

Additional file 1. Supplementary Methods, Supplementary Figure S1, and Supplementary Tables S1-S12

Associations of pandemic-era telework with cognition and depressive symptoms among middle-aged and older nonagricultural workers in China: a longitudinal CHARLS study

Cognition outcome harmonization

The primary cognition outcome was a harmonized 0-10 score comprising five orientation items (year, month, date, day of the week, and season; 0-5 points) and five serial-7 subtraction items (0-5 points). For the serial-7 component, the fixed-target scoring rule awarded one point for each response matching the expected sequence 93, 86, 79, 72, and 65. Higher total scores indicated better cognitive performance. The same harmonized scoring rule was applied across the 2015, 2018, and 2020 waves.

A process-scored cognition measure was examined as an outcome-definition sensitivity analysis. Under this rule, the first serial-7 response received one point if it equaled 93, and each subsequent numeric response received one point if it was exactly seven less than the participant's immediately preceding response. Thus, the process score assessed whether the serial subtraction sequence was continued correctly even when an earlier numerical response differed from the fixed target. This process-based serial-7 score was combined with the same five orientation items to form an alternative 0-10 cognition score.

Additional cognition sensitivities included a 0-11 score incorporating figure copying and a standalone serial-7 process score. These alternative definitions were used only as sensitivity outcomes and did not replace the primary fixed-target cognition score. Full outcome-definition sensitivity results are reported in Supplementary Table S5, and three-wave results for both cognition scoring approaches are reported in Supplementary Table S6.

Missing data and multiple imputation

Missing baseline covariate information in the primary analysis was addressed using 20 previously generated completed datasets. Telework status and the 2020 co-primary outcomes were not imputed. Before imputation, primary-sample eligibility additionally required sufficient 2018 baseline assessment information, operationalized as at least one valid response among the five 2018 orientation items, and complete information for baseline variables that were not included as imputation targets. Missing values in the imputation targets, including the complete 2018 cognition and CES-D-10 scores, did not themselves result in exclusion. Observed values were preserved in all available completed datasets.

Primary analyses were pooled across the 20 completed datasets using Rubin's rules. A paired fully conditional specification (FCS) sensitivity analysis used the same outcome-specific samples and eight variables with missing values (hukou, work class, poor self-rated health, IADL limitation, fair/poor vision, fair/poor hearing, 2018 cognition, and 2018 CES-D-10). S0 used the prespecified predictor set; S1 additionally included the corresponding 2015 same-domain outcome as a ninth imputation target and historical auxiliary. Nominal variables used stochastic multinomial-logistic models, binary variables used stochastic logistic models, and bounded scores or counts used predictive mean matching with five donors. Both scenarios used 20 completed datasets. Telework and the observed 2020 outcome were predictors but were not imputed.

Propensity-score overlap weighting and balance

Within each completed dataset and outcome-specific analytic sample, a propensity score $e(X)$ was estimated from the baseline covariates. Study-specific overlap weights were calculated as $1-e(X)$ for teleworkers and $e(X)$ for nonteleworkers. These weights were derived from the study propensity-score model and were not the official CHARLS survey weights. Official CHARLS individual survey weights were evaluated separately as a sensitivity analysis.

The propensity-score covariate set included age, sex, education, hukou status, partnered status, work class, poor self-rated health, chronic-condition count, activities of daily living and instrumental activities of daily living limitations, vision and hearing status, baseline cognition, and baseline CES-D-10. Balance was evaluated using absolute standardized mean differences before and after weighting, with $|SMD| < 0.10$ as the predefined balance target. Effective sample sizes and propensity-score/weight support were also examined. Full balance results are reported in Supplementary Table S1 and propensity-score/overlap-weight diagnostics in Supplementary Table S2.

Categorical variables were derived from the 2018 CHARLS questionnaire. Education (BD001_W2_4) was grouped as: code 0, no formal education/illiterate (reference); code 1, some primary education, Sishu/home school, or elementary school; code 2, middle school; code 3, high or vocational school; and code 4, college/associate degree or higher. Hukou (ZBC004) was grouped as: code 1, agricultural (reference); code 2, nonagricultural; and code 3, unified residence or no hukou. Work class was derived from XF1/XF7: code 0, not working or never worked in 2018; code 1, nonfarm employee (reference); code 2, nonfarm self-employed; code 3, unpaid family business worker; and code 4, farm employee. The no-hukou group was

combined with unified residence because it was very small. All groupings and reference categories were fixed before outcome modeling and were not selected according to the outcomes or statistical significance.

Self-rated health (DA002) used five response levels: very good, good, fair, poor, and very poor. Poor self-rated health equaled 1 for poor/very poor and 0 for very good/good/fair. Distance vision (DA033), near vision (DA034), and hearing (DA039) were rated excellent, very good, good, fair, or poor. Fair/poor vision equaled 1 if either distance or near vision was fair/poor; fair/poor hearing equaled 1 if hearing was fair/poor. Fair and poor ratings were combined to represent less-than-good self-reported sensory function, using a common binary definition for vision and hearing. These indicators retain the questionnaire response labels and combine the two vision items into one prespecified measure. Don't know, refusal, and other non-substantive responses were treated as missing.

Primary outcome models and sex interaction

For cognition, the 2020 cognition score was regressed on telework status, 2018 cognition, age, sex, education, hukou status, partnered status, work class, poor self-rated health, and chronic-condition count. The CES-D-10 model used the same covariate structure but adjusted for 2018 CES-D-10 instead of baseline cognition. Overlap weights entered the regression models directly.

Robust covariance matrices were clustered by 2020 community. Estimates were pooled across the 20 completed datasets using Rubin's rules, with Barnard-Rubin degrees of freedom. The two co-primary P values were adjusted using the Holm procedure. Sex heterogeneity was evaluated by adding a telework x sex interaction to the corresponding primary model; inference regarding heterogeneity was based on the interaction coefficient rather than on differences between subgroup-specific P values. Full interaction results are reported in Supplementary Table S7.

Three-wave longitudinal analyses

The longitudinal extension used harmonized 2015, 2018, and 2020 outcomes and was separate from the MI20 overlap-weighted primary analyses. In the complete-three-wave approach, participants with the relevant outcome observed at all three waves were retained, individual changes were calculated for 2015-2018, 2018-2020, and 2015-2020, and each change was compared according to later 2020 telework status using community-clustered robust standard errors.

A second all-available participant fixed-effects approach retained participants with at least two observed waves for the relevant outcome. Within-person demeaning with wave indicators and telework-by-wave terms was used to estimate differential changes while removing time-invariant participant characteristics. Because the 2015-2018 period preceded the 2020 telework measure, that contrast was used only to assess whether participants who later did and did not report telework were already following different outcome trajectories before the pandemic. Full longitudinal results are reported in Supplementary Table S6.

Sensitivity analyses

Sensitivity analyses examined alternative approaches to covariance estimation, empirical common support, missing data, employment eligibility, geographic adjustment, official survey weighting, outcome definition, outcome observation, longitudinal modelling, and propensity-score matching. No outcome-based outlier exclusion or winsorization was used in the primary analysis. Full sensitivity results are reported in Supplementary Table S5.

The 1:1 propensity-score-matching analysis was retained as a sensitivity analysis only because the predefined post-match balance criterion was not achieved (maximum |SMD|=0.168 for cognition and 0.151 for CES-D-10). Weighted HC3 covariance was also examined as a sensitivity specification. Propensity-score-matching balance and HC3 sensitivity details are reported in Supplementary Table S8.

