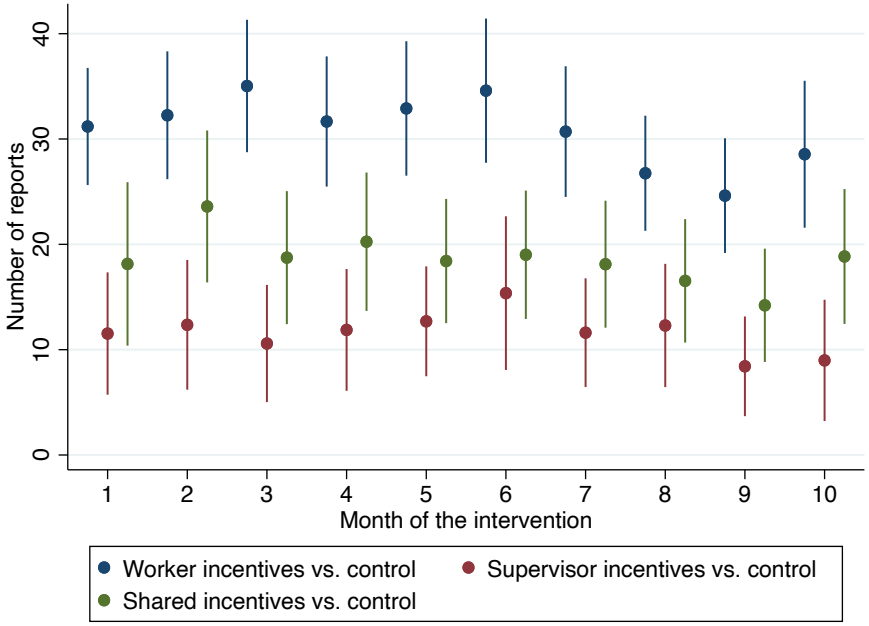
# Online Appendix (For Online Publication Only)

# A Appendix Figures and Tables

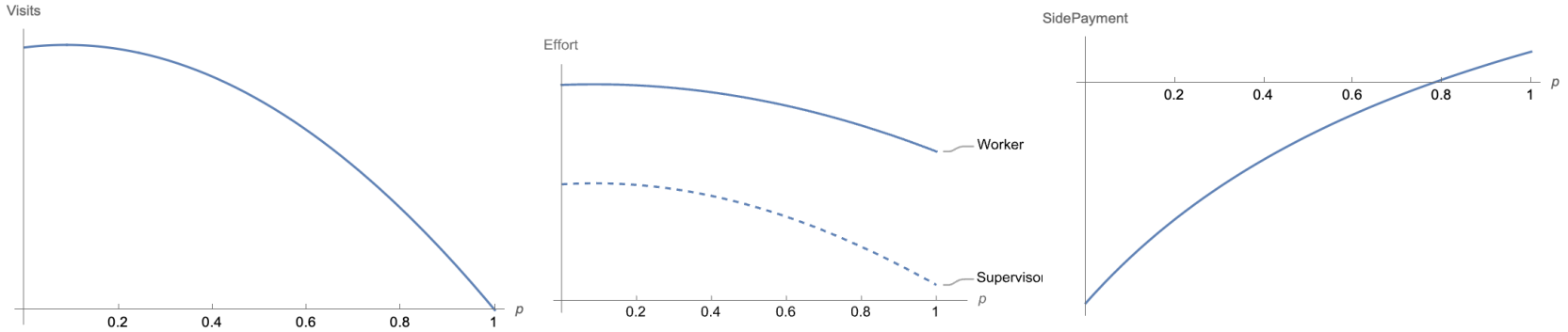
Figure A.1: SMS Reporting over Time



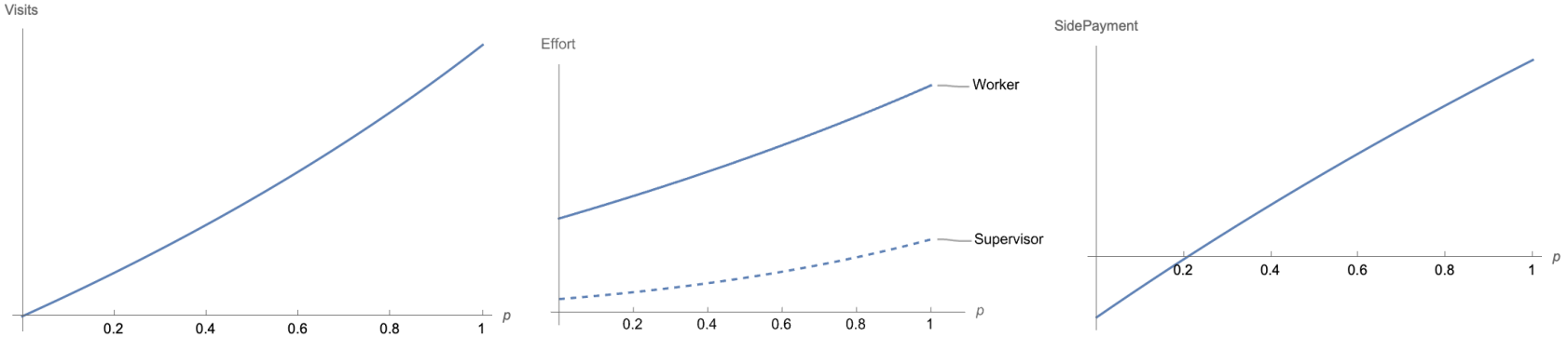

*Notes:* The figure plots the difference in the count of SMS reports between each treatment group and the control group. The coefficients are derived from a regression of the number of SMS reports for each individual month on the treatment dummies, controlling for the stratification variables and with standard errors clustered at the PHU level. The bars represent 95% confidence intervals.

Figure A.2: Model Predictions on Visits, Effort and Side Payments

Panel A: Weak Contractual Frictions




Panel B: Strong Contractual Frictions & Weak Effort Complementarities




Panel C: Strong Contractual Frictions & Strong Effort Complementarities

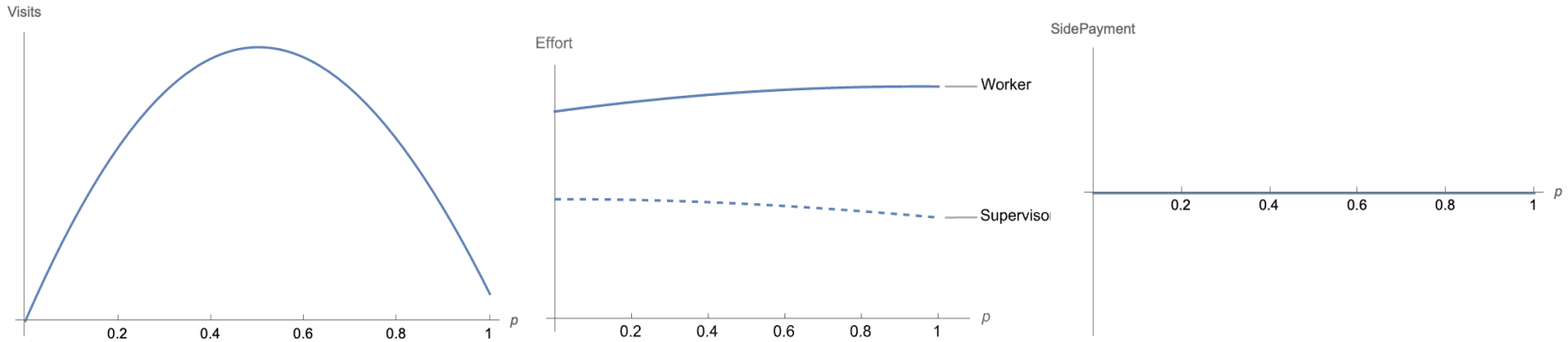


Figure A.3: Model Predictions on Supervisor Welfare with Strong Contractual Frictions and Weak Effort Complementarities

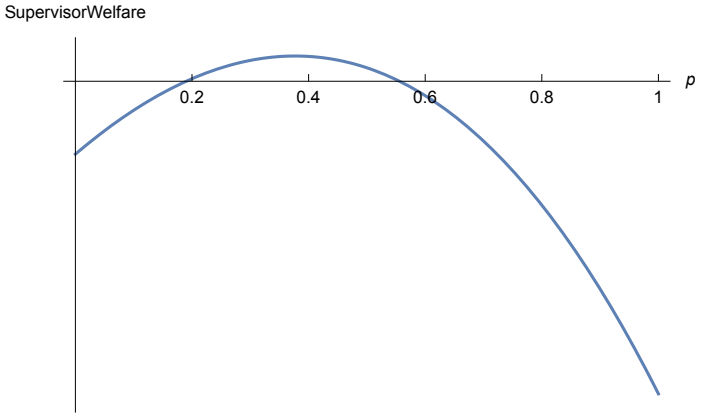
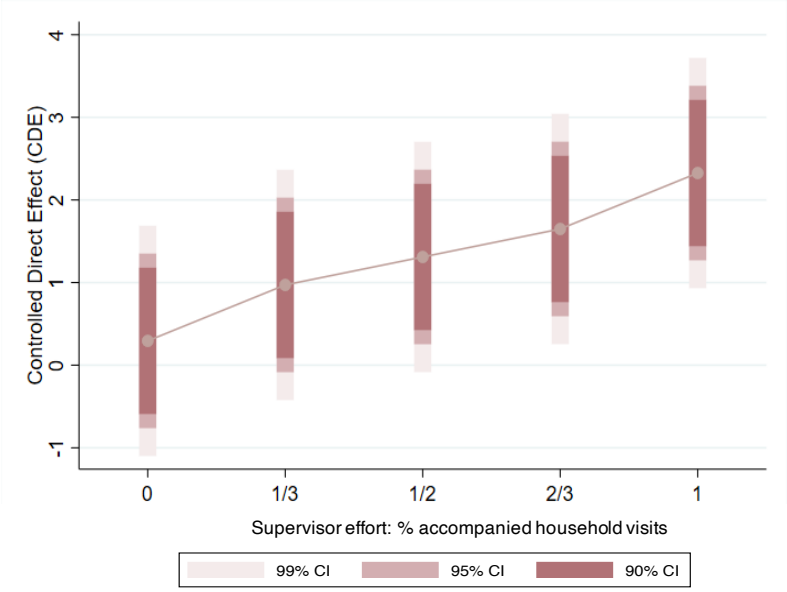
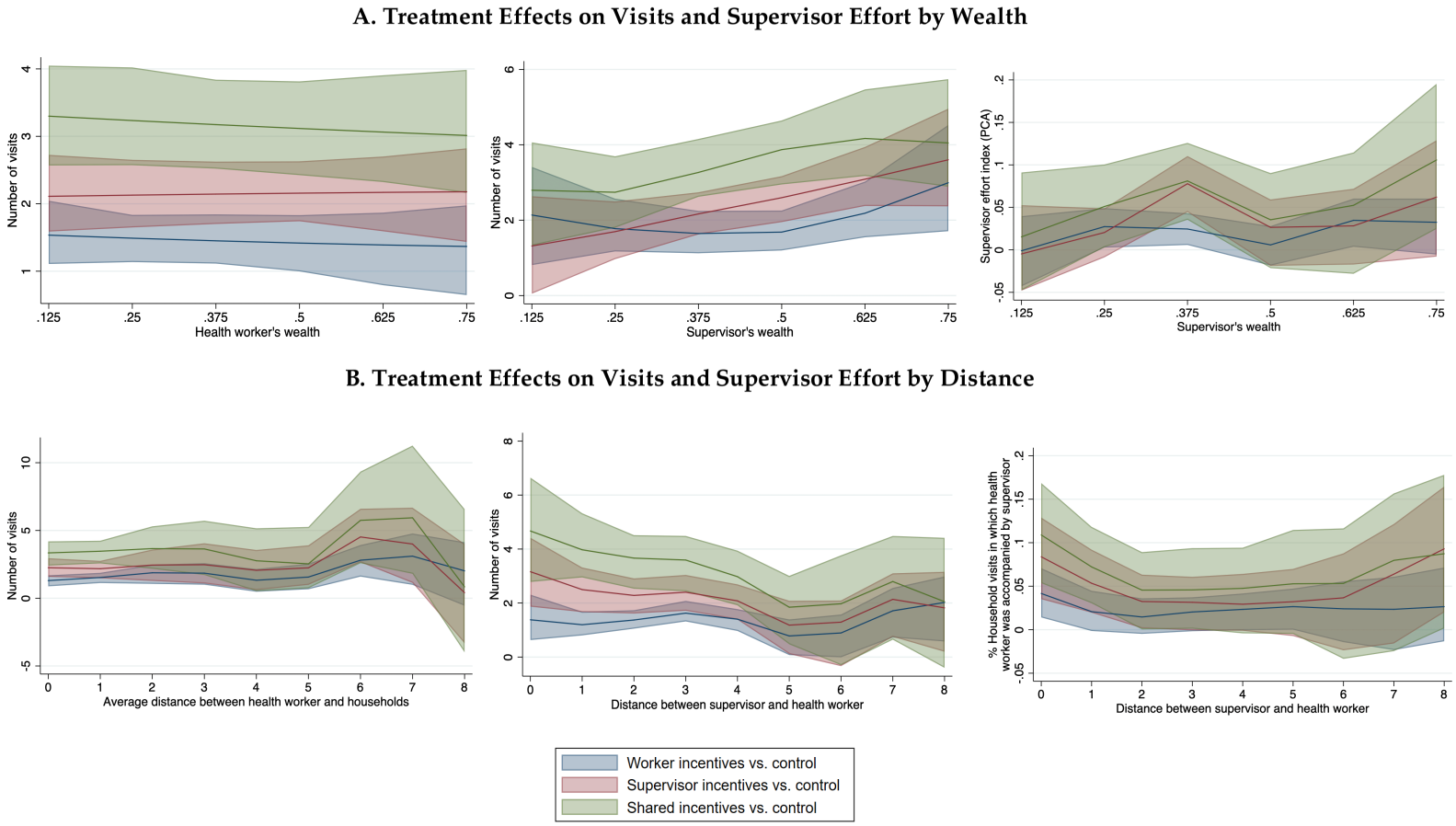



