No substitute for in-person interaction: changing modes of social contact during the coronavirus pandemic and effects on the mental health of adults in the UK

Supplementary File

Contents

S1: Study details

Datasets used

The studies included are: Millennium Cohort Study (born 2000-02), Next Steps (born 1989-90), the1970 British Cohort Study, and the 1958 National Child Development Study (NCDS)

Fieldwork dates for the three surveys were:

Wave 1: 4-30 May 2020

Wave 2: 10 Sep- 16 Oct 2020

Wave 3: 1 Feb-21 Mar 2021

The waves 1 and 2 surveys were conducted as web surveys only. Wave 3 was primarily a web survey but involved a telephone phase in which a sub-set of web-survey non-respondents were invited to take part via telephone to explore mode effects. Further details of potential mode effects using the wave 3 data are described in Brown et al. 2021.

Target population and issued sample response rates:

The target population of each cohort are the individuals born in the specified birth period of the cohort who are alive and still residing in the UK. At the time the Wave 1 survey was conducted, mass postal mailings were not possible in the UK, so the survey invitations had to be sent via email, resulting in a selection of participants for whom the Centre for Longitudinal Studies held an email address, provided that they a) had not permanently withdrawn from the study b) were not 'permanently untraced' and c) were not known to have died. At Waves 2 and 3 it was possible to send invitations via post, so it was possible to include those for whom no email address was held.

The total response rate of all cohort members with respect to the target population was 20.8% in Wave 1, 27.7% in Wave 2 and 31.2% in Wave 3, and the corresponding response rates for cohort members with respect to the issued sample of 37.5%, 39.1% and 43.8% respectively. The response rates of cohort members within the issued samples are comparable to those of similar web surveys undertaken at similar times

Attrition and non-response weights

Non-response is common in longitudinal surveys. The CLS has developed comprehensive advice on how to deal with missing data, capitalising on the rich data cohort members provided over the years before their non-response to the COVID surveys. These predictors of non-response include sex, ethnicity, parental social class, number of rooms at home/persons per room, cognitive ability, early life mental health, voting, membership in organisations, internet access prior to web survey, consent for biomarkers, consent for data linkages, educational qualifications, economic activity, partnership status, psychological distress, BMI, self-rated health, smoking status, maternal mental health, social capital/social support, income, number of non-responses across all previous waves, response to the COVID-19 wave 1 survey, response to the COVID-19 wave 2 survey.

To correct for non-response in the COVID-19 Wave 1, 2 and 3 surveys and facilitate analysis in all cohorts, non-response weights are derived, with sample corresponding to the target population (those alive and living in the UK). The CLS examined the effectiveness of the derived non-response weights in restoring sample representativeness by conducting several analyses. In relation to the distribution of sex in each cohort, which is observed at baseline in virtually all cohort members, the extent of bias in the estimated percentage of female cohort members caused by non-response to

the COVID-19 surveys at waves 1, 2 and 3 varied across cohorts, but was substantial in most cases. However, the application of the non-response weights greatly reduced this bias in all cohorts, essentially completely eliminating it in the NCDS, BCS70 and MCS samples so that the sample representativeness with respect to this variable was restored.

S2: Correlations between social contact measures at each wave

	Met people outside household in-person	Talked to family/friend via video/phone	Contacted family/friend by electronic msg	Took part in an online community activity
May 20 survey				
Met people outside household in-person	**1.00**			
Talked to family/friend via video/phone	0.08	**1.00**		
Contacted family/friend by electronic msg	0.01	0.45	**1.00**	
Took part in an online community activity	0.03	0.19	0.22	**1.00**
Sep/Oct 20 survey				
Met people outside household in-person	**1.00**			
Talked to family/friend via video/phone	0.20	**1.00**		
Contacted family/friend by electronic msg	0.19	0.43	**1.00**	
Took part in an online community activity	0.07	0.15	0.18	**1.00**
Feb/Mar 20 survey				
Met people outside household in-person	**1.00**			
Talked to family/friend via video/phone	0.15	**1.00**		
Contacted family/friend by electronic msg	0.12	0.42	**1.00**	
Took part in an online community activity	0.05	0.13	0.17	**1.00**

S3: Details of questions and recodes (if any) of explanatory variables in the statistical models

