_x})(\hat{\lambda}_i - \mathbf{x_i}'\boldsymbol{\pi})^2 - \hat{\sigma}_{\epsilon,i}^2}{\sum_i w_i}$$

where n_i is the number of interviewers in the sample, and n_x is the number of regressors in Equation 2. This parameter is estimated simultaneously with $\boldsymbol{\pi}$, in the iterative procedure described by Chandra et al. (2016).

Given these estimates, the empirical estimates of the Empirical Bayes-adjusted interviewer fixed effects are:

$$\hat{\lambda}_i^{EB} = (1-\hat{b}_i)\hat{\lambda}_i + \hat{b}_i\mathbf{x_i}'\hat{\boldsymbol{\pi}}$$

where the estimate of the shrinkage factor $\hat{b_i}$ with the relevant degrees of freedom adjustment is:

$$\hat{b_i} = (\frac{n_i - n_x - 2}{n_i - n_x})(\frac{\hat{\sigma}_{\epsilon,i}^2}{\hat{\sigma}_{\epsilon,i}^2 + \hat{\sigma}_a^2})$$

References

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health care exceptionalism? Performance and allocation in the US health care sector." *American Economic Review*, 106(8): 2110–44.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The impact of pollution on worker productivity." *American Economic Review*, 102(7): 3652–73.

Sheffield, Justin, Gopi Goteti, and Eric F. Wood. 2006. "Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling." *Journal of Climate*, 19: 3088–3111.