Figure A.4: Mediation Analysis




*Notes:* This figure plots the controlled direct effect (CDE) of the worker incentives treatment on the number of visits provided by a health worker for different values of supervisor's effort, as measured with the share of visits accompanied by the supervisor. We focus on the comparison between the worker incentives treatment and the control group since a mediation analysis performed on the other treatments would be confounded by the fact that in those treatments, the supervisor is directly incentivized to exert effort. We follow the steps outlined in Acharya et al. (2016) to perform the analysis. First, we regress the number of visits a health worker provides on the worker incentives treatment, the mediator (supervisor's effort), and their interaction. Second, we obtain a de-mediated outcome, defined as the difference between actual visits and the number of visits predicted by the regression model for a given level of the mediator. Third, we run a regression of the de-mediated outcome on the treatment. This regression identifies the CDE of the intervention for a given level of the mediator. We repeat this three-step procedure multiple times, changing the level at which we fix the mediator.

Figure A.5: Non-Parametric Estimates of Visits and Supervisor Effort by Wealth and Distance




*Notes:* This figure plots non-parametric estimates of the treatment effects on the number of visits (left and middle figures) and on the supervisor effort index (right figures) by wealth score and by distance, with 95% confidence intervals. We use a local linear estimator with Epanechnikov kernel function. Standard errors are bootstrapped for each value of the x-axis, with 100 repetitions and the re-sampling is with replacement.

Figure A.6: Counterfactual Exercise — Difference in household visits across treatments by years of experience

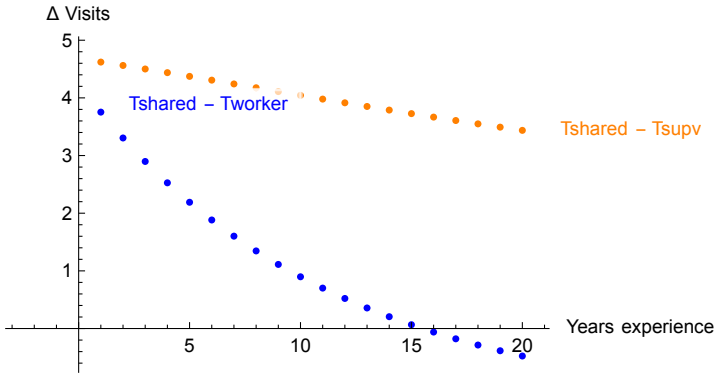
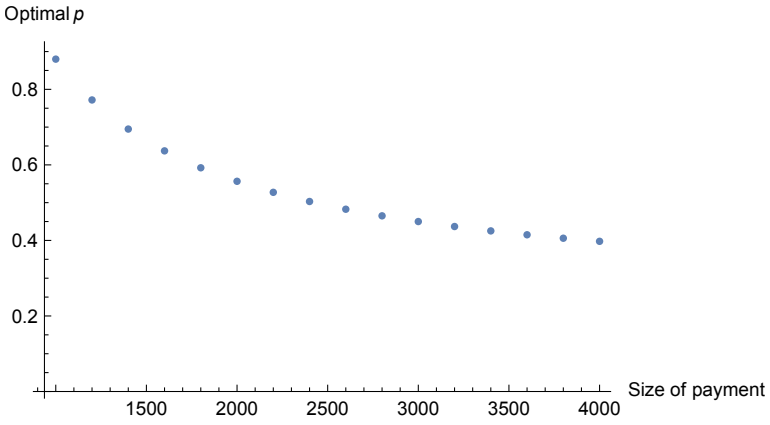



Figure A.7: Counterfactual Exercise — Optimal allocation as the size of the payment increases



# Table A.1: Balance Checks (Pairwise Treatment Comparisons)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | **P-values from Pairwise Treatment Comparisons** | | | | | |
| | Worker Incentives = Supervisor Incentives | Worker Incentives = Shared Incentives | Supervisor Incentives = Shared Incentives | Worker Incentives = Control | Supervisor Incentives = Control | Shared Incentives = Control |
| **A. Characteristics of the supervisors (N=372)** | | | | | | |
| Male = {0, 1} | 0.823 | 0.869 | 0.957 | 0.923 | 0.750 | 0.795 |
| Age (in years) | 0.664 | 0.151 | 0.074 | 0.615 | 0.370 | 0.371 |
| Completed primary education = {0, 1} | 0.399 | 0.454 | 0.109 | 0.592 | 0.748 | 0.195 |
| Completed secondary education = {0, 1} | 0.395 | 0.671 | 0.199 | 0.473 | 0.883 | 0.249 |
| Wealth score (0 to 8) | 0.901 | 0.285 | 0.215 | 0.371 | 0.295 | 0.888 |
| Number of health workers responsible for | 0.375 | 0.450 | 0.904 | 0.054 | 0.304 | 0.253 |
| **B. Characteristics of the health workers (N= 2,970)** | | | | | | |
| Male = {0, 1} | 0.912 | 0.218 | 0.170 | 0.678 | 0.749 | 0.104 |
| Age (in years) | 0.472 | 0.338 | 0.838 | 0.009 | 0.067 | 0.088 |
| Completed primary education = {0, 1} | 0.812 | 0.329 | 0.201 | 0.405 | 0.528 | 0.059 |
| Completed secondary education = {0, 1} | 0.944 | 0.708 | 0.738 | 0.666 | 0.590 | 0.397 |
| Wealth score (0 to 8) | 0.915 | 0.138 | 0.112 | 0.835 | 0.736 | 0.094 |
| Number of households responsible for | 0.532 | 0.353 | 0.711 | 0.291 | 0.138 | 0.096 |
| Distance to supervisor (in km) | 0.043 | 0.443 | 0.228 | 0.204 | 0.443 | 0.636 |
| **C. Characteristics of the female household respondent, aggregate at village level (N=2,970)** | | | | | | |
| Age (in years) | 0.851 | 0.477 | 0.388 | 0.099 | 0.080 | 0.347 |
| Completed primary education = {0, 1} | 0.072 | 0.257 | 0.440 | 0.065 | 0.923 | 0.469 |
| Completed secondary education = {0, 1} | 0.924 | 0.776 | 0.712 | 0.669 | 0.755 | 0.470 |
| Wealth score (0 to 8) | 0.785 | 0.581 | 0.324 | 0.122 | 0.141 | 0.015 |
| Distance to health worker (in km) | 0.727 | 0.907 | 0.818 | 0.184 | 0.327 | 0.225 |
| **D. Characteristics of the village (N=372)** | | | | | | |
| Accessible road to health facility = {0, 1} | 0.784 | 0.511 | 0.361 | 0.809 | 0.991 | 0.400 |
| Phone network available = {0, 1} | 0.715 | 0.341 | 0.210 | 0.361 | 0.222 | 0.955 |
| **E. Services provided by local health facilities per month (N=372)** | | | | | | |
| Number of pregnant women services | 0.669 | 0.539 | 0.811 | 0.311 | 0.467 | 0.637 |
| Number of institutional births | 0.749 | 0.740 | 0.565 | 0.229 | 0.432 | 0.206 |
| Number of fully immunized infants | 0.358 | 0.983 | 0.319 | 0.520 | 0.817 | 0.488 |
| Number of malaria cases treated | 0.345 | 0.102 | 0.458 | 0.076 | 0.368 | 0.860 |
| Number of diarrhea cases treated | 0.674 | 0.383 | 0.746 | 0.379 | 0.727 | 0.973 |

*Notes:* Each row presents p-values from pairwise treatment comparisons. These are calculated from a regression, where the variable is regressed on an indicator for worker, supervisor and shared incentives treatment. All regressions include stratification variables. Standard errors are clustered at the PHU level in worker/village level regressions and we use robust standard errors in PHU/supervisor level regressions. Data source is the supervisor survey in Panel A, the health worker survey in Panel B, the household survey in Panel C, the health worker's leaflet in Panel D, and the facility admin data in Panel E. The wealth score counts the number of items owned on a list of 8 household items (e.g., clothes, pair of shoes, cooking pots).

# Table A.2: P-values Corrected for Multiple Hypothesis Testing

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | **p-values with MHT correction** | | | | | | | | |
| | Romano and Wolf (2016) | | | Bonferroni correction | | | Benjamini, Krieger & Yekutieli (2006) | | |
| | Worker incentives | Supervisor incentives | Shared incentives | Worker incentives | Supervisor incentives | Shared incentives | Worker incentives | Supervisor incentives | Shared incentives |
| **Quantity of visits** | | | | | | | | | |
| *Index on quantity of visits (for next 2 variables)* | 0.004 | 0.004 | 0.004 | 0.002 | 0.000 | 0.000 | | | |
| Number of visits | 0.004 | 0.004 | 0.004 | 0.002 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 |
| % households visited | 0.004 | 0.004 | 0.004 | 0.040 | 0.009 | 0.000 | 0.004 | 0.002 | 0.001 |
| **Quantity of visits** | | | | | | | | | |
| *Index on quality of visits (for next 4 variables)* | 0.004 | 0.004 | 0.004 | 0.239 | 0.161 | 0.000 | | | |
| Number of visit types | 0.004 | 0.004 | 0.004 | 0.124 | 0.019 | 0.000 | 0.006 | 0.002 | 0.001 |
| Average visit length | 0.008 | 0.008 | 0.004 | 0.483 | 0.563 | 0.000 | 0.016 | 0.017 | 0.001 |
| Number of health topics discussed per visit | 0.048 | 0.048 | 0.004 | 1.000 | 1.000 | 0.001 | 0.051 | 0.051 | 0.001 |
| % households who trust the health worker as a health provider | 0.020 | 0.048 | 0.004 | 1.000 | 1.000 | 0.040 | 0.045 | 0.051 | 0.003 |
| **Health outcomes** | | | | | | | | | |
| *Index on pre- and post-natal care in past 2 years (for next 5 variables)* | 0.861 | 0.466 | 0.004 | 1.000 | 1.000 | 0.038 | 0.960 | 0.763 | 0.040 |
| % women who received at least 4 ante-natal visits before birth | 0.976 | 0.673 | 0.012 | 1.000 | 1.000 | 0.553 | 1.000 | 0.763 | 0.208 |
| % women with institutional birth | 0.869 | 0.175 | 0.175 | 1.000 | 1.000 | 1.000 | 0.960 | 0.489 | 0.489 |
| % women who received post-natal visit within 2 days of birth | 0.984 | 0.944 | 0.976 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1000 |
| % women with at least 6 months of breastfeeding | 0.984 | 0.976 | 0.319 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.629 |
| % households with up-to-date infant vaccination | 0.984 | 0.976 | 0.633 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.763 |
| *Index on disease incidence (for next 3 variables)* | 0.984 | 0.964 | 0.100 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.420 |
| % children under-5 who had fever | 0.649 | 0.984 | 0.012 | 1.000 | 1.000 | 0.238 | 0.763 | 1.000 | 0.131 |
| % children under-5 who had cough | 0.542 | 0.984 | 0.980 | 1.000 | 1.000 | 1.000 | 0.763 | 1.000 | 1.000 |
| % children under-5 who had diarrhea | 0.976 | 0.900 | 0.984 | 1.000 | 1.000 | 1.000 | 1.000 | 0.960 | 1.000 |