Community fixed-effects analyses were included to reduce confounding by time-invariant differences between communities. These analyses cannot remove unmeasured province- or city-specific factors that changed over time during the pandemic, such as local infection intensity or restrictions; these factors remain a limitation of the observational study.

Figure 2 model specifications

Empirical common support was defined within each completed dataset as the intersection of the observed propensity-score range among teleworkers and the observed range among nonteleworkers. The primary propensity score was retained; records outside the intersection received zero analysis weight. Community fixed effects and community-clustered robust covariance were treated as distinct model components: the community fixed-effects sensitivity was unweighted and entered the 2020 community identifier as a categorical factor in the outcome model while retaining community-clustered robust standard errors. The official individual CHARLS sampling weight was used alone in a separate sensitivity model and was not multiplied by the overlap or outcome-observation weights. For the working-in-2018, 2020-employees-only, and exclude-nonagricultural-self-employed sensitivities, propensity scores and overlap weights were re-estimated within each restricted analysis population.

Outcome-observation weighting

For outcome-observation weighting, $R=1$ denoted entry into the outcome-specific primary analytic cohort. The denominator model estimated $\Pr(R=1)$ conditional on the observed 2020 outcome, telework status, age, sex, and education; the numerator model included telework status only. The stabilized observation weight was the numerator probability divided by the denominator probability, and the final regression weight was the primary overlap weight multiplied by the stabilized observation weight. The model was fitted within the outcome-observed support population in which the required predictors were observed, and no truncation was applied. This analysis addresses selection into the primary analytic cohort within this support population and does not recover participants whose 2020 outcome itself was unobserved.

Mixed-exposure-community support analysis

A mixed-exposure community contained at least one teleworker and at least one nonteleworker. After restriction, the propensity score and overlap weights were re-estimated within each of the 20 completed datasets. The cognition analysis retained 1,853 participants, 261 teleworkers, and 169 communities; the CES-D-10 analysis retained 1,906 participants, 269 teleworkers, and 170 communities. The cognition estimate was $\beta=0.226395$ (95% CI -0.025370 to 0.478160) in mixed-exposure communities, compared with $\beta=0.277575$ (95% CI 0.039949 to 0.515201) in the full-cohort primary analysis. The restricted estimand pertains to the covariate-overlap population within communities in which both exposure states were empirically represented.

Random-intercept mixed-model sensitivity

The random-intercept sensitivity analysis was an unweighted long-format 2018–2020 model fitted by restricted maximum likelihood (REML) with a participant-specific random intercept. Fixed effects included time, telework, time $\times$ telework, age, sex, education, hukou status, partnered status, work class, poor self-rated health, and chronic-condition count. The model did not use community-clustered robust covariance. The time $\times$ telework coefficient estimates differential 2018–2020 change and is a different longitudinal estimand from the primary baseline-adjusted ANCOVA.

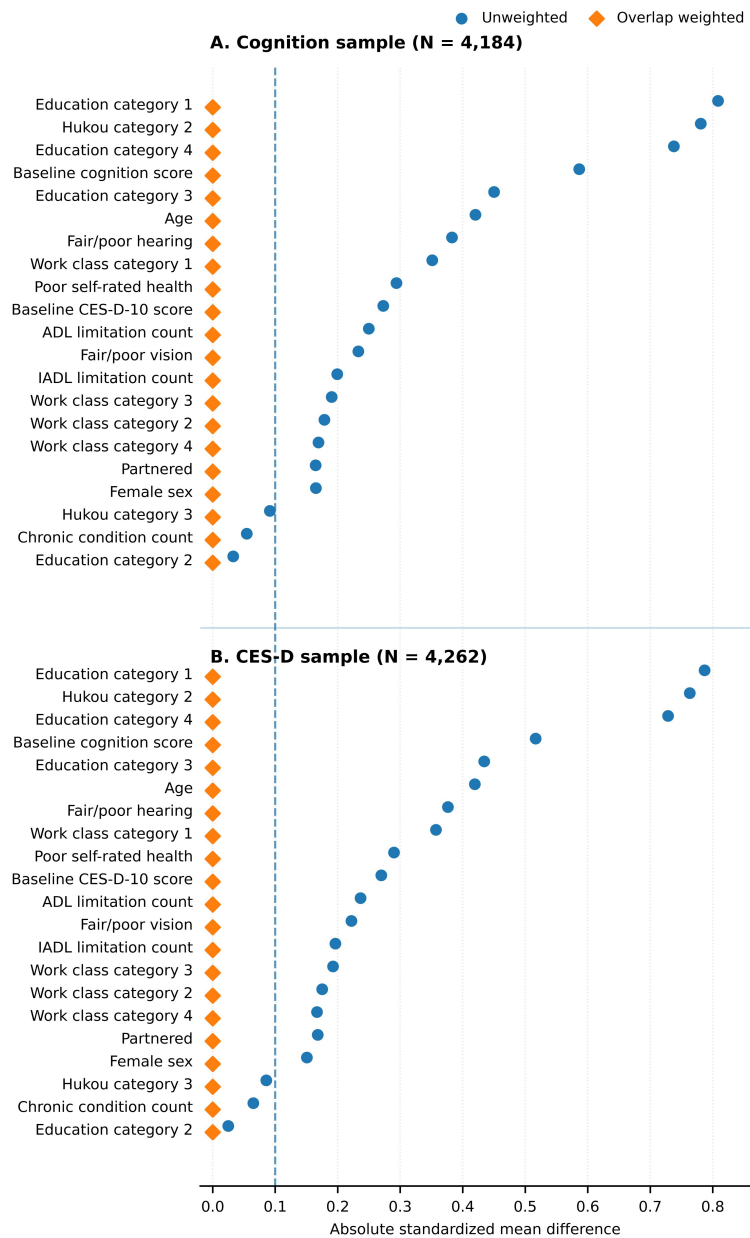
Expanded outcome-model sensitivity

The expanded outcome-model sensitivity retained the original overlap weights and added the complete propensity-score covariate set to each weighted outcome regression. Estimates were combined using Rubin's rules with Barnard-Rubin degrees of freedom. Detailed specifications for the analyses shown in Figure 2 are provided in Supplementary Table S12.

Community unit and missing-data details

The 2020 community identifier (commid2020) denotes the CHARLS village or urban residents' committee-level sampling unit, not a city or county. For every S0/S1 imputation target, the predictor set comprised telework, the observed 2020 outcome, age, sex, five-level education, partnered status, chronic-condition count, ADL limitation, current work status, and the health-response proxy; the other imputed targets also participated in the chained system. S1 additionally included the corresponding 2015 same-domain outcome. Both S0 and S1 used 20 completed datasets. Propensity-score, weighting, and outcome analyses were conducted using Stata/SE 15.1.

Supplementary Figure S1. Covariate balance before and after overlap weighting



Absolute standardized mean differences (SMDs) are shown for propensity-score covariate components in the cognition (N=4,184) and CES-D-10 (N=4,262) primary samples. Values are means across 20 completed datasets. Circles indicate unweighted balance and diamonds indicate overlap-weighted balance; the dashed line marks $|SMD|=0.10$. The plotted categorical components use numeric codes defined above. For compact display, the lowest-coded component was omitted; this plotting convention does not define the regression reference categories. All assessed components satisfied $|SMD|<0.10$ after weighting.