Explanatory Variables	Question	Categories	Recode	Recode
Met people in-person	In the last seven days, did you meet up in-person with any of your family or friends who do not live with you	(1) Every day; (2) 4-6 days; (3) 2-3 days; (4) 1 day; (5) Never	(0) Everyday to 2-3 days; (1) 1 day to Never (for H2)	(5) Every day; (4) 4-6 days; (3) 2-3 days; (2) 1 day; (1) Never (for H3)
Talked to family/friend via video/phone	In the last seven days, did you talk to family or friends who do not live with you via phone or video calls	(1) Every day; (2) 4-6 days; (3) 2-3 days; (4) 1 day; (5) Never		
Contacted family/friend by electronic msg	In the last seven days, did you keep in contact with family or friends you do not live with by email or text or other electronic messaging	(1) Every day; (2) 4-6 days; (3) 2-3 days; (4) 1 day; (5) Never		
Took part in an online community activity	In the last seven days, did you take part in an online community activity, e.g. an online community group, online chat group, street or neighbourhood social media group	(1) Every day; (2) 4-6 days; (3) 2-3 days; (4) 1 day; (5) Never		
COVID symptoms	Have you experienced any of the following symptoms in the past 2 weeks? List of symptoms including: Fever, Cough Sore Throat, Shortness of Breath, etc	(0) No; (1) At least one symptom		
Tested positive COVID-19 infection	Have you been tested for Coronavirus (Yes/No) & What was the result of your coronavirus test?	(0) No Coronavirus test or Negative/Inconclusive/Waiting for results; (1) Positive		
Self-rated health	In general, would you say your health is...	(0) Excellent; (1) Very Good; (3) Good; (4) Fair; (5) Poor	(0) Excellent/Very Good; (1) Good to Poor	
Finances	Overall, how do you feel your current financial situation compares to before the Coronavirus outbreak	(1) I'm much worse off; (2) I'm a little worse off; (3) I'm about the same; (4) I'm a little better off; (5) I'm much better off		
Household size	How many people do you currently live with? Please include yourself.	(0) One; (1) More than one		
Keyworker	Are you a Key worker, or has your work been classified as critical to the Covid-19 response?	(1) Not in paid work; (2) Yes; (3) No		
Region/Country	NB: N Ireland not included at wave 1 of COVID-19 surveys so N Irish participants from later waves were dropped	(1) North of England: North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, (2) East of England, London, South East, South West, (3) Wales, (4) Scotland		
Pre-pandemic social media use (MCS only)	On a normal weekday, how many hours do you spend on social networking or messaging sites or apps on the internet such as Facebook, Twitter, WhatsApp, Instagram and Snapchat?	(1) None; (2) <30 minutes; (3) 30-60 minutes; (4) 1-2 hours; (5) 2-3 hours; (6) 3-5 hours; (7) 5-7 hours; (8) 7-10 hours (9) >10 hours	(0) less than 2 hours a day; (1) 2 or more hours a day	

S4: Statistical models

S4.1. Fixed effects models: details

(from the STATA xtreg manual available here: https://www.stata.com/manuals/xtxtreg.pdf)

$$y_i = \alpha + x_i\beta + \varepsilon_i \quad (0)$$

In a linear regression model (0), the Ordinary Least Squares (OLS) estimator is unbiased if x_i is exogenous, which requires the covariates (x_i) to be uncorrelated with the error term from the same time period (ε_i).

In Stata, *xtreg* fits regression models to panel data. In particular, *xtreg* with the *fe* option, fits fixed-effects models (by using the within regression estimator).

$$y_{it} = \alpha + x_{it}\beta + v_i + \varepsilon_{it} \quad (1)$$

In model (1), $v_i + \varepsilon_{it}$ are the error terms that typically in a fixed effects model, we are not interested in; instead, we want to obtain estimates of β. v_i is the unit-specific error term; it differs between units (individuals in this analysis), but for any particular unit (individual), its value is constant.

We can derive the mean $\bar{y}_{it}$, $\bar{x}_{it}$ and $\bar{\varepsilon}$ across the panel data to estimate from (1)

$$\bar{y}_{it} = \alpha + \bar{x}_{it}\beta + v_i + \bar{\varepsilon}_{it} \quad (2)$$

Subtracting (2) from (1) we get

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (3)$$

These three equations provide the basis for estimating β. In particular, *xtreg, fe* provides what is known as the fixed-effects estimator- also known as the within estimator- and amounts to using OLS to perform the estimation of (3). Strict exogeneity with respect to the idiosyncratic error term $\bar{\varepsilon}_{it}$ is still assumed, which requires the covariates x from every time period to be uncorrelated with ε from every time period. Time-varying confounding and reverse causation could bias the fixed effects estimates.