*Notes:* This table presents multiple hypothesis testing corrected p-values for regressions of the row variable on the three treatment indicators, controlling stratification variables and clustering standard errors at the PHU level. Cols. (1) to (6) control for the familywise error rate, i.e. the probability of making any type 1 error. Cols. (7) to (9) control for the false discovery rate, i.e. the expected proportion of rejections that are type 1 errors. The adjusted p-values in cols. (1) to (3) are calculated following the Romano and Wolf (2016) step-down procedure, with 250 bootstrap resampling iterations. In cols. (4) to (6), a Bonferroni adjustement is applied by multiplying the original p-values by the number of outcomes in the table and capping the adjusted p-values at 1.000. Cols (7) to (9) apply the two-stage Benjamini, Krieger, and Yekutieli (2006)'s procedure.

Table A.3: Household Visits by Type

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | % households who received [visit type] from the health worker in the past 6 months | | | | | | | |
| | Natal-related visits | | | | Disease-related visits | | | |
| Dep. Var. | Pregancy visit | Accompanied woman for birth to the health facility | Pre and post-natal visit | Index (cols. 1-3) | Routine visit | Treatment / referrals of under-5 children | Follow-up visit of under-5 children | Index (cols. 5-7) |
| Worker incentives | 0.037** | -0.005 | 0.027 | 0.069 | 0.068** | 0.053** | 0.042* | 0.155*** |
| | (0.016) | (0.007) | (0.017) | (0.049) | (0.033) | (0.025) | (0.022) | (0.060) |
| Supervisor incentives | 0.027* | 0.004 | 0.037* | 0.092* | 0.089*** | 0.071** | 0.031 | 0.178*** |
| | (0.016) | (0.008) | (0.019) | (0.051) | (0.030) | (0.028) | (0.024) | (0.061) |
| Shared incentives | 0.064*** | 0.008 | 0.051*** | 0.168*** | 0.151*** | 0.111*** | 0.079*** | 0.324*** |
| | (0.017) | (0.008) | (0.017) | (0.051) | (0.029) | (0.024) | (0.021) | (0.056) |
| Unit | Worker | Worker | Worker | Worker | Worker | Worker | Worker | Worker |
| Observations | 2,926 | 2,926 | 2,926 | 2,926 | 2,926 | 2,926 | 2,926 | 2,926 |
| Mean dep. var. | 0.179 | 0.041 | 0.132 | 0.000 | 0.517 | 0.504 | 0.201 | 0.000 |
| Mean dep. var. in Control | 0.145 | 0.038 | 0.103 | -0.087 | 0.437 | 0.443 | 0.162 | -0.171 |
| p-value Worker = Supervisor | 0.543 | 0.193 | 0.604 | 0.650 | 0.540 | 0.527 | 0.669 | 0.733 |
| p-value Supervisor = Shared | 0.037 | 0.660 | 0.455 | 0.154 | 0.046 | 0.143 | 0.046 | 0.024 |
| p-value Worker = Shared | 0.132 | 0.069 | 0.151 | 0.056 | 0.014 | 0.020 | 0.094 | 0.008 |

*Notes:* Data source is the household survey, aggregate at the health worker level. The index in col. (4) [resp., col. (8)] estimates an equally weighted average of the z-scores of variables in cols. (1)-(3) [resp., cols. (5)-(7)]. All regressions include stratification variables. Standard errors clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Household Targeting

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. Var. | Number of visits to households with wealth score above vs. below the median | Number of visits to households with distance above vs. below the median | Number of visits to households who are family members vs. not family members | Number of visits to households who received vs. did not receive a visit accompanied by the supervisor |
| Worker incentives | -0.525 | 0.764 | -0.501 | -0.591 |
| | (0.522) | (0.624) | (0.544) | (0.556) |
| Supervisor incentives | 0.791 | -0.104 | -0.208 | -0.820 |
| | (0.603) | (0.598) | (0.544) | (0.620) |
| Shared incentives | -0.161 | 0.118 | -1.372** | -0.832 |
| | (0.663) | (0.692) | (0.577) | (0.555) |
| Unit | Worker | Worker | Worker | Worker |
| Observations | 2,970 | 2,970 | 2,970 | 2,970 |
| Mean dep. var. | 0.071 | -0.169 | -1.003 | -2.727 |
| Mean dep. var. in Control | 0.067 | -0.351 | -0.520 | -2.201 |
| p-value Worker = Supervisor | 0.041 | 0.241 | 0.631 | 0.752 |
| p-value Supervisor = Shared | 0.214 | 0.778 | 0.070 | 0.987 |
| p-value Worker = Shared | 0.605 | 0.428 | 0.176 | 0.719 |

*Notes:* Data source is the household survey, aggregated at the health worker level. For each health worker, we compute the differential in the number of visits received by households based on various criteria: those with a wealth score above versus below the median (col. 1), those whose distance from the health worker's home is above versus below the median (col. 2), family members versus non-family members of the worker (col. 3), those who received a visit accompanied by the supervisor vs. who never received such a visit (col. 4). In cases where the number of households in a subcategory equals zero, we assign a value of zero to the variable. The wealth score is determined by counting the ownership of items from a list of eight household items (e.g., clothes, pair of shoes, cooking pots). All regressions include stratification variables. Standard errors are clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: More Health Outcomes

| Dep. Var. | (1) Presence of a woman who gave birth in the past year in the household = {0, 1} | (2) % household respondents who know how to prevent malaria and diarrhea | (3) Under-5 mortality rate |
|---|---|---|---|
| Worker incentives | 0.036 | 0.062* | -0.318 |
| | (0.023) | (0.032) | (2.220) |
| Supervisor incentives | 0.035 | 0.053* | 3.083 |
| | (0.023) | (0.032) | (3.838) |
| Shared incentives | 0.010 | 0.086*** | -1.485 |
| | (0.024) | (0.030) | (2.093) |
| Unit | Worker | Worker | Worker |
| Observations | 2,970 | 2,970 | 2,824 |
| Mean dep. var. | 0.841 | 0.563 | 4.135 |
| Mean dep. var. in Control | 0.819 | 0.511 | 3.822 |
| p-value Worker = Supervisor | 0.976 | 0.796 | 0.380 |
| p-value Supervisor = Shared | 0.273 | 0.312 | 0.218 |
| p-value Worker = Shared | 0.263 | 0.460 | 0.571 |

*Notes:* Data source is the household survey, aggregate at the health worker level. In col. (2), a household respondent "knows how to prevent malaria and diarrhea" if she reports that children should (1) sleep under mosquito nets that is sprayed with mosquito repellent/insecticide, (2) use soap, (3) use toilet facility to defecate, and (4) drink clean water. Mortality under-5 is measured as child mortality per 1000 years of exposure to the risk of death (self-reported by households and aggregate at village level). The measure follows the method used Bjorkman Nyqvist et al. (2019). All regressions include stratification variables. Standard errors are clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: Health Facility Services

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | **Pre- and post-natal care at the health facility in the past month** | | | | **Disease treatments at the health facility in the past month** | | | |
| Dep. Var. | Index (cols. 2-4) | Number of pregnant women services | Number of institutional births | Number of fully immunized children | Index (cols. 6-8) | Number of malaria cases treated | Number of diarrhea cases treated | Number of fever cases treated |
| Worker incentives | 0.108 | 5.001 | 0.895 | 2.302 | 0.175* | 9.857* | 4.309** | 22.269** |
| | (0.094) | (5.511) | (1.020) | (1.553) | (0.099) | (5.839) | (2.102) | (9.982) |
| Supervisor incentives | 0.097 | 4.799 | 1.731 | 0.852 | 0.187** | 10.104* | 2.420 | 25.311*** |
| | (0.083) | (4.220) | (1.090) | (1.352) | (0.094) | (5.767) | (2.231) | (9.108) |
| Shared incentives | 0.244* | 13.918* | 2.552* | 3.042* | 0.223* | 9.455 | 6.309* | 24.337** |
| | (0.126) | (7.283) | (1.389) | (1.625) | (0.132) | (6.621) | (3.269) | (11.949) |
| Unit | PHU | PHU | PHU | PHU | PHU | PHU | PHU | PHU |
| Observations | 371 | 371 | 371 | 371 | 371 | 371 | 371 | 371 |
| Mean dep. var. | 0.000 | 41.889 | 13.776 | 12.406 | 0.000 | 57.464 | 18.936 | 132.026 |
| Mean dep. var. in Control | -0.110 | 36.063 | 12.513 | 10.892 | -0.155 | 49.595 | 15.582 | 113.274 |
| p-value Worker = Supervisor | 0.916 | 0.971 | 0.483 | 0.386 | 0.908 | 0.967 | 0.494 | 0.776 |
| p-value Supervisor = Shared | 0.269 | 0.209 | 0.584 | 0.203 | 0.800 | 0.923 | 0.291 | 0.937 |
| p-value Worker = Shared | 0.323 | 0.263 | 0.246 | 0.690 | 0.733 | 0.953 | 0.569 | 0.873 |

*Notes:* Data source is the admin data from the local health facilities (one per PHU). The index in col. (1) [resp., col. (5)] estimates an equally weighted average of the z-scores of variables in cols. (2)-(4) [resp., cols. (6)-(8)]. All regressions include stratification variables. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Supervisor Perceptions and Activities

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | **Supervisor perceptions about the reporting rate** | | **Other supervisor's activities** | | |
| Dep. Var. | Health workers regularly report visits | Perceived reporting rate | Number of times the health worker was visited in the past 6 months | Number of times the health worker was called in the past 6 months | Health worker was ever assisted with SMS reporting |
| Worker incentives | 0.619*** | 0.646*** | -0.050 | 0.031 | -0.000 |
| | (0.058) | (0.022) | (0.137) | (0.154) | (0.023) |
| Supervisor incentives | 0.373*** | 0.360*** | -0.041 | -0.048 | 0.016 |
| | (0.045) | (0.029) | (0.137) | (0.158) | (0.023) |
| Shared incentives | 0.403*** | 0.544*** | -0.043 | -0.194 | -0.003 |
| | (0.043) | (0.029) | (0.139) | (0.163) | (0.022) |
| Unit | Supervisor | Supervisor | Worker | Worker | Worker |
| Observations | 371 | 219 | 2,833 | 2,876 | 2,970 |
| Mean dep. var. | 0.640 | 0.605 | 1.375 | 1.785 | 0.088 |
| Mean dep. var. in Control | 0.291 | 0.230 | 1.417 | 1.847 | 0.084 |
| p-value Worker = Supervisor | 0.001 | <0.001 | 0.950 | 0.617 | 0.449 |
| p-value Supervisor = Shared | 0.608 | <0.001 | 0.987 | 0.382 | 0.354 |
| p-value Worker = Shared | 0.002 | 0.002 | 0.963 | 0.162 | 0.896 |

*Notes:* Data source is the endline supervisor survey in col. (1), the supervisor perception survey in col. (2) and the endline health worker survey in cols. (3)-(5). All regressions include stratification variables. Standard errors clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1.