Supplementary Table S1. Full covariate balance before and after propensity-score overlap weighting across 20 completed datasets

Covariate component	Mean SMD , unweighted	Range across 20 datasets, unweighted	Mean SMD , overlap weighted	Range across 20 datasets, overlap weighted
Panel A. Cognition primary sample (N=4,184; telework n=270)				
Age, years	0.420	0.420-0.420	1.21E-15	0-3.45E-15
Female sex	0.165	0.165-0.165	8.54E-16	0-2.03E-15
Education category 1	0.808	0.808-0.808	4.62E-16	4.26E-17-1.02E-15
Education category 2	0.033	0.033-0.033	4.96E-16	0-1.73E-15
Education category 3	0.450	0.450-0.450	5.39E-16	1.17E-16-1.52E-15
Education category 4	0.738	0.738-0.738	6.31E-16	7.74E-17-1.70E-15
Hukou category 2	0.781	0.741-0.816	8.39E-16	0-3.35E-15
Hukou category 3	0.091	0.078-0.140	3.09E-16	2.36E-17-1.16E-15
Partnered	0.165	0.165-0.165	6.75E-16	0-1.23E-15
Work class category 1	0.351	0.319-0.380	5.33E-16	0-1.23E-15
Work class category 2	0.179	0.144-0.208	4.42E-16	0-1.09E-15
Work class category 3	0.190	0.139-0.207	2.15E-16	1.83E-17-8.40E-16
Work class category 4	0.169	0.128-0.192	2.59E-16	7.76E-17-5.43E-16
Poor self-rated health	0.294	0.294-0.295	2.54E-15	1.06E-16-3.88E-15
Chronic condition count	0.055	0.055-0.055	8.77E-15	2.78E-15-1.33E-14
ADL limitation count	0.250	0.250-0.250	7.26E-16	5.73E-17-1.27E-15
IADL limitation count	0.199	0.160-0.223	8.38E-16	3.80E-17-1.56E-15
Fair/poor vision	0.233	0.232-0.233	1.04E-15	0-2.38E-15
Fair/poor hearing	0.383	0.378-0.387	1.49E-15	2.25E-16-3.04E-15
Baseline cognition score (0-10)	0.586	0.544-0.624	9.81E-16	0-1.82E-15
Baseline CES-D-10 score (0-30)	0.273	0.254-0.290	7.66E-16	0-2.69E-15
Panel B. CES-D primary sample (N=4,262; telework n=278)				
Age, years	0.419	0.419-0.419	1.67E-12	1.37E-12-2.09E-12
Female sex	0.151	0.151-0.151	9.60E-13	7.66E-13-1.31E-12
Education category 1	0.787	0.787-0.787	5.07E-13	4.08E-13-6.88E-13
Education category 2	0.025	0.025-0.025	1.19E-13	4.69E-14-2.45E-13
Education category 3	0.434	0.434-0.434	4.00E-13	3.34E-13-5.03E-13
Education category 4	0.728	0.728-0.728	1.34E-12	1.11E-12-1.61E-12
Hukou category 2	0.763	0.736-0.802	1.01E-12	8.90E-13-1.27E-12
Hukou category 3	0.086	0.077-0.115	8.76E-14	1.70E-15-2.77E-13
Partnered	0.168	0.168-0.168	4.69E-13	3.81E-13-5.30E-13
Work class category 1	0.357	0.322-0.378	1.65E-12	1.13E-12-2.56E-12
Work class category 2	0.175	0.133-0.196	2.37E-13	3.52E-14-5.13E-13
Work class category 3	0.193	0.142-0.208	3.74E-12	2.07E-13-9.26E-12
Work class category 4	0.167	0.143-0.194	8.18E-13	5.04E-13-1.28E-12
Poor self-rated health	0.290	0.290-0.291	3.20E-12	2.76E-12-3.68E-12
Chronic condition count	0.065	0.065-0.065	2.61E-12	2.30E-12-3.02E-12
ADL limitation count	0.236	0.236-0.236	2.93E-11	2.59E-11-3.50E-11
IADL limitation count	0.196	0.160-0.210	1.18E-11	9.67E-12-1.36E-11
Fair/poor vision	0.222	0.221-0.222	1.37E-12	1.21E-12-1.57E-12
Fair/poor hearing	0.376	0.373-0.381	1.25E-12	1.03E-12-1.54E-12
Baseline cognition score (0-10)	0.517	0.488-0.563	2.26E-12	1.73E-12-2.69E-12
Baseline CES-D-10 score (0-30)	0.270	0.251-0.286	3.86E-12	3.29E-12-4.43E-12

Notes: SMD, standardized mean difference. Entries are mean |SMD| and range across the 20 completed datasets. Supplementary Figure S1 uses |SMD|=0.10 as the balance reference. Overlap weights were $1-e(X)$ for teleworkers and $e(X)$ for nonteleworkers. The maximum observed unweighted |SMD| was 0.816 for cognition and 0.802 for CES-D; the corresponding maximum overlap-weighted |SMD| values were 1.33E-14 and 3.50E-11. Near-zero values are retained in scientific notation. Categorical component numbers follow the source-verified coding reported in Supplementary Methods; omitted components reflect the balance-display parameterization and do not redefine the regression reference categories.

Supplementary Table S2. Propensity-score and overlap-weighting diagnostics

Diagnostic	Cognition	CES-D
Analytic sample, N	4184	4262
Teleworkers, n	270	278
Propensity-score minimum, mean (range)	0.0004 (0.0003-0.0005)	0.0007 (0.0005-0.0009)
Propensity-score maximum, mean (range)	0.6774 (0.6620-0.8349)	0.6869 (0.6676-0.7081)
Empirical support lower limit, mean (range)	0.0069 (0.0062-0.0075)	0.0048 (0.0040-0.0059)
Empirical support upper limit, mean (range)	0.6570 (0.6499-0.6807)	0.6863 (0.6584-0.7081)
Common-support sample, N, mean (range)	3871.0 (3809-3959)	4128.7 (4043-4164)
Teleworkers in common support, n, mean (range)	269.0 (268-270)	277.9 (277-278)
Nonteleworkers in common support, n, mean (range)	3602.0 (3540-3689)	3850.8 (3766-3886)
Overlap-weight minimum, mean (range)	0.0004 (0.0003-0.0005)	0.0007 (0.0005-0.0009)
Overlap-weight maximum, mean (range)	0.9931 (0.9925-0.9938)	0.9952 (0.9941-0.9960)
Effective sample size, overall, mean (range)	845.6 (843.2-847.8)	873.2 (871.2-874.6)
Effective sample size, teleworkers, mean (range)	253.5 (253.0-253.8)	261.1 (260.8-261.5)
Effective sample size, nonteleworkers, mean (range)	1273.7 (1260.0-1289.2)	1332.0 (1319.1-1343.2)
Maximum SMD before weighting, mean (range)	0.814 (0.808-0.837)	0.797 (0.787-0.811)
Maximum SMD after overlap weighting, mean (range)	See Table S1	See Table S1
Datasets meeting PS/OW diagnostic checks	20/20	20/20

Notes: Values summarize the 20 completed datasets used for the primary propensity-score/overlap-weighting analyses and correspond to the primary and common-support rows in Supplementary Table S5. PS, propensity score; OW, overlap weight; ESS, effective sample size; SMD, standardized mean difference. Overlap weights were defined as $1-e(X)$ for teleworkers and $e(X)$ for nonteleworkers. Empirical common support was the intersection of the observed propensity-score ranges in teleworkers and nonteleworkers. Across all 20 completed datasets, propensity scores and overlap weights were complete and strictly between 0 and 1. The paired S0/S1 FCS analyses used separately generated completed datasets and are reported as missing-data sensitivity analyses rather than as primary diagnostics. Scope clarification for maximum |SMD|: Table 1 and Supplementary Table S1 report maxima across the components shown in the compact balance display. The pre-weighting maximum in Supplementary Table S2 was calculated across the full dummy-expanded propensity-score design, including components omitted only from that display; its range upper bounds (0.837 for cognition; 0.811 for CES-D-10) can therefore exceed the displayed-component maxima (0.816; 0.802). Post-weighting component-level balance is reported in Supplementary Table S1.