S4.2 Instrumental variables models: details

(from the STATA xtivreg manual available here: https://www.stata.com/manuals/xtxtreg.pdf)

An instrumental variable (IV) is a variable (say Z) that is highly correlated with one of the independent variables (say X) but is uncorrelated with the error term (e). The researcher uses an instrumental variable in case the model suffers from an endogeneity problem. In this study, it is possible that there are common causes of both in-person social contact and mental health, such as personality types or individual preferences. A valid instrument needs to satisfy several criteria (see Fig S1 for an illustration):

1. Z has a casual effect on X (relevance criteria);
2. Z affects the outcome variable Y only through X (Z does not have a direct influence on Y which is referred to as the exclusion restriction);
3. There is no confounding for the effect of Z on Y (exogeneity criteria).

Although IVs can control for confounding and measurement error in observational studies, they have some limitations. We must be careful when dealing with many confounders and also if the correlation between the IV and the exposure variables is small. Both weak instruments and confounders produce large standard error which results in imprecise and biased results.

The two-stage Ordinary Least Squares IV model is similar to a standard regression model, but with two regression models estimated. The first-stage regression (4) regresses the endogenous variable y_{it} with another independent variable x_{1it} in the model along with an instrumental variable z_{1it}.

The second-stage regression regresses the dependent variable Y_{it} with predicted endogenous variable y_{it} and independent variable x_{1it}.

In STATA, *xtivreg* with the *fe* option uses the two-stage least-squares within estimator for fitting panel-data models in which some of the right-hand-side covariates are endogenous.

Consider an equation of the form

$$y_{it} = x_{1it}\beta_1 + z_{1it}\beta_2 + u_{it} \quad (4)$$

$$Y_{it} = y_{it}\beta_3 + x_{1it}\beta_4 + v_{it} \quad (5)$$

Here Y_{it} is the dependent variable for the i^{th} observation for unit t, y_{it} represents the endogenous regressors, x_{1it} represents the included exogenous regressors, and z_{1it} represents the excluded exogenous regressors. x_{1it} and z_{1it} are collectively called the instruments. u_{it} and v_{it} are zero-mean error terms, and there is some correlation between u_i and v_{it} (resulting in the endogeneity of y_{it} in equation (5). The within estimator (FE2SLS) fits the model by removing the panel-level means from each variable in equation (4) (see *xtreg* in S1.1).

S5.1 Examining the exogeneity criteria

In this section, we examine the plausibility of the exogeneity criteria for the instrumental variable used in the analysis of in-person social contact: regional/country differences over time in pandemic related social restrictions. Ideally the exogenous (instrumental) variable is a random process that allocates individuals into groups of higher or lower in-person social contact. We argue that changing policies across the pandemic and between countries and regions of the UK represent a natural experiment that is similar to a random process.

The UK government and devolved authorities imposed a national lockdown on March 23, 2020, with residents across regions and different countries of the UK required to stay at home except for essential reasons.

The first stay-at-home order ended in England and Northern Ireland on 3 May 2020, ending two weeks later in Scotland on 29 May, and Wales on 1 June. In Figure 1 (main manuscript), we see that there were very little differences between regions and countries of the UK in the May 2020 survey in the mean levels of the in-person social contact variable, with most people at the highest (never) level. However, several English regions remained under "local lockdown" orders from May 2020 onwards because of a higher infection rate- these included regions in the North, North-West, North-East and the Midlands. No areas in the South of England were included in these local lockdown orders with the exception of Luton where the restrictions were revoked by 1 August 2020. These local lockdowns continued until 14 October 2020 when a new tier system was introduced across England. In addition to regional and tiered restrictions, a 'rule of 6' was implemented at various times during the pandemic in different regions and countries of the UK, that banned all social mixing between households. On September 14 in England and Scotland, gatherings of more than six people were illegal in any setting either indoors, outdoors, at home or a pub or restaurant. In Wales, people were permitted to meet in groups of six or less indoors (as long as they formed an extended household) while up to 30 people could meet outdoors.