# Table A.8: Heterogeneity by Worker Experience

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dep. Var. | Number of visits | | Supervisor effort: % accompanied household visits | | Side transfers: Supervisor shared incentive with health worker = {0, 1} | | Reporting rate | |
| **A. Treatment effects for workers with experience below the median:** | | | | | | | | |
| Low experience * Worker incentives | 2.054*** | 2.395*** | 0.030 | 0.030 | -0.004 | -0.011 | 0.716*** | 0.713*** |
| | (0.628) | (0.722) | (0.025) | (0.027) | (0.027) | (0.032) | (0.142) | (0.158) |
| Low experience * Supervisor incentives | 2.576*** | 2.661*** | 0.067** | 0.067** | 0.122*** | 0.135** | 0.151 | 0.159 |
| | (0.598) | (0.646) | (0.029) | (0.031) | (0.045) | (0.053) | (0.108) | (0.120) |
| Low experience * Shared incentives | 4.022*** | 4.335*** | 0.092*** | 0.104*** | 0.057 | 0.066 | 0.239** | 0.232** |
| | (0.684) | (0.751) | (0.026) | (0.027) | (0.046) | (0.052) | (0.095) | (0.102) |
| **B. Treatment effects for workers with experience above the median:** | | | | | | | | |
| High experience * Worker incentives | 2.246*** | 2.056*** | 0.030 | 0.031 | 0.014 | 0.023* | 0.613*** | 0.659*** |
| | (0.780) | (0.756) | (0.031) | (0.032) | (0.010) | (0.014) | (0.207) | (0.204) |
| High experience * Supervisor incentives | 1.720** | 1.657** | 0.045 | 0.045 | 0.257*** | 0.283*** | 0.197 | 0.238 |
| | (0.669) | (0.643) | (0.030) | (0.031) | (0.063) | (0.067) | (0.167) | (0.179) |
| High experience * Shared incentives | 2.583*** | 2.638*** | 0.030 | 0.022 | 0.152*** | 0.175*** | 0.305* | 0.290* |
| | (0.608) | (0.670) | (0.030) | (0.032) | (0.047) | (0.052) | (0.174) | (0.170) |
| High experience | 1.057** | 1.141* | 0.017 | 0.033 | -0.055** | -0.067** | -0.081 | -0.058 |
| | (0.532) | (0.594) | (0.025) | (0.028) | (0.027) | (0.034) | (0.128) | (0.132) |
| Unit | Worker | Worker | Worker | Worker | Worker | Worker | Worker | Worker |
| Extra Controls | No | Yes | No | Yes | No | Yes | No | Yes |
| Observations | 2,909 | 2,552 | 2,902 | 2,547 | 2,898 | 2,509 | 2,607 | 2,288 |
| Mean Dep. Var. | 7.296 | 7.296 | 0.204 | 0.204 | 0.084 | 0.084 | 0.530 | 0.530 |
| Mean Dep. Var. in Control & Low experience | 4.749 | 4.749 | 0.131 | 0.131 | 0.022 | 0.022 | 0.222 | 0.222 |
| ***Treatment comparisons in Panel A (Low experience)*** | | | | | | | | |
| p-value Worker=Supv | 0.455 | 0.733 | 0.226 | 0.236 | 0.005 | 0.004 | 0.001 | 0.003 |
| p-value Worker=Shared | 0.011 | 0.026 | 0.029 | 0.010 | 0.174 | 0.105 | 0.003 | 0.006 |
| p-value Supv=Shared | 0.057 | 0.038 | 0.431 | 0.248 | 0.268 | 0.287 | 0.486 | 0.587 |
| ***Treatment comparisons in Panel B (High experience)*** | | | | | | | | |
| p-value Worker=Supv | 0.551 | 0.630 | 0.643 | 0.676 | 0.000 | 0.000 | 0.039 | 0.037 |
| p-value Worker=Shared | 0.684 | 0.492 | 0.990 | 0.781 | 0.003 | 0.003 | 0.138 | 0.056 |
| p-value Supv=Shared | 0.234 | 0.192 | 0.632 | 0.482 | 0.169 | 0.187 | 0.521 | 0.763 |
| ***Treatment comparisons across Panels (Low vs. High experience)*** | | | | | | | | |
| p-value for Worker incentives | 0.824 | 0.716 | 0.994 | 0.973 | 0.542 | 0.309 | 0.670 | 0.821 |
| p-value for Supervisor incentives | 0.270 | 0.218 | 0.535 | 0.572 | 0.015 | 0.019 | 0.808 | 0.689 |
| p-value for Shared incentives | 0.094 | 0.077 | 0.086 | 0.039 | 0.061 | 0.056 | 0.716 | 0.733 |

*Notes:* The table reports the coefficients from a fully interacted model in which the treatment dummies are interacted with a dummy for whether the worker's experience is high or low. "Low experience" is an indicator that takes value one if the health worker has less than the median number of experience (i.e., less than 4 years of experience) as a health worker at baseline. Data source for the dependent variable is the household survey, aggregated at the health worker level in cols. (1)-(4); and the health worker survey in cols. (5)-(6). The reporting rate in col. (7)-(8) is calculated by dividing the number of SMS reports (from the SMS admin data) by the total monthly visit count, adjusted by the inverse of the sampling probability (obtained from the household survey). The even columns control for the health worker characteristics that are significantly correlated (p<.1) with experience -- i.e., gender, age, wealth score, distance to supervisor -- interacted with the treatment dummies. All regressions include stratification variables. Standard errors are clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Evidence of Contractual Frictions

| | (1) | (2) | (3) |
|---|---|---|---|
| Dep. Var. | Side-payment: Supervisor shared incentive with health worker = {0, 1} | | Net financial gain for the supervisor per health worker (incentives payments - net transfer to the worker*z) |
| Sample: | Workers with higher education than their supervisor | Workers with higher earnings from other job than their supervisor | All workers |
| Worker incentives | 0.032 | -0.033 | 0.219 |
| | (0.028) | (0.065) | (0.260) |
| Supervisor incentives | 0.216*** | 0.248** | 20.896*** |
| | (0.066) | (0.103) | (1.826) |
| Shared incentives | 0.161*** | 0.029 | 17.599*** |
| | (0.062) | (0.080) | (1.274) |
| Unit | Worker | Worker | Worker |
| Observations | 512 | 293 | 2,488 |
| Mean dep. var. | 0.096 | 0.116 | 9.734 |
| Mean dep. var. in Control | 0.000 | 0.044 | 0.057 |
| p-value Worker = Supervisor | 0.004 | 0.020 | <0.001 |
| p-value Supervisor = Shared | 0.532 | 0.088 | 0.139 |
| p-value Worker = Shared | 0.036 | 0.323 | <0.001 |

*Notes:* The sample in cols. (1) and (2) is restricted to workers with higher outside option than supervisor, as proxied with whether the worker has a strictly higher education level than the supervisor or an average hourly earnings from any other job which is higher than the one of their supervisor (conditional on both the worker and the supervisor being engaged in an outside job with a positive income). The dependent variable in col. (3) is equal to (the average number of visits reported by the health worker per month) * (the incentive allocated to the supervisor - net transfer from the supervsior to the worker*z), where z = 3.6 (as estimated in the structural model). The variable is expressed in 1,000 SLL. All regressions include stratification variables. Standard errors are clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Job Satisfaction

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Health worker is satisfied with […] = {0, 1} | | | Supervisor is satisfied with […] = {0, 1} | | | | |
| Dep. Var. | Incentives payments | Organization paying the incentives | Job | Incentives payments | Organization paying the incentives | Job | Health worker perceives the work environment as non-competitive | Health worker self-identifies herself through her job = {0, 1} |
| Worker incentives | 0.344*** | 0.074*** | -0.039 | 0.086 | 0.012 | -0.067 | -0.046 | 0.006 |
| | (0.032) | (0.024) | (0.036) | (0.056) | (0.049) | (0.061) | (0.044) | (0.026) |
| Supervisor incentives | -0.003 | 0.014 | -0.040 | 0.305*** | 0.082* | -0.073 | -0.023 | -0.019 |
| | (0.030) | (0.028) | (0.038) | (0.062) | (0.042) | (0.062) | (0.043) | (0.030) |
| Shared incentives | 0.198*** | 0.048** | -0.054 | 0.329*** | 0.092** | -0.040 | -0.024 | -0.018 |
| | (0.036) | (0.024) | (0.036) | (0.063) | (0.041) | (0.059) | (0.043) | (0.029) |
| Unit | Worker | Worker | Worker | Supervisor | Supervisor | Supervisor | Worker | Worker |
| Observations | 2,709 | 2,825 | 2,876 | 360 | 364 | 359 | 2,923 | 2,923 |
| Mean dep. var. | 0.357 | 0.870 | 0.793 | 0.311 | 0.909 | 0.738 | 0.727 | 0.829 |
| Mean dep. var. in Control | 0.219 | 0.837 | 0.828 | 0.132 | 0.860 | 0.787 | 0.746 | 0.838 |
| p-value Worker = Supervisor | <0.001 | 0.023 | 0.986 | 0.001 | 0.098 | 0.929 | 0.598 | 0.379 |
| p-value Supervisor = Shared | <0.001 | 0.186 | 0.709 | 0.738 | 0.757 | 0.605 | 0.987 | 0.963 |
| p-value Worker = Shared | <0.001 | 0.254 | 0.679 | <0.001 | 0.048 | 0.664 | 0.609 | 0.394 |

*Notes:* Data source is the health worker survey in cols. (1)-(3) and (7)-(11), and the supervisor survey in cols. (4)-(6). A worker/supervisor is defined as unsatisfied with the incentive payment if she reports that the incentive she is paid per valid SMS report is "not fair" (too little). A worker/supervisor is defined as unsatisfied with the environment if she reports that the environment is competitive rather than cooperative. ``Health worker self-identifies herself through her job" is a dummy variable that takes value one if the health worker answers "my job as a community health worker" to the following question: "We have spoken with many people in Sierra Leone and they identify themselves to different groups. Some people self identify themselves as belong to an ethnic group, a language, a religion, etc. Others identify themselves describe themselves in terms of their job. Besides being a citizen of Sierra Leone, which specific group do you feel you belong to first and foremost?". Questions in cols. (7) and (8) were not asked to the supervisor. The sample size changes across columns because a number of health workers and supervisors answered "don't know" to the questions. All regressions include stratification variables. Standard errors clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1

Table A.11: Parameter Estimates (No Effort Complementarities)

| | (1) |
|---|---|
| Worker return to effort $\alpha$ | 24.0090 |
| Supervisor return to effort $\beta$ | 253.7800 |
| Worker baseline incentive $b_1$ | 0.0000 |
| Supervisor baseline incentive $b_2$ | 27.2470 |
| Calibrated friction $z$ | 3.6 height 1pt |

*Notes:* This table shows parameter estimates obtained using minimum distance estimation.

Table A.12: Moment Fit (No Effort Complementarities)

| Moments | Targeted Real | Simulated |
|---|---|---|
| Supervisor effort in worker incentives group | 0.198 | 0.173 |
| Supervisor effort in supervisor incentives group | 0.225 | 0.248 |
| Supervisor effort in shared incentives group | 0.228 | 0.222 |
| Supervisor effort in control group | 0.164 | 0.173 |
| Output in worker incentives group | 59.679 | 61.438 |
| Output in supervisor incentives group | 58.896 | 59.746 |
| Output in shared incentives group | 66.895 | 60.895 |
| Output in control group | 41.040 | 42.977 |
| Value loss function | 18.9 | |

*Notes:* The table shows the targeted empirical moments used for minimum distance estimation as well as the simulated moments.