Supplementary Table S3. Multiple-imputation diagnostics and distribution summaries

Outcome	MI specification	Imputed variable	Missing, n	Observed mean (SD)	Imputed mean (SD)	Imputed range	SD of means across M=20	Trace shift / SD	Datasets meeting diagnostic checks
Cognition	S0 paired sensitivity	Hukou category (2018)	134	1.197 (0.403)	1.267 (0.452)	1.0-3.0	0.033	0.678	20/20
Cognition	S0 paired sensitivity	Work class category (2018)	570	1.303 (0.842)	1.577 (0.890)	0.0-4.0	0.039	0.330	20/20
Cognition	S0 paired sensitivity	Poor self-rated health (2018)	1	0.148 (0.355)	0.050 (0.224)	0.0-1.0	0.224	0.894	20/20
Cognition	S0 paired sensitivity	IADL limitation count (2018)	67	0.160 (0.590)	0.250 (0.683)	0.0-6.0	0.087	0.754	20/20
Cognition	S0 paired sensitivity	Fair/poor vision (2018)	4	0.764 (0.425)	0.863 (0.347)	0.0-1.0	0.151	0.661	20/20
Cognition	S0 paired sensitivity	Fair/poor hearing (2018)	4	0.563 (0.496)	0.463 (0.502)	0.0-1.0	0.247	0.405	20/20
Cognition	S0 paired sensitivity	Cognition score (2018)	767	7.329 (2.418)	6.531 (2.629)	0.0-10.0	0.056	0.728	20/20
Cognition	S0 paired sensitivity	CES-D-10 score (2018)	295	7.001 (5.686)	8.139 (6.079)	0.0-30.0	0.357	0.313	20/20
Cognition	S1 history-informed	Hukou category (2018)	134	1.197 (0.403)	1.258 (0.443)	1.0-3.0	0.032	1.485	20/20
Cognition	S1 history-informed	Work class category (2018)	570	1.303 (0.842)	1.566 (0.879)	0.0-4.0	0.047	0.603	20/20
Cognition	S1 history-informed	Poor self-rated health (2018)	1	0.148 (0.355)	0.000 (0.000)	0.0-0.0	0.000	NA*	20/20
Cognition	S1 history-informed	IADL limitation count (2018)	67	0.160 (0.590)	0.246 (0.692)	0.0-6.0	0.070	0.171	20/20
Cognition	S1 history-informed	Fair/poor vision (2018)	4	0.764 (0.425)	0.887 (0.318)	0.0-1.0	0.151	0.000	20/20
Cognition	S1 history-informed	Fair/poor hearing (2018)	4	0.563 (0.496)	0.463 (0.502)	0.0-1.0	0.203	0.984	20/20
Cognition	S1 history-informed	Cognition score (2018)	767	7.329 (2.418)	6.437 (2.666)	0.0-10.0	0.071	0.077	20/20
Cognition	S1 history-informed	CES-D-10 score (2018)	295	7.001 (5.686)	8.268 (6.232)	0.0-30.0	0.351	0.010	20/20
Cognition	S1 history-informed	Cognition score (2015)	451	7.616 (2.417)	7.535 (2.494)	0.0-10.0	0.071	0.986	20/20
CES-D	S0 paired sensitivity	Hukou category (2018)	140	1.197 (0.404)	1.258 (0.442)	1.0-3.0	0.033	0.219	20/20
CES-D	S0 paired sensitivity	Work class category (2018)	578	1.303 (0.840)	1.565 (0.884)	0.0-4.0	0.040	0.328	20/20
CES-D	S0 paired sensitivity	Poor self-rated health (2018)	1	0.149 (0.356)	0.100 (0.308)	0.0-1.0	0.308	0.000	20/20
CES-D	S0 paired sensitivity	IADL limitation count (2018)	64	0.155 (0.582)	0.195 (0.586)	0.0-5.0	0.087	0.252	20/20
CES-D	S0 paired sensitivity	Fair/poor vision (2018)	4	0.765 (0.424)	0.875 (0.333)	0.0-1.0	0.152	0.989	20/20
CES-D	S0 paired sensitivity	Fair/poor hearing (2018)	4	0.560 (0.496)	0.500 (0.503)	0.0-1.0	0.243	1.027	20/20
CES-D	S0 paired sensitivity	Cognition score (2018)	815	7.316 (2.403)	6.806 (2.562)	0.0-10.0	0.071	0.737	20/20
CES-D	S0 paired sensitivity	CES-D-10 score (2018)	274	6.996 (5.693)	8.045 (6.131)	0.0-30.0	0.230	0.101	20/20

Outcome	MI specification	Imputed variable	Missing, n	Observed mean (SD)	Imputed mean (SD)	Imputed range	SD of means across M=20	Trace shift / SD	Datasets meeting diagnostic checks
CES-D	S1 history-informed	Hukou category (2018)	140	1.197 (0.404)	1.271 (0.456)	1.0-3.0	0.031	1.202	20/20
CES-D	S1 history-informed	Work class category (2018)	578	1.303 (0.840)	1.552 (0.873)	0.0-4.0	0.045	0.717	20/20
CES-D	S1 history-informed	Poor self-rated health (2018)	1	0.149 (0.356)	0.100 (0.308)	0.0-1.0	0.308	0.650	20/20
CES-D	S1 history-informed	IADL limitation count (2018)	64	0.155 (0.582)	0.212 (0.634)	0.0-5.0	0.069	1.083	20/20
CES-D	S1 history-informed	Fair/poor vision (2018)	4	0.765 (0.424)	0.787 (0.412)	0.0-1.0	0.233	0.429	20/20
CES-D	S1 history-informed	Fair/poor hearing (2018)	4	0.560 (0.496)	0.400 (0.493)	0.0-1.0	0.188	0.265	20/20
CES-D	S1 history-informed	Cognition score (2018)	815	7.316 (2.403)	6.818 (2.580)	0.0-10.0	0.067	0.040	20/20
CES-D	S1 history-informed	CES-D-10 score (2018)	274	6.996 (5.693)	8.237 (6.166)	0.0-30.0	0.274	0.471	20/20

Notes: MI, multiple imputation. S0 denotes the paired eight-target FCS sensitivity; S1 uses the identical framework and adds the same-domain 2015 outcome as a ninth historical auxiliary. Nominal variables were imputed using stochastic multinomial-logistic models, binary variables using stochastic logistic models, and bounded scores/counts using predictive mean matching with k=5. Each specification used M=20 completed datasets. Telework and 2020 outcomes were predictors but were not imputed, and observed target values were preserved. Trace shift/SD is descriptive only. *For cognition S1 poor self-rated health, one missing value produced identical draws, so the ratio is undefined.