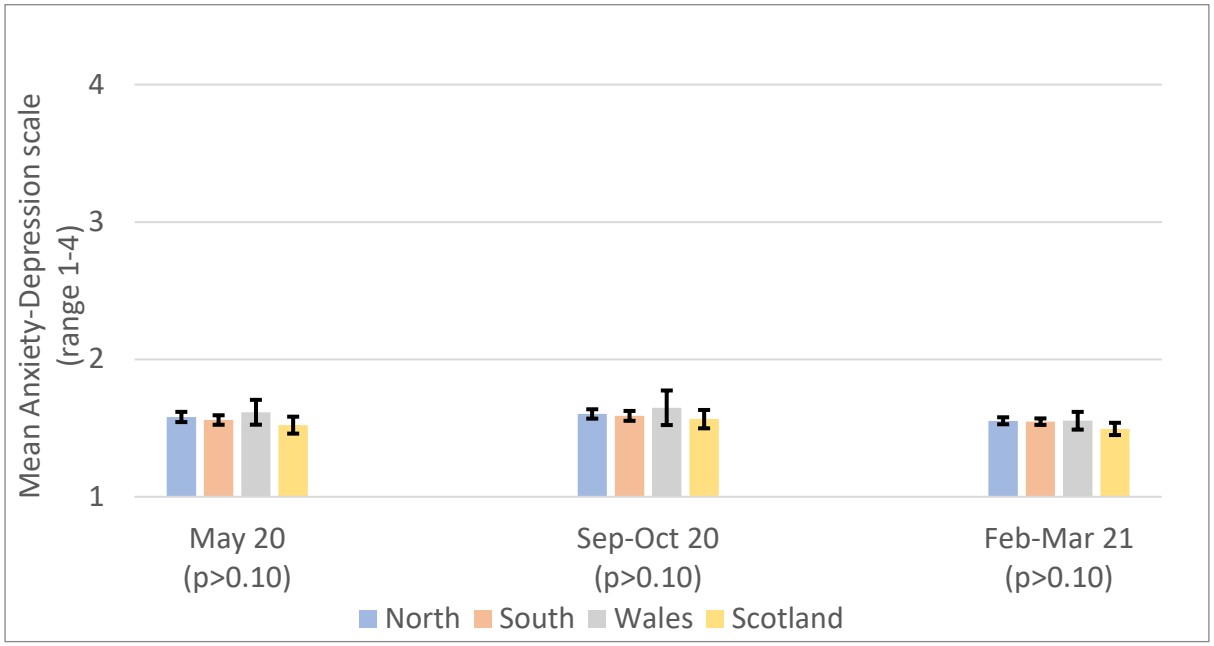
By the end of the October '20 survey, we not only see a reduction in mean in-person social contact variable (increased meeting of family/friends outside the home) compared to May '20, but also some regional and country differences with people in the North of England and Midlands most likely to never meet anyone outside their home, while those in Wales more likely to meet family/friends outside their home.

In England, household mixing was prohibited once again on 5th Jan 2021 until 29 March 2021, whereas the lifting of this ban was a little earlier in Scotland (on 12 March), Wales (13 March) when some limited outdoor gathering was allowed. Compared to the Sep-October '20 survey, we observe less frequent in-person social contact across all regions/countries.

S5.2 Examining the exclusion restriction criteria

The exclusion restriction criteria of IV analysis requires that the instrument (the region by time differences in this analysis) are only associated with the outcome (anxiety-depression) through the endogenous variable (in-person social contact). There can be no other pathways by which the instrument can affect anxiety-depression. While there is no direct test of the exclusion restriction criteria, an analysis of the association of the instrument with the outcome shown in S2.4 below suggests that there is no association between the instrument and the outcome variable. As the instrument is causally associated with the endogenous variable (see section S2.1), we believe this present strong evidence for that the exclusion restriction criteria of IV analysis holds for the IV analysis.

S5.3 Country/region difference in mean (95% confidence intervals) anxiety-depression: figure



S6: Distribution of respondents at waves 1, 2 and 3 of COVID-19 surveys

Non-response survey weights

The Centre for Longitudinal Studies (CLS) team developed comprehensive advice on how to deal with missing data in the COVID-19 surveys. They derived non-response survey weights to correct for non-response in the COVID-19 Wave 1, 2 and 3 surveys that also facilitate analysis in all the cohorts comprising the COVID-19 surveys. The target population of the non-response weights are individuals born in the specified birth period of the cohort who are alive and still residing in the UK.