Table A.13: Model Robustness

| Moments | Headline | Alternative | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| $\Delta$ in marginal product of worker effort (shared incentive) | 144% | 143% | 142% | 140% | 143% |
| $\Delta$ in marginal product of worker effort (no incentive) | 99% | 99% | 98% | 96% | 98% |
| Visits in $T_{shared}$ - visits in $T_{worker}$ | 2.796 | 2.754 | 2.816 | 2.915 | 2.884 |
| Visits in $T_{shared}$ - visits in $T_{supv}$ | 2.923 | 2.820 | 2.968 | 3.161 | 3.103 |

*Notes:* The table shows estimated parameters and simulated moments for different assumptions on the reporting threshold used in $T_{shared}$ and $T_{worker}$. Headline: workers in $T_{worker}$ use the non-strategic threshold, workers in $T_{shared}$ use a threshold that is 10 percent higher than the non-strategic threshold. Alternative 1: workers in $T_{worker}$ and $T_{shared}$ use the non-strategic threshold. Alternative 2: workers in $T_{worker}$ and $T_{shared}$ use a threshold that is 10 percent higher than the non-strategic threshold. Alternative 3: workers in $T_{worker}$ and $T_{shared}$ use a threshold that is 20 percent higher than the non-strategic threshold. Alternative 4: workers in $T_{worker}$ use a threshold that is 10 percent higher than the non-strategic threshold, workers in $T_{shared}$ use a threshold that is 20 percent higher than the non-strategic threshold.

Table A.14: Counterfactual Experiments about Reporting

| Simulated Moments | Main model (1) | Experiment 1 (2) | Experiment 2 (3) | Experiment 3 (4) |
|---|---|---|---|---|
| Supervisor effort in worker incentives | 0.203 | 0.203 (+0.000) | 0.202 (-0.001) | 0.234 (+0.031) |
| Supervisor effort in supervisor incentives | 0.228 | 0.241 (+0.013) | 0.240 (+0.012) | 0.269 (+0.041) |
| Supervisor effort in shared incentives | 0.226 | 0.231 (+0.005) | 0.232 (+0.006) | 0.274 (+0.048) |
| Supervisor effort in control group | 0.157 | 0.126 (-0.031) | 0.119 (-0.038) | 0.155 (-0.002) |
| Worker effort in worker incentives | 3.925 | 3.926 (+0.001) | 3.913 (-0.012) | 4.530 (+0.605) |
| Worker effort in supervisor incentives | 3.662 | 3.434 (-0.228) | 3.312 (-0.350) | 3.976 (+0.314) |
| Worker effort in shared incentives | 3.851 | 3.790 (-0.061) | 3.746 (-0.105 ) | 4.449 (+0.598) |
| Worker effort in control group | 3.033 | 2.438 (-0.595) | 2.304 (-0.729) | 2.996 (-0.037) |
| Output in worker incentives | 60.841 | 60.849 (+0.008) | 60.547 (-0.294) | 76.318 (+15.477) |
| Output in supervisor incentives | 60.714 | 58.933 (-1.781) | 56.603 (-4.111) | 73.030 (+12.316) |
| Output in shared incentives | 63.637 | 63.366 (-0.271) | 62.837 (-0.800) | 82.662 (+19.025) |
| Output in control group | 40.979 | 29.719 (-11.260) | 27.393 (-13.586) | 40.236 (-0.743) |

*Notes:* The table shows simulated moments for our headline model (column 1) and for three counterfactual experiments (column 2, 3 and 4). The difference between a moment simulated in the experiment and the relative moment from our headline model is reported in parenthesis. In the first experiment, the principal forces the worker to raise the reporting threshold in all groups to the level used in $T_{worker}$. In the second experiment, the principal eliminates under-reporting, by forcing the worker to raise the reporting threshold in all groups to a level that guarantees 100 percent reporting. In the third experiment, the principal reduces the mean of the cost distribution by 50 percent, by shifting downwards the upper parameter of the uniform distribution. The changes simulated in the counterfactual experiment are common knowledge among agents.

Table A.15: Sample of Health Workers without Promotion Incentives

| | (1) | (4) | (5) | (2) | (3) | (6) |
|---|---|---|---|---|---|---|
| | Household visits provided by the health worker in the past 6 months | | | | | Trust |
| Dep. Var. | Number of visits | Average visit length | Number of health topics discussed per visit | % households visited | Number of visit types | % households who trust the health worker as a health provider |
| Worker incentives | 2.357*** | 3.915*** | 0.330** | 0.119*** | 0.345*** | 0.051 |
| | (0.815) | (1.359) | (0.165) | (0.040) | (0.132) | (0.033) |
| Supervisor incentives | 2.350*** | 1.988 | 0.362** | 0.138*** | 0.452*** | 0.048 |
| | (0.686) | (1.233) | (0.171) | (0.038) | (0.129) | (0.038) |
| Shared incentives | 3.122*** | 4.267*** | 0.603*** | 0.154*** | 0.636*** | 0.060* |
| | (0.655) | (1.251) | (0.165) | (0.037) | (0.134) | (0.035) |
| Unit | Worker | Worker | Worker | Worker | Worker | Worker |
| Observations | 960 | 960 | 960 | 960 | 960 | 960 |
| Mean dep. var. | 6.753 | 13.702 | 2.168 | 0.682 | 1.660 | 0.730 |
| Mean dep. var. in Control | 4.777 | 11.060 | 1.811 | 0.571 | 1.268 | 0.680 |
| p-value Worker = Supervisor | 0.993 | 0.147 | 0.868 | 0.600 | 0.425 | 0.928 |
| p-value Supervisor = Shared | 0.258 | 0.062 | 0.209 | 0.661 | 0.178 | 0.746 |
| p-value Worker = Shared | 0.344 | 0.794 | 0.146 | 0.330 | 0.036 | 0.780 |

*Notes:* Data source is the household survey, aggregate at the health worker level. Sample restricted to health workers who did not experience any change in promotion incentives. All regressions include stratification variables. Standard errors clustered at the PHU level. *** p<0.01, ** p<0.05, * p<0.1

Table A.16: Treatment Assignments

| | (1) | (2) |
|---|---|---|
| | Meritocratic promotions treatment | Info treatment |
| Worker incentives | -0.011 | 0.000 |
| | (0.074) | (0.074) |
| Supervisor incentives | -0.011 | 0.011 |
| | (0.074) | (0.074) |
| Shared incentives | -0.000 | 0.011 |
| | (0.074) | (0.074) |
| Unit | PHU | PHU |
| Observations | 372 | 372 |

*Notes:* Robust standard errors. The info treatment provided information about supervisor's fixed salary to workers.
*** p<0.01, ** p<0.05, * p<0.1

# Table A.17: Balance Checks — All Pairwise Treatment Comparisons

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | P-values from Pairwise Treatment Comparisons | | | | | |
| | Meritocraticy & Worker Incentives = No Meritocracy & Worker Incentives | Meritocraticy & Supv Incentives = No Meritocracy & Supv Incentives | Meritocraticy & Shared Incentives = No Meritocracy & Shared Incentives | Info & Worker Incentives = No Info & Worker Incentives | Info & Supv Incentives = No Info & Supv Incentives | Info & Shared Incentives = No Info & Shared Incentives |
| **A. Characteristics of the supervisors (N=372)** | | | | | | |
| Male = {0, 1} | 0.748 | 0.836 | 0.769 | 0.593 | 0.474 | 0.583 |
| Age (in years) | 0.599 | 0.765 | 0.970 | 0.313 | 0.731 | 0.822 |
| Completed primary education = {0, 1} | 0.397 | 0.807 | 0.270 | 0.505 | 0.241 | 0.872 |
| Completed secondary education = {0, 1} | 0.397 | 0.759 | 0.270 | 0.505 | 0.241 | 0.627 |
| Wealth score (0 to 8) | 0.391 | 0.487 | 0.327 | 0.454 | 0.754 | 0.395 |
| Number of health workers responsible for | 0.794 | 0.889 | 0.849 | 0.726 | 0.839 | 0.408 |
| **B. Characteristics of the health workers (N= 2,970)** | | | | | | |
| Male = {0, 1} | 0.782 | 0.755 | 0.271 | 0.794 | 0.456 | 0.378 |
| Age (in years) | 0.733 | 0.808 | 0.202 | 0.275 | 0.049 | 0.343 |
| Completed primary education = {0, 1} | 0.291 | 0.301 | 0.422 | 0.715 | 0.828 | 0.925 |
| Completed secondary education = {0, 1} | 0.581 | 0.981 | 0.532 | 0.796 | 0.717 | 0.301 |
| Wealth score (0 to 8) | 0.771 | 0.417 | 0.490 | 0.726 | 0.785 | 0.768 |
| Number of households responsible for | 0.115 | 0.418 | 0.133 | 0.762 | 0.789 | 0.437 |
| Distance to supervisor (in km) | 0.557 | 0.104 | 0.577 | 0.431 | 0.284 | 0.478 |
| **C. Characteristics of the female household respondent, aggregate at village level (N=2,970)** | | | | | | |
| Age (in years) | 0.052 | 0.424 | 0.563 | 0.508 | 0.607 | 0.346 |
| Completed primary education = {0, 1} | 0.008 | 0.045 | 0.405 | 0.187 | 0.635 | 0.227 |
| Completed secondary education = {0, 1} | 0.136 | 0.904 | 0.830 | 0.327 | 0.805 | 0.587 |
| Wealth score (0 to 8) | 0.377 | 0.617 | 0.788 | 0.800 | 0.871 | 0.494 |
| Distance to health worker (in km) | 0.065 | 0.246 | 0.209 | 0.548 | 0.958 | 0.506 |
| **D. Characteristics of the village (N=372)** | | | | | | |
| Accessible road to health facility = {0, 1} | 0.443 | 0.146 | 0.327 | 0.779 | 0.458 | 0.324 |
| Phone network available = {0, 1} | 0.045 | 0.085 | 0.736 | 0.235 | 0.558 | 0.678 |
| **E. Services provided by local health facilities per month (N=372)** | | | | | | |
| Number of pregnant women services | 0.081 | 0.113 | 0.407 | 0.946 | 0.422 | 0.037 |
| Number of institutional births | 0.772 | 0.660 | 0.251 | 0.602 | 0.193 | 0.484 |
| Number of fully immunized infants | 0.322 | 0.339 | 0.318 | 0.731 | 0.058 | 0.799 |
| Number of malaria cases treated | 0.629 | 0.472 | 0.616 | 0.979 | 0.539 | 0.197 |
| Number of diarrhea cases treated | 0.868 | 0.686 | 0.720 | 0.486 | 0.704 | 0.233 |

*Notes:* Each row presents p-values from pairwise treatment comparisons. These are calculated from a regression, where the variable is regressed on 9 indicators that interact the worker, supervisor and shared incentives treatments with the meritocratic promotions treatment or the information treatment (information about supervisor's fixed salary to workers). All regressions include stratification variables. Standard errors are clustered at the PHU level in worker/village level regressions and we use robust standard errors in PHU/supervisor level regressions. Data source is the supervisor survey in Panel A, the health worker survey in Panel B, the household survey in Panel C, the health worker's leaflet in Panel D, and the facility admin data in Panel E. The wealth score counts the number of items owned on a list of 8 household items (e.g., clothes, pair of shoes, cooking pots).