Supplementary Table S4. Missingness and observation/completeness diagnostics

Diagnostic	Nontelework observed/complete	Telework observed/complete	Difference, percentage points	Telework OR for availability (95% CI)	P	Model
Cognition fixed-target: 2015 observed	4257/5141 (82.80%)	277/366 (75.68%)	-7.12	0.712 (0.556-0.913)	.007480	Age + 2020 sex
Cognition fixed-target: 2018 observed	3441/5141 (66.93%)	264/366 (72.13%)	+5.20	1.302 (1.013-1.674)	.039106	Age + 2020 sex
Cognition fixed-target: 2020 observed	4446/5141 (86.48%)	317/366 (86.61%)	+0.13	1.067 (0.775-1.470)	.689131	Age + 2020 sex
CES-D-10 complete: 2015 observed	4118/5141 (80.10%)	272/366 (74.32%)	-5.78	0.760 (0.597-0.967)	.025722	Age + 2020 sex
CES-D-10 complete: 2018 observed	3994/5141 (77.69%)	286/366 (78.14%)	+0.45	1.021 (0.775-1.346)	.880399	Age + 2020 sex
CES-D-10 complete: 2020 observed	4545/5141 (88.41%)	327/366 (89.34%)	+0.94	1.071 (0.760-1.508)	.695386	Age + 2020 sex
Cognition fixed-target: complete 3-wave	2863/5141 (55.69%)	216/366 (59.02%)	+3.33	1.196 (0.959-1.492)	.111269	Age + 2020 sex
CES-D-10 complete: complete 3-wave	3256/5141 (63.33%)	225/366 (61.48%)	-1.86	0.936 (0.743-1.179)	.573759	Age + 2020 sex
Complete baseline-rich covariates	4119/5141 (80.12%)	269/366 (73.50%)	-6.62	0.688 (0.533-0.888)	.004188	Unadjusted
Complete baseline-full covariates incl. work class	3538/5141 (68.82%)	254/366 (69.40%)	+0.58	1.028 (0.812-1.301)	.820988	Unadjusted

Notes: Percentages use the exposure-eligible backbone (nontelework N=5141; telework N=366). The OR compares the odds of being observed/complete in teleworkers with nonteleworkers; OR<1 indicates lower availability among teleworkers. For outcome and complete-three-wave rows, ORs are from the full-N descriptive model adjusted for harmonized age and 2020 sex. For baseline-completeness rows, ORs are unadjusted. All observation models used community-clustered robust covariance and are descriptive, not causal retention models. Missingness was nonmonotone, and complete-case deletion may introduce additional selection. Outcome-observation weighting has a positivity limitation that includes a 596-person missing-education stratum. For the CES-D ≥8-item prorated sensitivity, the 2018 observation model showed exposure/target separation in the baseline-rich and baseline-full-with-workclass specifications; these models were therefore not estimated. These diagnostics do not establish that the missing-at-random assumption holds.

Supplementary Table S5. Full sensitivity-analysis results for cognition and CES-D

Analysis family	Outcome	Analysis	N	Telework, n	Beta (95% CI)	P	Interpretation / diagnostic note
Primary analysis	Cognition	Overlap-weighted MI20 ANCOVA, community-clustered SE	4184	270	0.278 (0.040 to 0.515)	.021953	Primary cognition estimate; positive association with limited robustness in longitudinal analyses. Primary CES-D-10 estimate; negative association with broadly consistent direction across major sensitivity analyses. Eight-target FCS comparator; does not replace the primary analysis. Adds 2015 cognition as an imputation auxiliary; estimate changed minimally relative to matched S0. Eight-target FCS comparator; does not replace the primary analysis. Adds 2015 CES-D-10 as an imputation auxiliary; estimate changed minimally relative to matched S0. Consistent with the primary estimate; descriptive P outside the Holm family. Consistent with the primary estimate; descriptive P outside the Holm family. Consistent with the primary estimate; descriptive P outside the Holm family. Direction stable; confidence interval includes the null; descriptive P outside the Holm family. Direction stable; confidence interval includes the null; descriptive P outside the Holm family.
Primary analysis	CES-D	Overlap-weighted MI20 ANCOVA, community-clustered SE	4262	278	-0.792 (-1.317 to -0.266)	.003215	
Missing-data sensitivity	Cognition	S0 paired FCS, OW clustered ANCOVA	4184	270	0.282 (0.045 to 0.518)	.019688	
Missing-data sensitivity	Cognition	S1 history-informed paired FCS, OW clustered ANCOVA	4184	270	0.286 (0.051 to 0.522)	.017256	
Missing-data sensitivity	CES-D	S0 paired FCS, OW clustered ANCOVA	4262	278	-0.799 (-1.327 to -0.271)	.003124	
Missing-data sensitivity	CES-D	S1 history-informed paired FCS, OW clustered ANCOVA	4262	278	-0.789 (-1.316 to -0.262)	.003414	
Core sensitivity	Cognition	Primary OW MI20 (HC3 sensitivity)	4184	270	0.278 (0.035 to 0.520)	.024811	
Core sensitivity	Cognition	Common-support restricted OW (MI20)	3809-3959	268-270	0.272 (0.034 to 0.510)	.025017	
Core sensitivity	Cognition	Unweighted baseline-adjusted ANCOVA (MI20)	4184	270	0.263 (0.015 to 0.511)	.037916	
Core sensitivity	Cognition	Complete-case re-estimated OW	2677	206	0.245 (-0.003 to 0.493)	.052448	
Core sensitivity	Cognition	Complete-case unweighted ANCOVA	2865	215	0.215 (-0.039 to 0.469)	.097539	

Analysis family	Outcome	Analysis	N	Telework, n	Beta (95% CI)	P	Interpretation / diagnostic note
Core sensitivity	CES-D	Primary OW MI20 (HC3 sensitivity)	4262	278	-0.792 (-1.308 to -0.275)	.002665	Consistent with the primary estimate; descriptive P outside the Holm family.
Core sensitivity	CES-D	Common-support restricted OW (MI20)	4043-4164	277-278	-0.783 (-1.309 to -0.257)	.003628	Consistent with the primary estimate; descriptive P outside the Holm family.
Core sensitivity	CES-D	Unweighted baseline-adjusted ANCOVA (MI20)	4262	278	-0.880 (-1.424 to -0.336)	.001593	Consistent with the primary estimate; descriptive P outside the Holm family.
Core sensitivity	CES-D	Complete-case re-estimated OW	2718	214	-0.810 (-1.404 to -0.216)	.007670	Consistent with the primary estimate; descriptive P outside the Holm family.
Core sensitivity	CES-D	Complete-case unweighted ANCOVA	3331	240	-0.872 (-1.442 to -0.303)	.002761	Consistent with the primary estimate; descriptive P outside the Holm family.
Population restriction	Cognition	Exclude proxy interviews	4181	270	0.279 (0.042 to 0.516)	.021403	Essentially unchanged from the primary estimate.
Population restriction	Cognition	Exclude fair/poor vision or hearing	774-776	74	0.791 (0.360 to 1.222)	.000356	Narrow sensory-intact subgroup; estimate differs from the primary analysis.
Population restriction	Cognition	Exclude sex mismatch	4175	269	0.280 (0.042 to 0.517)	.021266	Essentially unchanged from the primary estimate.
Population restriction	Cognition	Restrict to working in 2018	3819	255	0.295 (0.055 to 0.534)	.015906	Employment-restricted sensitivity; direction consistent with the primary estimate.
Population restriction	Cognition	Restrict to 2020 employees	3039	219	0.296 (0.022 to 0.570)	.034521	Employment-restricted sensitivity; direction consistent with the primary estimate.
Population restriction	Cognition	Exclude nonagricultural self-employed	3337	226	0.289 (0.015 to 0.562)	.038692	Employment-restricted sensitivity; direction consistent with the primary estimate.
Population restriction	Cognition	0-11 cognition in measurement-clean subset	604-606	70-71	0.912 (0.476 to 1.347)	.000050	Very narrow measurement-clean subgroup; estimate larger than in the primary analysis.
Population restriction	CES-D	Exclude proxy interviews	4259	278	-0.795 (-1.320 to -0.270)	.003092	Essentially unchanged from the primary estimate.