The CLS team examined how effective the derived non-response weights were in restoring sample representativeness to the target population. They demonstrated that the bias caused by non-response in the COVID-19 surveys was greatly reduced in relation to a number of key variables such as sex, persons per room, number of rooms, and psychological distress, when the non-response weights were taken into account.

Table S6.1: Distribution of the respondents in the COVID-19 survey in waves 1 to 3 by unit and item-non-response

	wave 1	wave 2	wave 3	n obs
Respondents in survey	13,948	18,532	16,199	48,679
Respondents with non-zero w3 weights (unit non-response at waves 1 and 2)	11,581	15,340	16,011	42,932
% missing with non-zero w3 weights	17.0%	17.2%	1.2%	11.8%
Respondents with non-zero w3 weights & no missing data (item non-response at any wave)	10,146	14,309	15,008	39,463
% missing with non-zero w3 weights & no missing data	12.4%	6.7%	6.3%	8.1%

Table S6.1 shows how analytical sample of 39,463 observations (used in the fixed-effect regression models in Table 2) was derived from respondents to the COVID-19 surveys at waves 1, 2 and 3. The row "% missing with non-zero w3 weights" refers to unit non-response at waves 1 and 2, conditional on having a valid (non-response) weight at wave 3. The row "% missing with non-zero w3 weights & no missing data" refers to item non-response as a proportion of the respondents at a particular wave with non-zero w3 weights.

We can observe that unit non-response at waves 1 and 2 was much larger than item non-response. This was because most of the unit non-response at waves 1 and 2 was due to the addition of respondents from Northern Ireland at wave 3, who were missing from earlier waves. In relation to the unit non-response, the variable with largest amount of missing data at any wave was the mental health variable. Moreover, respondents with poor mental health were also more likely to have missing data on other covariates. However, after taking into account the wave 3 weights, poor mental health no longer predicted greater chances of having missing data on the other covariates. This suggests that the non-response weights were useful not just for accounting for unit non-response, but also for item non-response.

S7 Fixed effect models estimates (95% confidence intervals) of anxiety-depression with interactions between age-cohort and frequency of social contact: table