# B Additional Material on the Context and Intervention

## B.1 Context: Checklists

Workers are expected to follow a checklist when they visit a household. The checklist differs depending on the type of visit the health worker conducts:
(i) Prenatal visits to a pregnant woman: Health workers are asked to visit expecting mothers at least four times over the course of a pregnancy. During these visits, health workers should first make sure not only the pregnant woman but also her husband or other decision-makers in the family are present. Second, they assess the pregnant woman for danger signs (e.g., convulsion or fever) that would require an immediate referral to the PHU. Third, they use the Mother, Newborn, and Child Health Card to assess previously agreed actions and current health practices related to the pregnancy with the family. Fourth, health workers present new visit-specific information to the family (e.g., helping with planning for the birth including arranging transportation so the woman can give birth at the PHU). Fifth, health workers and families identify barriers together and agree on an action plan until the next visit. Finally, health workers must fill a register that documents what they have done during the visit.
(ii) Accompanying a pregnant woman to the PHU for childbirth: The health workers should accompany pregnant women to the PHU to give birth. At the PHU, the health worker should help the family obtain all necessary drugs and other supplies. In case a woman delivers at her home rather than at the PHU, the health worker should assist during the birth, communicate the birth to the head of the PHU, and accompany the woman for a post-natal visit at the PHU as soon as possible after birth.
(iii) Postnatal visits within one month of birth: Health workers are asked to visit mothers with newborn babies at least four times during the first month after birth. During these visits, health workers first assess the mother and baby for the presence of any danger signs (e.g., fever or convulsions) that would require a referral to the PHU. Second, they discuss with the family how well they were able to implement health practices agreed upon with the health worker during the previous visit. Third, health workers present new visit-specific information about health behaviors relevant to the mother and baby (e.g., telling the mother to keep the baby warm and only breastfeed the baby). Fourth, they go over a checklist of recommended health behaviors and check whether or not the family knows about and follows them. Fifth, for the items on the checklist that the family does not follow yet, health workers discuss barriers and possible solutions with the family and make a new action plan to be discussed during the next visit. Finally, health workers fill out a register that documents what they have done during the visit.
(iv) Child health checkup visits: Health workers are asked to visit mothers and their young children five times between the age of 1 - 15 months. During these visits, health workers first

assess the child for danger signs (e.g., convulsions or being unable to breastfeed) that would require an immediate referral to the PHU. Second, they use the Mother, Newborn, and Child Health Card to assess previously agreed actions and current health practices related to the pregnancy with the family. Third, health workers present visit-specific information to the mother (e.g., advising the mother on how to transition from exclusive breastfeeding to other foods after the age of 6 months or reminding the mother of scheduled vaccinations for the child). Fourth, health workers and families identify barriers together and agree on an action plan until the next visit. Finally, health workers must fill a register that documents what they have done during the visit.

(v) Visits in which a disease is diagnosed and the patient is either treated or referred to the health facility: The main focus of health workers is on children who are younger than 5 years. They are trained to identify whether a child has diarrhea, malaria, or pneumonia and decide whether or not the child can be treated by the health worker or whether it needs to be referred to the PHU. First, health workers assess the child for general danger signs (e.g., convulsions or the child being unable to breastfeed or drink) which would require an immediate referral to the PHU. Second, they assess the child for the three conditions above (e.g., they count the breaths per minute and compare this to age-specific threshold values in order to assess a child for pneumonia) and decide whether or not the child requires treatment and whether or not the child needs to be referred to the PHU. Health workers also should always assess children for malnutrition.

(vi) Follow-up visits of sick patients: For sick children that were not referred to the PHU, health workers are supposed to do at least two follow-up visits at the child's home on the third and sixth days after the start of the treatment. During these follow-up visits, health workers re-assess the sick child following the same steps as during the initial visit. They also should discuss the condition of the child with the caregiver and counsel the caregiver on disease-specific steps they need to undertake as well as general recommended health behaviors (e.g., hand washing or bed net use).

(vii) Routine household visits: First, health workers introduce themselves and the purpose of the visit. Second, they use the Family Health Card and assess previously agreed upon actions as well as current household health practices with the family. Third, health workers present new health information (e.g., on topics like hand washing and sanitation, bed net use, or family planning) to the family. Finally, health workers and families identify barriers together and agree on an action plan until the next routine household visit by the health worker.

## B.2 Context: Supervision

Supervisors have three main tools to train and advise health workers:

(i) Monthly trainings: Supervisors host a monthly meeting at the PHU, which all health workers under their supervision are supposed to attend. During these trainings, supervisors provide information on health knowledge (how to prevent diseases, recognize danger signs). Central to these monthly meetings is the facilitation of mutual learning among health workers. They are asked to share both successes and barriers they experienced during their work in the previous month. Depending on the number of affected health workers, supervisors help them individually or collectively find solutions for the barriers that have been identified. This often involves re-training health workers on the checklists mentioned above or advising them on effective communication strategies health workers can use with households.
(ii) One-to-one trainings: Supervisors are asked to visit each health worker under their supervision in their village once per month. During these field visits, supervisors go through the records of health workers and randomly select three recent households the health worker provided a service to. For each of these three cases, supervisors ask the health worker about the detailed actions the health worker took and validate whether the steps on the checklists mentioned above have been followed. Supervisors then provide detailed feedback in which they identify gaps in the health worker's knowledge and explain again in detail how to provide the health services correctly.
(iii) In-the-field supervision / direct observation: Supervisors are asked to accompany the health worker to household visits and directly observe how the health worker conducts the visit. During these household visits, supervisors identify both the strengths and weaknesses of the health worker and raise awareness about the importance of her work with the family. After the household visit, supervisors provide personal feedback to the health worker in private.

## B.3 Intervention: Choice of the Treatments and the Incentive Amount

Theoretically, the set of possible splits an organization can select from is larger than the three splits in our design (100%-0%, 50%-50% or 0%-100%). An organization could for instance decide to give 25% of the incentive to the worker and 75% to the supervisor (or vice-versa). Due to the limited sample size of the experiment, we could not test the effect of a wider set of possible splits. We chose the 50%-50% split because informal discussions we had with supervisors (outside of our experimental areas) and government officials indicated that this split was the most natural in our setting. More precisely, we asked these informants how they would split an incentive of 2,000 SLL between supervisors and workers such that the number of visits provided in the PHU is maximized. 63% of the respondents answered that the supervisor should be assigned half of the incentive (1,000 SLL), 8% answered that they should be assigned 60% of the incentive (1,200 SLL), 21% answered that they should be assigned 75% of the incentive (1,500 SLL), and the remaining 8% chose another split. In line

with this, our structural model confirms that the optimal split is indeed very close to the 50%-50% one: see Section 5.

An alternative design to ours to study effort complementarities would offer (i) an incentive of 2,000 SLL to the worker and the supervisor in the shared incentives treatment, (ii) an incentive of 2,000 SLL to the worker in the worker incentives treatment, and (iii) an incentive of 2,000 SLL to the supervisor in the supervisor incentives treatment. In this design, complementarities can be detected by testing whether the impact of the two-sided incentives is greater than the sum of the impacts of the one-sided incentives. However, this would entail varying the amount of the incentive per visit disbursed by the organization, and hence would not shed light on the key question we aim to answer, i.e., how a given incentive should be allocated across the layers of an organization with limited liquidity.

The Ministry of Health and Sanitation selected an incentive amount of 2,000 SLL. This amount was deemed convenient due to its easy divisibility by two in shared incentives scenarios, all the while maintaining its status as a round number. The Ministry regarded any amount above 2,000 as excessive, given that the variable component of a health worker's pay was projected to exceed one fourth of their fixed pay.

## B.4 Intervention: Location of the Experiment and Randomization

Our experiment takes place in 372 PHUs across six districts of Sierra Leone. One district is located in the south (Bo), one in the east (Kenema), three in the north (Bombali, Tonkolili and Kambia) and one in the west (Western Area Rural). Out of the existing 823 PHUs across the six districts, we excluded half because no up-to-date and verified list of health workers was available, and selected 372 PHUs from the remaining eligible PHUs to be part of the experiment.

The 372 PHUs were randomized into four groups of equal size: the worker incentives treatment, the supervisor incentives treatment, the shared incentives treatment, and the control group. The randomization was stratified by: district, average distance between the residence of the supervisor and the health workers in the PHU being above or below the median, and the number of health workers in the PHU being above or below the median. We stratify by these variables because these are key predictors of our main outcome variables. These variables were measured at baseline by surveying the supervisor and the health workers before the randomization took place.

## B.5 Intervention: Promotion Incentives

A random sample of health workers in this study experienced a change in the promotion system. More specifically, six months after the start of the experiment which is the focus of this paper, the promotion system became more meritocratic in half of the 372 PHUs while

the rest of the PHUs kept the status-quo system (in which the promotion decision is at the discretion of the PHU in-charge). See **?** for more details. Importantly, Table A.15 shows that the effects of our incentives treatments persist if we restrict the analysis to the sub-sample of health workers that did not take part in this separate study. Table A.16 shows that the incentives treatments in this study are orthogonal to the treatments of the other study, while Table A.17 presents balance checks on all pairwise treatment comparisons.

## B.6 Intervention: Reporting System

The reporting system works in three steps:
(i) Each time a household visit is provided, the health worker is asked to send an SMS to a toll-free number indicating the date of the service, the name and phone number of the patient, and a one-letter code corresponding to the service type. If the SMS does not include all the required information, the system returns an error message.[1] All health workers in our study (including those in the control group) are asked to report their visits. The incentive was only paid for household visits that fall into one of these categories: (i) prenatal visits to a pregnant woman, (ii) accompanying a pregnant woman to the PHU for childbirth, (iii) postnatal visits within 1 month of birth, (iv) child health checkup visits (for children 1-15 months), (v) visits in which a disease is diagnosed and the patient is either treated or referring to the health facility, (vi) follow-up visits of sick patients, (vii) routine household visits (e.g., providing health education on how to prevent diseases).
(ii) The SMS information is automatically uploaded to a server from which the performance incentives are calculated on a monthly basis and are paid without delay.
(iii) The SMS information is continuously back-checked by a team of monitors who contact a random 25% of households each week either by phone or in-person (unannounced visits), and ask them to confirm the date and the type of the household visit.
All health workers were promised a fixed bonus of SLL 10,000 conditional on truthful reporting at the end of the experiment. Despite this, we show in the paper that the reporting rate is low in all treatments.

## B.7 Intervention: Training

Before initiating the intervention, workers and supervisors from each PHU were invited to attend a 2-days training session together. These sessions usually occurred within or near the main health facility. Across all 372 PHUs, regardless of treatment, the training first provided instructions on the reporting process (i.e., how to send the SMS, at which cost, what to include in the SMS for the report to be "valid"). It further detailed the back-checks

[1] When the patient is a child, the health worker reports the name and phone number of the primary care giver. When the household does not have a phone, the health worker reports the phone number of a neighbor.

and potential consequences of dishonest reporting. For the worker incentives and shared incentives treatments, all attendees, both workers and supervisors, were informed about the reporting incentives to be given to the workers. In the supervisor incentives treatment and control group, attendees were informed that no incentive would be given to the worker. This approach makes supervisors aware of the workers' incentives, and how workers' reporting behavior affects their own pay.

At the conclusion of the training, supervisors were isolated into a separate room (without the workers), where they were informed about their incentives per report (if any). This information was not provided to the workers. Supervisors were then explicitly made aware that they held the discretion to share these incentives, either wholly or partially, with their workers. While we did not suggest a specific amount to be transferred, we told the supervisors that these transfers could be viewed as potential incentives to stimulate worker effort.

### B.8 Data: Variables Definitions

*Accompanied visits*: Each household was asked whether they had received a visit from the health worker (on her own), the supervisor (on her own), the health worker and the supervisor jointly. 71% of household report having received a visit from the health worker, of which 16% are accompanied by the supervisor. No household reports having received a visit from the supervisor without the presence of the health worker. Among households who received at least one accompanied visit, 97% received one accompanied visit only, and 3% received two accompanied visits. Overall, this implies that roughly one-fifth of the households have seen the supervisor once in the past six months, and the vast majority of the remaining households have never seen the supervisor.

*Health knowledge of health workers*: We administered a quiz on health knowledge to all health workers at baseline and endline. The quiz counts the number of correct answers given to the following five questions: (i) Assuming there are otherwise no danger signs, after how many days of fever should the health worker refer a child under 5 to the PHU? (ii) Assuming there are otherwise no danger signs, after how many days of loose or watery stools should the health worker refer a child under 5 to the PHU? (iii) For a child between 2 and 11 months of age, how many breaths per minute is the threshold from which it counts as fast breathing? (iv) For how many years after giving birth should a woman wait before falling pregnant again? (v) Is a composting toilet an improved or unimproved sanitation facility?