Analysis family	Outcome	Analysis	N	Telework, n	Beta (95% CI)	P	Interpretation / diagnostic note
Population restriction	CES-D	Exclude fair/poor vision or hearing	790-793	75	-0.212 (-1.042 to 0.617)	.614879	Narrow sensory-intact subgroup; estimate differs from the primary analysis.
Population restriction	CES-D	Exclude sex mismatch	4253	277	-0.781 (-1.305 to -0.256)	.003640	Essentially unchanged from the primary estimate.
Population restriction	CES-D	Restrict to working in 2018	3897	264	-0.753 (-1.276 to -0.230)	.004896	Employment-restricted sensitivity; direction consistent with the primary estimate.
Population restriction	CES-D	Restrict to 2020 employees	3101	225	-0.933 (-1.505 to -0.360)	.001466	Employment-restricted sensitivity; direction consistent with the primary estimate.
Population restriction	CES-D	Exclude nonagricultural self-employed	3401	231	-0.958 (-1.525 to -0.391)	.000970	Employment-restricted sensitivity; direction consistent with the primary estimate.
Outcome definition	Cognition	Cognition process-scored 0-10	3417	245	0.172 (0.001 to 0.343)	.048465	Positive estimate, with the lower confidence limit near zero.
Outcome definition	Cognition	Cognition 0-11 incl. figure copying	3246	242	0.256 (0.000 to 0.512)	.049641	Positive estimate, with the lower confidence limit near zero.
Outcome definition	Cognition	Cognition change score (2020-2018)	4184	270	0.278 (-0.022 to 0.577)	.069000	Positive direction; confidence interval includes the null.
Outcome definition	Cognition	Standalone serial-7 process score	3749	253	0.077 (-0.052 to 0.206)	.241788	No clear association in this outcome-definition sensitivity.
Outcome definition	CES-D	CES-D >=8-item prorated	4208	278	-0.708 (-1.232 to -0.185)	.008082	Direction consistent with the primary CES-D-10 estimate.
Outcome definition	CES-D	CES-D change score (2020-2018)	4262	278	-0.792 (-1.402 to -0.181)	.011164	Direction consistent with the primary CES-D-10 estimate.
Design alternative	Cognition	1:1 propensity-score matching	504-514	252-257	0.232 (-0.096 to 0.560)	.165040	Direction consistent; prespecified post-match balance target not met.
Design alternative	Cognition	Official 2020 individual-weight regression	3684	225	0.217 (-0.101 to 0.535)	.179942	Positive but imprecise under official survey weighting.
Design alternative	Cognition	Long-format random-intercept time x telework mixed model	4184	270	0.058 (-0.255 to 0.371)	.715719	Mixed-model cognition estimate near the null.

Analysis family	Outcome	Analysis	N	Telework, n	Beta (95% CI)	P	Interpretation / diagnostic note
Design alternative	CES-D	1:1 propensity-score matching	516-528	258-264	-0.785 (-1.627 to 0.058)	.067764	Direction consistent; prespecified post-match balance target not met.
Design alternative	CES-D	Official 2020 individual-weight regression	3748	231	-1.259 (-1.924 to -0.595)	.000222	Negative association under official survey weighting.
Design alternative	CES-D	Long-format random-intercept time x telework mixed model	4262	278	-0.819 (-1.481 to -0.157)	.015331	Mixed-model CES-D-10 estimate consistent in direction.
Geography sensitivity	Cognition	Unweighted community fixed-effects ANCOVA	4184	270	0.251 (-0.033 to 0.536)	.083388	Positive but imprecise under community fixed effects; time-varying local confounding may remain.
Geography sensitivity	CES-D	Unweighted community fixed-effects ANCOVA	4262	278	-0.917 (-1.510 to -0.324)	.002521	Negative association within communities; time-varying local confounding may remain.
Outcome-observation weighting	Cognition	Overlap weight x stabilized outcome-observation weight	4184	270	0.284 (0.047 to 0.521)	.018935	Essentially unchanged within empirical observation support; positivity limitation and residual selection bias may remain.
Outcome-observation weighting	CES-D	Overlap weight x stabilized outcome-observation weight	4262	278	-0.806 (-1.331 to -0.281)	.002686	Essentially unchanged within empirical observation support; positivity limitation and residual selection bias may remain.

Notes: Beta is the telework coefficient or the analysis-specific contrast shown. The primary rows use the primary MI20 completed datasets. S0 and S1 are paired FCS missing-data sensitivities; S1 differs from S0 only by adding the corresponding 2015 same-domain outcome as an imputation auxiliary. Except for the two co-primary outcomes, sensitivity P values are descriptive and are outside the Holm multiplicity family. The CES-D ≥8-item prorated analysis required observed prorated scores in both 2018 and 2020: all 4,262 primary CES-D participants had the 2020 prorated score, whereas 54 lacked the 2018 prorated baseline, yielding N=4,208. PSM is retained as a sensitivity despite residual post-match imbalance: maximum |SMD|=0.168 for cognition and 0.151 for CES-D-10. The outcome-observation-weighting rows have a structural positivity limitation involving a 596-person missing-education stratum with zero outcome-sample representation. Longitudinal 2015-2020 trajectory results and formal sex-interaction results are reported separately in subsequent supplementary tables.