	Model 1: cohort*met in-person	Model 2: cohort*talked by video/phone	Model 3: cohort*contact by email/txt	Model 4: cohort*online community activity
Met people in-person (ref: 4-6 days)				
Every day	0.01 (-0.02, 0.05)	0.02 (-0.01, 0.05)	0.02 (-0.01, 0.05)	0.02 (-0.01, 0.05)
2-3 days	0.02 (-0.01, 0.04)	**0.02 (0.004 0.05)**	**0.02 (0.003, 0.05)**	**0.02 (0.003, 0.05)**
1 day	0.02 (-0.01, 0.04)	**0.04 (0.02, 0.06)**	**0.04 (0.02, 0.06)**	**0.04 (0.02, 0.06)**
Never	0.03 (0.01, 0.06)	**0.04 (0.02, 0.06)**	**0.04 (0.02, 0.06)**	**0.04 (0.02, 0.06)**
Talked to family/friend via video/phone (ref: Every day)				
4-6 days	-0.003 (-0.02, 0.02)	-0.01 (-0.03, 0.01)	-0.004 (-0.02, 0.02)	-0.003 (-0.02, 0.02)
2-3 days	0.002 (-0.02, 0.02)	-0.002 (-0.03, 0.02)	0.001 (-0.02, 0.02)	0.002 (-0.02, 0.02)
1 day	**0.03 (0.01, 0.06)**	0.01 (-0.02, 0.04)	**0.03 (0.01, 0.06)**	**0.03 (0.01, 0.06)**
Never	**0.04 (0.005, 0.07)**	0.01 (-0.03, 0.06)	**0.04 (0.003, 0.07)**	**0.04 (0.003, 0.07)**
Contacted family/friend by electronic msg (ref: Every day)				
4-6 days	<0.001 (-0.02, 0.02)	0.001 (-0.02, 0.02)	0.002 (-0.02, 0.02)	0.001 (-0.02, 0.02)
2-3 days	-0.001 (-0.02, 0.02)	<0.001 (-0.02, 0.02)	-0.01 (-0.03, 0.02)	<0.001 (-0.02, 0.02)
1 day	0.01 (-0.02, 0.04)	0.01 (-0.02, 0.04)	-0.01 (-0.04, 0.03)	0.01 (-0.02, 0.04)
Never	-0.01 (-0.05, 0.02)	-0.01 (-0.04, 0.02)	0.01 (-0.04, 0.05)	-0.01 (-0.04, 0.02)
Took part in an online community activity (ref: Every day)				
4-6 days	-0.003 (-0.03, 0.03)	-0.004 (-0.03, 0.03)	-0.004 (-0.03, 0.03)	0.02 (-0.02, 0.06)
2-3 days	-0.01 (-0.04, 0.01)	-0.02 (-0.04, 0.01)	-0.02 (-0.04, 0.01)	0.01 (-0.03, 0.04)
1 day	0.01 (-0.02, 0.03)	0.01 (-0.02, 0.03)	0.01 (-0.02, 0.03)	0.02 (-0.02, 0.05)
Never	0.01 (-0.01, 0.04)	0.01 (-0.01, 0.04)	0.01 (-0.01, 0.04)	0.03 (-0.01, 0.06)
Cohort (ref: age 63)*contact mode				
4-6 days*age 51	0.003 (-0.06, 0.07)	0.01 (-0.03, 0.05)	-0.01 (-0.05, 0.02)	0.02 (-0.05, 0.08)
4-6 days*age 31	0.004 (-0.11, 0.12)	0.03 (-0.03, 0.09)	0.05 (-0.01, 0.1)	-0.08 (-0.18, 0.02)
4-6 days*age 20	0.03 (-0.07, 0.13)	-0.003 (-0.08, 0.07)	-0.02 (-0.09, 0.05)	**-0.11 (-0.22, -0.005)**
2-3 days*age 51	-0.005 (-0.05, 0.04)	-0.01 (-0.05, 0.03)	0.01 (-0.03, 0.05)	0.004 (-0.05, 0.06)
2-3 days*age 31	0.02 (-0.07, 0.10)	0.02 (-0.04, 0.08)	0.002 (-0.06, 0.06)	-0.03 (-0.12, 0.05)
2-3 days*age 20	0.04 (-0.04, 0.13)	0.002 (-0.07, 0.07)	0.02 (-0.06, 0.1)	**-0.12 (-0.21, -0.03)**
1 day*age 51	0.01 (-0.04, 0.05)	0.01 (-0.04, 0.06)	0.02 (-0.03, 0.07)	0.002 (-0.05, 0.06)
1 day*age 31	0.05 (-0.04, 0.13)	0.04 (-0.04, 0.11)	0.08 (-0.02, 0.18)	-0.01 (-0.09, 0.07)
1 day*age 20	0.10 (0.02, 0.18)	0.07 (-0.03, 0.17)	0.05 (-0.13, 0.24)	-0.06 (-0.16, 0.04)
Never*age 51	-0.02 (-0.07, 0.03)	-0.02 (-0.09, 0.06)	-0.05 (-0.12, 0.02)	-0.01 (-0.06, 0.04)
Never*age 31	0.05 (-0.03, 0.14)	0.07 (-0.03, 0.17)	-0.02 (-0.16, 0.13)	0.002 (-0.08, 0.08)
Never*age 20	0.04 (-0.04, 0.11)	0.08 (-0.02, 0.18)	0.01 (-0.14, 0.16)	-0.06 (-0.13, 0.02)
Rsq				
Within	0.0123	0.0121	0.0121	0.0124
Between	0.1729	0.1296	0.1045	0.0256
Overall	0.1232	0.092	0.075	0.022

Bold coefficients indicate statistical significance at $p<0.01$
All models control for variables described in S3.

S8 Fixed effect models estimates (95% confidence intervals) of anxiety-depression and Kessler score with interactions between online community activity*pre-pandemic social media use: table