## C Research Ethics

Following **?**, we detail key aspects of research ethics.

## C.1 IRB

The project received IRB from the Universitat Pompeu Fabra (Parc de Salut MAR: 2018/7834/I), Northwestern University (ID: STU00207110) and from the Sierra Leone Ethics and Scientific Review Committee (no IRB number was assigned by this local institution).

We obtained informed consent from all participants prior to the study. The consent form described the participants' risks and rights, confidentiality, and contact information. Research staff and enumerator teams were not subject to additional risks in the data collection process. None of the researchers have financial or reputation conflicts of interest with regard to the research results. No contractual restrictions were imposed on the researchers limiting their ability to report the study findings.

The interventions under study did not pose any potential harm to participants and non-participants. The intervention rollout took place according to the evaluation protocol. Our data collection and research procedures adhered to protocols around privacy, confidentiality, risk-management, and informed consent. Participants were not considered particularly vulnerable (beyond some households residing in poverty). Besides individual consent from study participants, consultations were conducted with local representatives at the district levels. All the enumerators involved in data collection were aware about implicit social norms in these communities.

The findings from this project were presented to the MoHS in Sierra Leone. No activity for sharing results with participants in each study village is planned due to resource constraints. We do not foresee risks of the misuse of research findings.

## C.2 AEA RCT Registry

The experimental design and main outcomes of this study were pre-registered in the AEA RCT Registry with the number AEARCTR-0003345. This paper follows the pre-registration closely. The outcome variables we use in the paper were all pre-registered.

The only deviation is that we do not study treatment effects on the "number of hours that the workers/supervisors self-report dedicating to their job," as a measure of their effort. This is because self-reported hours exhibit limited variability in the data, likely because of a self-reporting bias (as mentioned by enumerators in the field). Moreover, self-reported hours worked by the health workers do not correlate with the average number of hours households report having been visited by the worker (number of visits $\times$ average visit length): the correlation is -0.019 and is not statistically significant. Similarly, self-reported hours worked by the supervisors only weakly correlates with the likelihood that a household reports having received a joint visit in which the supervisor was present. For these reasons, we refrain from using self-reported hours as outcome variables in our analysis and use instead an objective measure of output reported by a third party (health visits, as reported by households).

# D Model Appendix

In this section, we first introduce the headline model, which we use in our structural analysis (Section 5). This model employs a production function that accounts for effort complementarities, contractual frictions, endogenous under-reporting, and potential supervisor misperceptions.[2] We then present an alternative version of the model, which employs a production function that excludes effort complementarities. This alternative model is structurally estimated in Section 5 to demonstrate the importance of including effort complementarities in our analysis.

## D.1 Headline Model

There are three players, a worker (player 1), a supervisor (player 2) and a principal (player 3). The worker and the supervisor choose how much effort to exert ($e_1$ and $e_2$ correspondingly) to generate output $y = \alpha e_1 + \gamma e_1 e_2$, where $\alpha$ represents the worker's productivity and $\gamma$ the complementarity between the effort of the supervisor and the worker in producing $y$. Each of the agents incurs a cost of $c_i e_i^2$ from exerting effort. For each unit of output, they obtain some intrinsic utility, $b_i$, as well as a payment from the principal, which is conditional on the unit of output being reported (by the worker to the principal). In particular, if a unit of output is reported, the principal pays a total of $m$ to the worker and supervisor jointly, such that a share $pm$ goes to the worker and a share $(1-p)m$ goes to the supervisor. The principal's role is to choose $p \in [0,1]$ with the objective of maximizing output.

For each unit of output, the worker can choose to report it ($q = 1$) or not ($q = 0$). Reporting is costly to the worker, in the sense that if $q = 1$, she incurs a cost of $k$. The cost of reporting, $k \in \mathbb{R}$, is a random variable, which may be different for each unit of output.

Lastly, the supervisor is allowed to incentivize the worker with a transfer $s$ per unit of reported output. For the worker to get a transfer of $s$, the supervisor needs to pay $sz$ (with $z > 0$). $z$ captures the relative effectiveness of the supervisor in eliciting worker effort. If $z > 1$, the supervisor is relatively less effective than the principal due to contractual frictions. If $z < 1$, the supervisor is more effective than the principal due to a motivation effect. Further, we assume that these side transfers can only be paid by the supervisor to the worker but not the other way around. Also, we assume that $z > 0$.

The timing is as follows. First, the principal chooses $p$. Then, the worker announces a reporting rule $q : \mathbb{R} \to \{0,1\}$ and commits to it. After, the supervisor chooses her effort and the incentive offered to the worker. Lastly, the worker makes her choice of effort. At the end of the game, the variable $k$ is realized for each unit of output, the reporting rule $q$ is implemented and players obtain their payoffs.

[2]Note that, for simplicity, supervisor misperceptions are not included in the main model presented in Section 4.1, but they are added in this section.

The worker's payoff for each extra unit of output, with a realization of the reporting cost, $k_v$, is the following:

$$q(k_v)(mp + s - k_v) + b_1.$$

We assume that the optimal reporting rule $q : \mathbb{R} \to \{0, 1\}$ follows a threshold rule, such that the visit is reported whenever $k_v$ is below some threshold $\tilde{k}$ and not reported otherwise. Thus, the worker's expected payoff per unit of output, for a given $\tilde{k}$, is:

$$F(\tilde{k})(mp + s - E(k|k < \tilde{k})) + b_1$$

where $F$ is the cumulative density function of $k$ and $F(\tilde{k})$ stands for the probability that $k$ is below $\tilde{k}$. To simplify notation, we denote $F(\tilde{k}) := \bar{q}$ and $E(k|k < \tilde{k}) := \bar{k}$. Note that $\bar{k}$ is not $k$'s mean, but a truncated mean.

Moreover, instead of believing the worker's announced reporting rate, the supervisor has a potentially biased perception of the threshold $\tilde{k}$. We denote all the perceived variables with a hat, namely, $\hat{\tilde{k}}$ and $\hat{\bar{q}}$. Note that in the main model presented in Section 4.1, we assume no supervisor misperceptions (i.e., we assume that $\hat{\tilde{k}} = \bar{k}$ and $\hat{\bar{q}} = \bar{q}$ ).

Given these assumptions, the worker's expected payoff can be expressed as:

$$E(\pi_1) = (\alpha e_1 + \gamma e_1 e_2)(\bar{q}(mp + s - \bar{k}) + b_1) - c_1 e_1^2$$

Meanwhile, the expected payoff for the supervisor is:

$$E(\pi_2) = (\alpha e_1 + \gamma e_1 e_2)(\hat{\bar{q}}(m(1 - p) - sz) + b_2) - c_2 e_2^2.$$

The principal's payoff is $\alpha e_1 + \gamma e_1 e_2$.

Now, let us solve the model by backward induction:

Period 4:

The maximization problem of the worker in the last period is:

$$\max_{e_1}(\alpha e_1 + \gamma e_1 e_2)(\bar{q}(mp + s - \bar{k}) + b_1) - c_1 e_1^2.$$

By the FOC, the worker's optimal level of effort is:

$$e_1 = \frac{(\alpha + \gamma e_2)(\bar{q}(mp + s - \bar{k}) + b_1)}{2c_1}.$$

Period 3:

Internalizing the worker's optimal strategy in the third period (but with the misperceived

reporting rule), the supervisor solves the following problem:

$$\max_{e_2,s} \frac{(\alpha+\gamma e_2)^2}{2c_1}(\hat{\bar{q}}(mp+s-\hat{\bar{k}})+b_1)(\hat{\bar{q}}(m(1-p)-sz)+b_2)-c_2e_2^2,$$

which is solved by:[3]

$$e_2 = \frac{\alpha\gamma(\hat{\bar{q}}(mp+s-\hat{\bar{k}})+b_1)(\hat{\bar{q}}(m(1-p)-sz)+b_2)}{2c_1c_2-\gamma^2(\hat{\bar{q}}(mp+s-\hat{\bar{k}})+b_1)(\hat{\bar{q}}(m(1-p)-sz)+b_2)}$$

$$s = \begin{cases} \frac{b_2+\hat{\bar{q}}m-zb_1-\hat{\bar{q}}(mp(1+z)-\hat{\bar{k}}z)}{2\hat{\bar{q}}z}, & p < \frac{b_2+\hat{\bar{q}}m-zb_1+\hat{\bar{q}}\hat{\bar{k}}z}{\hat{\bar{q}}m(1+z)} \\ 0, & otherwise \end{cases}$$

When the principal pays a relatively high incentive to the worker (compared to the supervisor), $p \geq \frac{b_2+\hat{\bar{q}}m-zb_1+\hat{\bar{q}}\hat{\bar{k}}z}{\hat{\bar{q}}m(1+z)}$, the supervisor chooses not to incentivize the worker herself ($s=0$) and provide an effort level:

$$e_2 = \frac{\alpha\gamma(\hat{\bar{q}}(mp-\hat{\bar{k}})+b_1)(\hat{\bar{q}}m(1-p)+b_2)}{2c_1c_2-\gamma^2(\hat{\bar{q}}(mp-\hat{\bar{k}})+b_1)(\hat{\bar{q}}m(1-p)+b_2)}.$$

Plugging $s$and $e_2$ into $e_1$, we obtain the following worker's effort level:

$$e_1 = \frac{\alpha(\bar{q}(mp-\bar{k})+b_1)c_2}{2c_1c_2-\gamma^2(\hat{\bar{q}}(mp-\hat{\bar{k}})+b_1)(\hat{\bar{q}}m(1-p)+b_2)}$$

When the principal pays instead a relatively low incentive to the worker (compared to the supervisor), $p < \frac{b_2+\hat{\bar{q}}m-zb_1+\hat{\bar{q}}\hat{\bar{k}}z}{\hat{\bar{q}}m(1+z)}$, the supervisor chooses to incentivize the worker herself with

$$s = \frac{b_2+\hat{\bar{q}}m-zb_1-\hat{\bar{q}}(mp(1+z)-\hat{\bar{k}}z)}{2\hat{\bar{q}}z}$$

and exerts effort:

$$e_2 = \frac{\alpha\gamma\eta^2}{8zc_1c_2-\gamma^2\eta^2}$$

where $\eta = b_2+zb_1+\hat{\bar{q}}(m(1+p(z-1))-\hat{\bar{k}}z)$. In that case, the worker's effort is given by:

$$e_1 = \frac{4\alpha zc_2}{8zc_1c_2-\gamma^2\eta^2}\left(\frac{\bar{q}(\eta+2\hat{\bar{q}}z(\hat{\bar{k}}-\bar{k}))}{2\hat{\bar{q}}z}+\frac{b_1(\hat{\bar{q}}-\bar{q})}{\hat{\bar{q}}}\right).$$

Period 2:

Recognizing that the reporting rule will influence the supervisor's perceived reporting

[3]We are indirectly assuming that the parameters $c_1, c_2$ and $\gamma$ are such that the solution to the optimal effort levels is positive.

rule, $\hat{\tilde{k}}$,[4] and consequently both the supervisor's choices and the worker's own choice of effort, the worker strategically decides on the threshold $\tilde{k}$. Given the distinctions between scenarios where the supervisor does not incentivize the worker and those where they do, we are presented with two distinct problems for the worker to solve.

First,

$$\max_{\tilde{k}} \frac{\alpha^2(\bar{q}(mp-\bar{k})+b_1)^2c_1c_2^2}{(2c_1c_2-\gamma^2(\hat{\bar{q}}(mp-\hat{\tilde{k}})+b_1)(\hat{\bar{q}}m(1-p)+b_2))^2}$$

subject to the constraint that $p \geq \frac{b_2+\hat{\bar{q}}m-zb_1+\hat{\bar{q}}\hat{\tilde{k}}z}{\hat{\bar{q}}m(1+z)}$.