Supplementary Table S6. Full longitudinal trajectory analyses

Outcome definition	Analysis	Contrast	N	Telework, n	Communities, n	β (SE)	95% CI	df	P
Cognition fixed-target 0-10	Complete3	2015-2018	3079	216	431	0.064 (0.155)	-0.240 to 0.367	430	.680972
Cognition fixed-target 0-10	Complete3	2018-2020	3079	216	431	0.001 (0.154)	-0.302 to 0.304	430	.994241
Cognition fixed-target 0-10	Complete3	2015-2020	3079	216	431	0.065 (0.135)	-0.201 to 0.331	430	.632742
Cognition fixed-target 0-10	All-available participant FE	2015-2018	4660	297	438	0.075 (0.142)	-0.204 to 0.354	437	.597228
Cognition fixed-target 0-10	All-available participant FE	2018-2020	4660	297	438	0.027 (0.138)	-0.244 to 0.297	437	.846849
Cognition fixed-target 0-10	All-available participant FE	2015-2020	4660	297	438	0.102 (0.123)	-0.140 to 0.343	437	.409182
Cognition process-scored 0-10	Complete3	2015-2018	3079	216	431	0.228 (0.111)	0.009 to 0.446	430	.041434
Cognition process-scored 0-10	Complete3	2018-2020	3079	216	431	-0.040 (0.108)	-0.253 to 0.172	430	.708175
Cognition process-scored 0-10	Complete3	2015-2020	3079	216	431	0.187 (0.100)	-0.010 to 0.384	430	.062922
Cognition process-scored 0-10	All-available participant FE	2015-2018	4660	297	438	0.263 (0.102)	0.062 to 0.465	437	.010463
Cognition process-scored 0-10	All-available participant FE	2018-2020	4660	297	438	-0.060 (0.100)	-0.256 to 0.137	437	.549383
Cognition process-scored 0-10	All-available participant FE	2015-2020	4660	297	438	0.204 (0.094)	0.019 to 0.388	437	.030658
CES-D-10 complete	Complete3	2015-2018	3481	225	430	0.226 (0.332)	-0.427 to 0.878	429	.497280
CES-D-10 complete	Complete3	2018-2020	3481	225	430	-0.703 (0.328)	-1.348 to -0.059	429	.032600
CES-D-10 complete	Complete3	2015-2020	3481	225	430	-0.478 (0.355)	-1.176 to 0.220	429	.179132
CES-D-10 complete	All-available participant FE	2015-2018	4766	311	437	0.386 (0.317)	-0.237 to 1.008	436	.224023
CES-D-10 complete	All-available participant FE	2018-2020	4766	311	437	-0.746 (0.288)	-1.314 to -0.179	436	.009991
CES-D-10 complete	All-available participant FE	2015-2020	4766	311	437	-0.361 (0.330)	-1.010 to 0.288	436	.275075
CES-D >=8-item prorated	Complete3	2015-2018	3887	242	432	0.081 (0.334)	-0.574 to 0.737	431	.807931
CES-D >=8-item prorated	Complete3	2018-2020	3887	242	432	-0.618 (0.312)	-1.232 to -0.005	431	.048230
CES-D >=8-item prorated	Complete3	2015-2020	3887	242	432	-0.537 (0.339)	-1.204 to 0.130	431	.114093
CES-D >=8-item prorated	All-available participant FE	2015-2018	4925	316	438	0.232 (0.310)	-0.376 to 0.841	437	.453673
CES-D >=8-item prorated	All-available participant FE	2018-2020	4925	316	438	-0.629 (0.283)	-1.185 to -0.073	437	.026662

Outcome definition	Analysis	Contrast	N	Telework, n	Communities, n	β (SE)	95% CI	df	P
CES-D $\geq$ 8-item prorated	All-available participant FE	2015-2020	4925	316	438	-0.397 (0.317)	-1.020 to 0.227	437	.211665

Notes: β is the differential longitudinal outcome change associated with later 2020 telework status relative to nontelework for the indicated wave contrast. Complete3 analyses include participants with the relevant outcome observed in 2015, 2018, and 2020. All-available participant fixed-effects (FE) analyses include participants with at least two observed waves and use all available outcome observations. The 2015-2018 contrast represents the pre-pandemic trajectory, the 2018-2020 contrast the pandemic-period trajectory, and the 2015-2020 contrast the overall three-wave interval. Higher cognition scores indicate better performance; higher CES-D scores indicate more depressive symptoms. All models are unweighted longitudinal extensions with community-clustered robust standard errors. These supplementary trajectory analyses are not co-primary tests and no multiplicity adjustment was applied. They use the harmonized three-wave longitudinal file rather than the MI20/overlap-weighted primary-analysis files and should be interpreted as associative trajectory analyses, not causal effects.

Supplementary Table S7. Full telework-by-sex interaction results

Outcome	Contrast	N	Telework represented, n	β (SE)	Barnard-Rubin df	95% CI	P
Cognition	Male marginal association	4184	186	0.348 (0.141)	415.1	0.071 to 0.626	.014085
Cognition	Female marginal association	4184	84	0.124 (0.206)	429.8	-0.280 to 0.529	.546035
Cognition	Female - male interaction	4184	270	-0.224 (0.240)	429.0	-0.695 to 0.248	.351891
CES-D	Male marginal association	4262	191	-0.868 (0.288)	429.6	-1.434 to -0.302	.002724
CES-D	Female marginal association	4262	87	-0.626 (0.512)	425.8	-1.633 to 0.380	.222134
CES-D	Female - male interaction	4262	278	0.242 (0.562)	423.7	-0.863 to 1.348	.666872

Notes: Marginal associations are the estimated telework coefficients within the indicated sex from the formal telework-by-sex interaction model. The interaction contrast is female minus male. Models use the primary MI20 overlap-weighted baseline-adjusted ANCOVA specification, community-clustered covariance, Rubin pooling, and Barnard-Rubin degrees of freedom. The formal interaction P values are 0.351891 for cognition and 0.666872 for CES-D; therefore, there is no clear evidence that the telework association differs by sex for either outcome. A statistically significant marginal association in one sex and a nonsignificant marginal association in the other must not be interpreted as evidence of sex heterogeneity. All estimates are associative and should not be interpreted causally.

Supplementary Table S8. PSM balance and HC3 sensitivity diagnostics

Diagnostic	Cognition	CES-D	Interpretation / diagnostic note
PSM estimate, beta (95% CI)	0.232 (-0.096 to 0.560)	-0.785 (-1.627 to 0.058)	Direction consistent, but imprecise
PSM P value	.165040	.067764	Descriptive sensitivity P value
Matched sample N across completed datasets	504-514	516-528	1:1 propensity-score matching
Matched teleworkers, n across completed datasets	252-257	258-264	One matched control per matched teleworker
Matched fraction	0.941	0.939	Matched fraction among eligible teleworkers
Maximum post-match SMD	0.168	0.151	Prespecified balance target: <0.10
Post-match balance target met	No	No	PSM retained as sensitivity only
HC3 beta (95% CI)	0.278 (0.035 to 0.520)	-0.792 (-1.308 to -0.275)	Weighted HC3 covariance sensitivity
HC3 SE	0.123627	0.263346	Rubin-pooled across MI20
HC3 Barnard-Rubin df	3655.266	4193.902	Complete-data df based on N minus model parameters
HC3 P value	.024811	.002665	Sensitivity inference; not in Holm family

Notes: PSM, propensity-score matching; SMD, standardized mean difference; HC3, heteroskedasticity-consistent covariance estimator type 3; MI20, 20 completed multiple-imputation datasets. The 1:1 PSM analysis was retained as a sensitivity analysis because the predefined post-match |SMD|<0.10 target was not achieved. Weighted HC3 and the primary community-clustered analyses were calculated in Stata/SE 15.1, with the weighted sandwich covariance specified in Mata. No result was selected according to statistical significance.