	Anxiety-Depression	Kessler
Met people in-person (ref: 4-6 days)		
Every day	0.05 (-0.05, 0.15)	0.04 (-0.09, 0.16)
2-3 days	0.06 (-0.02, 0.14)	0.05 (-0.03, 0.13)
1 day	**0.11 (0.03, 0.19)**	0.05 (-0.03, 0.12)
Never	0.07 (-0.001, 0.15)	0.06 (-0.01, 0.14)
Talked to family/friend via video/phone (ref: Every day)		
4-6 days	-0.004 (-0.08, 0.07)	-0.004 (-0.08, 0.07)
2-3 days	-0.002 (-0.07, 0.06)	-0.02 (-0.08, 0.05)
1 day	0.08 (-0.0002, 0.17)	0.05 (-0.03, 0.14)
Never	0.08 (-0.02, 0.17)	-0.02 (-0.11, 0.07)
Contacted family/friend by electronic msg (ref: Every day)		
4-6 days	-0.02 (-0.09, 0.05)	0.02 (-0.04, 0.09)
2-3 days	-0.02 (-0.10, 0.05)	0.03 (-0.04, 0.10)
1 day	0.03 (-0.13, 0.19)	0.12 (-0.05, 0.29)
Never	-0.11 (-0.25, 0.03)	0.03 (-0.12, 0.18)
Took part in an online community activity (ref: Every day)		
4-6 days	-0.08 (-0.28, 0.12)	-0.13 (-0.32, 0.07)
2-3 days	**-0.20 (-0.36, -0.05)**	-0.20 (-0.36, -0.05)
1 day	-0.01 (-0.21, 0.20)	-0.06 (-0.27, 0.15)
Never	-0.12 (-0.25, 0.01)	-0.16 (-0.29, -0.03)
Online community activity*pre-pandemic social media use (ref:< 2 hrs/day)		
4-6 days*>2 hrs/day	0.01 (-0.22, 0.24)	0.04 (-0.19, 0.27)
2-3 days*>2 hrs/day	0.14 (-0.04, 0.33)	**0.16 (-0.03, 0.35)**
1 day*>2 hrs/day	-0.03 (-0.27, 0.20)	-0.04 (-0.28, 0.20)
Never*>2 hrs/day	0.15 (-0.002, 0.30)	**0.15 (-0.005, 0.31)**
Rsq		
Within	0.0324	0.0362
Between	0.0792	0.0946
Overall	0.0577	0.0768

Bold coefficients indicate statistical significance at $p<0.01$
All models control for variables described in S3.

S9.1 Testing assumptions of IV models + sensitivity analyses

Although IVs can control for confounding and measurement error in observational studies, they have some limitations. They are not useful when dealing with many confounders and also if the correlation between the IV and the exposure variables is small. Both weak instruments and confounders produce large standard errors which results in imprecise and biased results.

Table S4.1 shows the results of the first stage of the IV analysis. The coefficients show that the exogenous predictors of in-person social contact, namely the wave*region interaction terms are significantly associated with in-person social contact. Compared to people living in the North of England either at waves 1 or 3, people living in the South of England, Scotland and Wales had significantly higher frequency of in-person social contact at wave 2. This pattern corroborates the pattern shown in Figure S2.2 (based on mean in-person social contact frequency). The p-value for the F-test for excluded instruments was less than 0.001, indicating that there was a strong association between the exogenous variables (the country/region by wave interaction) and in-person social contact, fulling the relevance criteria of the IV analysis.

The test for over-identification (Hansen's J statistic =1.591, chi-sq p-val =0.45) indicates little evidence against the null-hypothesis, which is that the instruments are valid instruments i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. There was little evidence that the first stage equation was under-identified with the p-value of the Kleibergen-Paap rk LM statistic (2830.1) being less than 0.001. A low p value indicates we can reject the hypothesis that the 1st stage equation is under identified. Weak identification is a potential source of bias for IV analysis. However, the Cragg-Donald Wald F statistic (1498.6) was much larger than the largest Stock Yogo critical value, indicating that the null hypothesis of weak instruments can be rejected.

For sensitivity analyses, we conducted additional cross-sectional IV analyses specific for each wave using region as the instrument to predict in-person social contact at each wave. Results from the 2nd stage models are shown in Table S9.3. At wave 1, the effect of more frequent in-person social contact on anxiety-depression was positive, although not statistically different from zero. This positive association between more frequent in-person social contact and higher levels of anxiety-depression could have arisen because people who were allowed to mix outside their home (such as keyworkers) were initially worried about the consequences of in-person social contact on the risks of infection and subsequent ill-health. By waves 2 and 3, the coefficient of in-person social contact turned negative and was statistically significant at wave 3- more frequent in-person social contact resulted in lower anxiety-depression. The pattern of these cross-sectional IV estimates of in-person social contact are very similar to the fixed effects IV results shown in Table 3- lower frequency of in-person social contact driven by changes in pandemic restrictions was associated with an increase in anxiety-depression. Hence unobserved time-varying confounders were unlikely to bias the inference from the fixed effects IV analysis.