Next,

$$\max_{\tilde{k}} \frac{16\alpha^2z^2c_1c_2^2}{(8zc_1c_2-\gamma^2\eta^2)^2}\left(\frac{\bar{q}(\eta+2\hat{\bar{q}}z(\hat{\tilde{k}}-\bar{k}))}{2\hat{\bar{q}}z}+\frac{b_1(\hat{\bar{q}}-\bar{q})}{\hat{\bar{q}}}\right)^2$$

subject to $p < \frac{b_2+\hat{\bar{q}}m-zb_1+\hat{\bar{q}}\hat{\tilde{k}}z}{\hat{\bar{q}}m(1+z)}$.

The worker chooses the solution for $\tilde{k}$ that gives the higher utility out of the two.

Period 1:

Lastly, the principal takes all of the above optimal strategies into account to choose $p$ such that total output is maximized.

## D.2 Model with No Complementarities

We now present a model where the production function does not feature strategic effort complementarity: $y = \alpha e_1 + \beta e_2$. Apart from this modification, the model is identical to the model presented in section D.1. The worker's expected payoff becomes:

$$E(\pi_1) = (\alpha e_1 + \beta e_2)(\bar{q}(mp+s-\bar{k})+b_1) - c_1e_1^2$$

while the expected payoff for the supervisor is:

$$E(\pi_2) = (\alpha e_1 + \beta e_2)(\hat{\bar{q}}(m(1-p)-sz)+b_2) - c_2e_2^2.$$

The principal's payoff is $\alpha e_1 + \beta e_2$.

As before, we solve the model by backward induction:

[4] We assume that $\hat{\tilde{k}}$ is an increasing function of $\tilde{k}$.

Period 4:

The maximization problem of the worker in the last period is:

$$\max_{e_1}(\alpha e_1 + \beta e_2)(\bar{q}(mp + s - \bar{k}) + b_1) - c_1 e_1^2.$$

By the FOC, the worker's optimal level of effort is:

$$e_1 = \frac{\alpha(\bar{q}(mp + s - \bar{k}) + b_1)}{2c_1}.$$

Period 3:

Internalizing the worker's optimal strategy in the third period (but with the misperceived reporting rule), the supervisor solves the following problem:

$$\max_{e_2,s}\left(\frac{\alpha^2(\hat{\bar{q}}(mp + s - \hat{\bar{k}}) + b_1)}{2c_1} + \beta e_2\right)(\hat{\bar{q}}(m(1-p) - sz) + b_2) - c_2 e_2^2,$$

which is solved by:

$$e_2 = \frac{\beta(\hat{\bar{q}}(m(1-p) - sz) + b_2)}{2c_2}$$

$$s = \begin{cases} \frac{\alpha^2(b_2 - zb_1 + \hat{\bar{q}}(m(1-p(1+z)) + \hat{\bar{k}}z)) - 2c_1 z\beta e_2}{2\alpha^2 \hat{\bar{q}} z}, & p < \frac{\alpha^2(b_2 - b_1 z + \hat{\bar{q}}(m + \hat{\bar{k}}z)) - 2c_1 z\beta e_2}{\alpha^2 \hat{\bar{q}} m(1+z)} \\ 0, & otherwise \end{cases}$$

When the principal pays a relatively high incentive to the worker (compared to the supervisor), $p \geq \delta$, the supervisor chooses not to incentivize the worker herself ($s = 0$) and provides an effort level:

$$e_2 = \frac{\beta(\hat{\bar{q}} m(1-p) + b_2)}{2c_2}.$$

And the worker's effort level in this case is:

$$e_1 = \frac{\alpha(\bar{q}(mp - \bar{k}) + b_1)}{2c_1}$$

When the principal instead pays a relatively low incentive to the worker (compared to the supervisor), $p < \delta$, the supervisor chooses to incentivize the worker herself with

$$s = \frac{\alpha^2 c_2 \psi - \beta^2 c_1 z(\hat{\bar{q}} m(1-p) + b_2)}{\hat{\bar{q}} z(2\alpha^2 c_2 - \beta^2 c_1 z)}$$

where $\psi = b_2 - zb_1 + \hat{\bar{q}}(m(1-p(1+z)) + \hat{\bar{k}}z)$, and exerts effort:

$$e_2 = \frac{\alpha^2\beta\eta}{2(2\alpha^2c_2 - \beta^2zc_1)}$$

where $\eta = b_2 + zb_1 + \hat{\bar{q}}(m(1-p(1-z)) - \hat{\bar{k}}z)$.

In that case, the worker's effort is given by:

$$e_1 = \frac{\alpha(b_1 + \bar{q}(mp - \bar{k}))}{2c_1} + \frac{\alpha\bar{q}}{2c_1}\frac{\alpha^2c_2\psi - \beta^2c_1z(\hat{\bar{q}}m(1-p) + b_2)}{\hat{\bar{q}}z(2\alpha^2c_2 - \beta^2c_1z)}$$

The threshold $\delta$ corresponds to the value of $p$ for which the last equation for $s$ converges to 0, namely:

$$\delta = \frac{\alpha^2c_2(b_2 - zb_1 + \hat{\bar{q}}(m + \hat{\bar{k}}z)) - \beta^2c_1z(\hat{\bar{q}}m + b_2)}{\hat{\bar{q}}m(\alpha^2c_2(1+z) - \beta^2c_1z}$$

Period 2:

Recognizing that the reporting rule will influence the supervisor's perceived reporting rule, $\hat{\bar{k}}$, and consequently both the supervisor's choices and the worker's own choice of effort, the worker strategically decides on the threshold $\tilde{k}$. Given the distinctions between scenarios where the supervisor does not incentivize the worker and those where they do, we are presented with two distinct problems for the worker to solve.

First,

$$\max_{\tilde{k}} \left(\frac{\alpha^2(\bar{q}(mp - \bar{k}) + b_1)}{2c_1} + \frac{\beta^2(\hat{\bar{q}}m(1-p) + b_2)}{2c_2}\right)(\bar{q}(mp - \bar{k}) + b_1) - c_1\left(\frac{\alpha^2(\bar{q}(mp - \bar{k}) + b_1)}{2c_1}\right)^2$$

subject to the constraint that $p \geq \delta$.

Next,

$$\max_{\tilde{k}} \frac{\alpha^2}{4c_1}\left(b_1 + \bar{q}(mp - \bar{k}) + \frac{\bar{q}(\alpha^2c_2\psi - \beta^2c_1z(\hat{\bar{q}}m(1-p) + b_2))}{\hat{\bar{q}}z(2\alpha^2c_2 - \beta^2c_1z)}\right)^2$$
$$+\frac{\alpha^2\beta^2\eta}{2(2\alpha^2c_2 - \beta^2zc_1)}\left(b_1 + \bar{q}(mp - \bar{k}) + \frac{\bar{q}(\alpha^2c_2\psi - \beta^2c_1z(\hat{\bar{q}}m(1-p) + b_2))}{\hat{\bar{q}}z(2\alpha^2c_2 - \beta^2c_1z)}\right).$$

Finally, the worker chooses the solution for $\tilde{k}$ that gives a higher utility out of the two.

Period 1:

Lastly, the principal takes all of the above optimal strategies into account to choose $p$ such that total output is maximized.

# E Prediction Survey

In collaboration with the Social Science Prediction Platform,[5] we invited social scientists to forecast how our treatments affect household visits compared to the control group. All participants made their forecasts before the results of this study were made public (e.g., before we presented the paper at conferences and before we circulated this draft). Participants were paid to participate in the survey, which included 31 individuals. Of these, 93% are economists, 43% are faculty members, and 54% are graduate students.

Participants were asked to forecast the average number of household visits health workers conduct in $T_{worker}$, $T_{supv}$, and $T_{shared}$ after giving them a 700-words description of the study — i.e., the organization, the role of health workers, the role of supervisors — and informing them about the average number of household visits and its standard deviation for control group workers:

> *"We are interested to hear your predictions about the effects of the different incentive schemes on the main outcome variable, the number of household visits conducted by the community health worker in the previous 6 months as reported by the household's female primary caregiver during the endline household survey. Control Group Reference: As a reference point, community health workers in the control group conducted on average 5.3 visits per household in the 6 months preceding the endline survey, with a standard deviation of 5.6. We would like you to predict the number of visits that the health workers conducted in the other three experimental conditions:*
>
> 1. *How many visits do you think the health workers carried out when the 2,000 incentive was paid in full to the community health worker?*
> 2. *How many visits do you think the health workers carried out when the 2,000 incentive was paid in full to the supervisor?*
> 3. *How many visits do you think the health workers carried out when the 2,000 incentive was shared equally between the community health worker and the supervisor?"*

The average forecasts for the number of household visits by survey participants are 7.73 in $T_{worker}$ (compared to 7.42 we find in the data), 6.28 in $T_{supv}$ (7.48 in the data), and 7.41 in $T_{shared}$ (8.7 in the data). Half (52%) of participants forecasted $T_{worker}$ to be the most effective treatment in our paper, 3% chose $T_{supv}$, 26% chose $T_{shared}$, and 19% forecasted either two or all three treatments to have the same effect.

[5] See https://socialscienceprediction.org. This prediction platform enables the systematic collection and assessment of expert forecasts of the effects of untested social programs.

After asking social scientists to forecast the results from our field experiment, we also asked them to evaluate the expected levels of effort complementarities and contractual frictions in our study setting against four hypothetical organizations commonly studied in low-income countries: a private retail firm, a garment factory, a public school, and a standard centralized government job. We asked:

> *We would like you to forecast the magnitude of two key parameters that we estimate structurally through a basic model of service provision. The model has two key features. First, the efforts of both the supervisor and the health worker are strategic complements. The more effort the supervisor exerts, the higher the marginal productivity of health worker's effort. Second, contractual frictions prevent the two subjects from redistributing the payment costlessly. We capture these frictions with a parameter that measures the additional cost that the supervisor has to pay to make a side payment of one unit to the worker (e.g., if the supervisor has to pay 1.5 units in order to transfer 1 unit, then the contractual friction increases the cost of side payments by 50%).*
>
> 1. *By how much do you think the effort of the supervisor raises the marginal return to the effort of the health worker in the shared incentives treatment? Please express this in percentage terms.*
> 2. *By how much do you think that contractual frictions raise the cost to the supervisor of providing side incentives to the worker (i.e., what is the value of the contractual friction parameter)? Please express this in percentage terms.*
>
> *Finally, we would like to ask you about your expectations on how relevant effort complementarities and contractual frictions between supervisors and frontline workers are in different settings:*
>
> *Private firm in the retail sector:*
>
> 1. *By how much do you think the effort of a supervisor raises the marginal return to the effort of a frontline worker? Please express this in percentage terms. By how much do you think contractual frictions raise the cost to the supervisor of providing side incentives to the worker? Please express this in percentage terms.*
>
> *Garment factory:*
>
> 1. *By how much do you think the effort of a line manager raises the marginal return to the effort of a frontline worker? Please express this in percentage terms.*

2. *By how much do you think contractual frictions raise the cost to the line manager of providing side incentives to the worker? Please express this in percentage terms.*

*Public school:*

1. *By how much do you think the effort of a school principal raises the marginal return to the effort of a teacher? Please express this in percentage terms.*
2. *By how much do you think contractual frictions raise the cost to the school principal of providing side incentives to the teacher? Please express this in percentage terms.*

*Central ministerial administration:*

1. *By how much do you think the effort of a senior bureaucrat raises the marginal return to the effort of a bureaucrat? Please express this in percentage terms.*
2. *By how much do you think contractual frictions raise the cost to the senior bureaucrat of providing side incentives to the bureaucrat? Please express this in percentage terms.*

The participants predicted effort complementarities and contractual frictions to be present in all these different four organizations but they also expected a degree of heterogeneity. They ranked the complementarities as follows: private retail firm $>$ garment factory $>$ our organization $>$ public school $>$ government ministry. For contractual frictions, their ranking was instead: public school $>$ garment factory $>$ government ministry $>$ our organization $>$ private retail firm.