Supplementary Table S9. Expanded outcome-model covariate sensitivity

Outcome	Model specification	beta	SE	95% CI	P value
Cognition	Primary overlap-weighted model	0.278	0.121	0.040 to 0.515	0.022
Cognition	Expanded full PS-covariate outcome model	0.278	0.117	0.048 to 0.507	0.018
CES-D-10	Primary overlap-weighted model	-0.792	0.267	-1.317 to -0.266	0.003
CES-D-10	Expanded full PS-covariate outcome model	-0.792	0.267	-1.315 to -0.268	0.003

Note. The expanded models retained the original overlap weights and additionally entered ADL and IADL limitations, fair/poor vision, fair/poor hearing, and the other-domain baseline outcome, together with the primary covariates. P values shown here are unadjusted sensitivity-analysis P values; Holm adjustment applies to the two co-primary analyses, not to sensitivity analyses.

Supplementary Table S10. Community exposure support and mixed-community sensitivity analysis

Outcome	Total communities	Non-TW only	Mixed	TW only	Restricted N (%)	Teleworkers retained	beta	95% CI	P
Cognition	436	260	169	7	1853 (44.29%)	261/270 (96.67%)	0.226	-0.025 to 0.478	.077650
CES-D-10	436	259	170	7	1906 (44.72%)	269/278 (96.76%)	-0.870	-1.424 to -0.315	.002304

Note. Mixed communities contained at least one teleworker and one nonteleworker. Propensity scores and overlap weights were re-estimated after restriction in every completed dataset. The restricted estimand pertains to participants in communities with empirical representation of both exposure states; restriction alone does not eliminate community-level confounding.

Supplementary Table S11. Restricted-sample diagnostics across 20 completed datasets

Diagnostic	Cognition	CES-D-10
Restricted sample	1853; 261 teleworkers	1906; 269 teleworkers
Maximum weighted SMD	<0.001 in every dataset	<0.001 in every dataset
Overall ESS, mean (range)	752.3 (749.9-754.8)	778.7 (777.2-780.5)
Teleworker ESS, mean (range)	244.2 (243.7-244.6)	251.7 (251.4-252.1)
Nonteleworker ESS, mean (range)	818.0 (810.3-827.8)	859.3 (853.2-863.6)
Propensity-score range across datasets	0.001190-0.750368	0.002495-0.696472
Validity	All PS/OW strictly between 0 and 1	All PS/OW strictly between 0 and 1

Supplementary Table S12. Specifications of analyses presented in Figure 2

Figure 2 model / outcome	Sample and target population	Weight	PS treatment	Outcome model	Community / covariance	Pooling
Primary overlap-weighted ANCOVA	Cognition N=4,184, TW=270; CES-D-10 N=4,262, TW=278; 436 communities. Covariate-overlap population.	Primary OW	PS estimated in each MI dataset	2018 baseline-adjusted ANCOVA with primary covariates.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
HC3 covariance	Same primary samples/target. Cognition N=3,809–3,959; CES-D-10 N=4,043–4,164. Individual PS-range-intersection population.	Primary OW	Primary PS retained	Same primary ANCOVA.	Weighted HC3 sandwich; not community-clustered.	Rubin + Barnard-Rubin; complete-data df N-18
Common-support OW	Cognition N=1,853, TW=261, 169 communities; CES-D-10 N=1,906, TW=269, 170 communities. Mixed-community covariate-overlap population.	Primary OW inside empirical PS intersection; zero outside	Primary PS retained; no re-estimation	Same primary ANCOVA.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Mixed-exposure-community OW	Full outcome-specific cohorts. Cognition N=2,677, TW=206; CES-D-10 N=2,718, TW=214. Complete-case covariate-overlap population.	OW re-estimated after restriction	PS re-estimated in each restricted MI dataset	Same primary ANCOVA after restriction.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Unweighted ANCOVA	Full outcome-specific cohorts. Cognition N=2,677, TW=206; CES-D-10 N=2,718, TW=214. Complete-case covariate-overlap population.	None	Not applicable	Unweighted baseline-adjusted ANCOVA.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Complete-case OW	Full outcome-specific cohorts with community indicators.	OW re-estimated	PS/OW re-estimated in complete cases	Primary ANCOVA in complete cases.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Community fixed effects	Full-sample covariate-overlap population. Cognition N=4,184; CES-D-10 N=4,262. Primary cohort within outcome-observed support.	None; unweighted	Not applicable	Baseline-adjusted ANCOVA + categorical community indicators.	Community indicators + community-clustered robust covariance.	Rubin + Barnard-Rubin
Expanded full-PS covariates	Cognition N=3,684, TW=225; CES-D-10 N=3,748, TW=231. Nonmissing official-weight cohort.	Primary OW retained	Primary PS retained	Weighted outcome model with full PS covariate set.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Outcome-observation weighting	Cognition N=3,039, TW=219; CES-D-10 N=3,101, TW=225. Restricted covariate-overlap population.	Primary OW × stabilized observation weight	Primary PS retained; separate observation model	Primary ANCOVA with combined final weight.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Official 2020 weight	Cognition N=3,337, TW=226; CES-D-10 N=3,401, TW=231. Restricted covariate-overlap population.	Official CHARLS individual weight alone	Study PS/OW not used	Primary ANCOVA with official survey weight.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
2020 employees only		New OW after restriction	PS re-estimated; new OW	Primary ANCOVA in restricted sample.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Exclude nonagricultural self-employed		New OW after restriction	PS re-estimated; new OW	Primary ANCOVA in restricted sample.	Community-clustered robust covariance.	Rubin + Barnard-Rubin

Figure 2 model / outcome	Sample and target population	Weight	PS treatment	Outcome model	Community / covariance	Pooling
Working in 2018	Cognition N=3,819, TW=255; CES-D-10 N=3,897, TW=264. Restricted covariate-overlap population.	New OW after restriction	PS re-estimated; new OW	Primary ANCOVA in restricted sample.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Process-scored cognition	Cognition N=3,417, TW=245.	Primary OW retained	Primary PS retained	Primary model with alternative cognition outcome.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Cognition 0–11 + figure copying	Cognition N=3,246, TW=242.	Primary OW retained	Primary PS retained	Primary model with alternative cognition outcome.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
CES-D ≥8-item prorated	CES-D-10 N=4,208, TW=278.	Primary OW retained	Primary PS retained	Primary model with prorated CES-D outcome.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Change score (2020–2018)	Cognition N=4,184; CES-D-10 N=4,262.	Primary OW retained	Primary PS retained	Outcome model for 2020-minus-2018 change.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Serial-7 process score	Cognition N=3,749, TW=253.	Primary OW retained	Primary PS retained	Primary model with serial-7 process score. 2018–2020 long format: time + telework + time×telework + baseline demographic/employment/health covariates; participant random intercept; REML.	Community-clustered robust covariance.	Rubin + Barnard-Rubin
Random-intercept mixed model	Cognition N=4,184; CES-D-10 N=4,262. Two-wave longitudinal population.	None; unweighted	Not applicable		Participant random-intercept covariance; no community-clustered robust covariance.	Not MI20-pooled; REML
1:1 PSM†	Cognition matched N=504–514; CES-D-10 matched N=516–528 across MI datasets.	Matched sample	PS estimated for matching; no OW	Matched-sample outcome analysis.	Matched-sample model.	Rubin + Barnard-Rubin

Note. PS, propensity score; OW, overlap weight; MI, multiple imputation; TW, teleworker; REML, restricted maximum likelihood. Sensitivity P values were descriptive and were not treated as additional confirmatory tests. The mixed-exposure-community analysis pertains to the covariate-overlap population within communities in which both exposure states were observed.