S9.2 1st stage of instrumental variable model predicting more frequent in-person social contact: table

	Estimates (95% CI)
Talked to family/friend via phone/video calls (ref: Every day)	
4-6 days	**0.06 (0.01, 0.11)**
2-3 days	0.01 (-0.04, 0.06)
1 day	0.03 (-0.03, 0.09)
Never	0.06 (-0.02, 0.14)
Contact family/friend by electronic msg (ref: Every day)	
4-6 days	0.01 (-0.03, 0.06)
2-3 days	-0.01 (-0.06, 0.03)
1 day	-0.04 (-0.10, 0.03)
Never	**-0.10 (-0.19, -0.01)**
Took part in an online community activity (ref: Every day)	
4-6 days	-0.03 (-0.12, 0.05)
2-3 days	**-0.10 (-0.17, -0.03)**
1 day	**-0.11 (-0.18, -0.04)**
Never	0.02 (-0.04, 0.09)
COVID infection (ref: no)	
Yes	**-0.10 (-0.15, -0.06)**
Tested positive COVID-19 infection (ref: no)	
Yes	-0.06 (-0.15, 0.03)
Self-rated health (ref: Excellent/Very Good)	
Good to Poor	0.004 (-0.04, 0.05)
Finances (ref: much worse off)	
A little worse off	-0.01 (-0.08, 0.05)
About the same	0.01 (-0.06, 0.08)
A little better off	-0.02 (-0.10, 0.05)
Much better off	-0.01 (-0.11, 0.10)
Household size (ref: more than 1 person)	
Single person household	**0.13 (0.07, 0.20)**
Keyworker (ref: not in work)	
In work: keyworker	**0.20 (0.14, 0.26)**
In work: not keyworker	**0.20 (0.15, 0.26)**
Wave-Region (ref: North/Midlands England or waves 1 or 3)	
Southern England wave 2	**0.98 (0.95, 1.01)**
Wales wave 2	**0.99 (0.89, 1.09)**
Scotland wave 2	**0.85 (0.77, 0.92)**

F test of excluded instruments:
F(3, 23348) = 1467.41
Prob > F = 0.0000
Hansen J statistic (overidentification test of all instruments): 1.591
Chi-sq(2) P-val = 0.4513

Bold coefficients indicate statistical significance at $p<0.01$

S9.3 2nd stage of instrumental Variables model estimates (95% confidence intervals) of anxiety-depression- cross-sectional models for each wave (with country/region as the exogenous variable): table

	May-20	Sep-Oct 20	Feb-Mar 21
Predicted more frequent meetings in-person	0.48 (-0.16, 1.12)	-0.07 (-0.21, 0.08)	**-0.22 (-0.43, -0.02)**
Talked to family/friend via phone/video calls (ref: Every day)			
4-6 days	0.02 (-0.04, 0.08)	0.001 (-0.04, 0.05)	-0.02 (-0.06, 0.03)
2-3 days	0.08 (-0.01, 0.18)	-0.01 (-0.07, 0.05)	-0.01 (-0.06, 0.05)
1 day	**0.19 (0.07, 0.31)**	0.05 (-0.02, 0.13)	**0.08 (0.002, 0.15)**
Never	**0.33 (0.12, 0.55)**	**0.13 (0.04, 0.22)**	**0.12 (0.01, 0.22)**
Contact family/friend by electronic msg (ref: Every day)			
4-6 days	**-0.13 (-0.18, -0.08)**	**-0.14 (-0.18, -0.10)**	**-0.15 (-0.19, -0.11)**
2-3 days	**-0.21 (-0.27, -0.15)**	**-0.19 (-0.24, -0.14)**	**-0.22 (-0.27, -0.17)**
1 day	**-0.23 (-0.31, -0.15)**	**-0.21 (-0.28, -0.14)**	**-0.26 (-0.32, -0.20)**
Never	**-0.28 (-0.43, -0.13)**	**-0.16 (-0.24, -0.07)**	**-0.24 (-0.32, -0.17)**
Took part in an online community activity (ref: Every day)			
4-6 days	0.05 (-0.04, 0.15)	-0.02 (-0.09, 0.06)	-0.03 (-0.12, 0.05)
2-3 days	0.05 (-0.03, 0.12)	-0.02 (-0.08, 0.04)	-0.02 (-0.09, 0.04)
1 day	0.05 (-0.02, 0.13)	-0.05 (-0.11, 0.01)	-0.04 (-0.11, 0.03)
Never	0.06 (-0.01, 0.14)	-0.05 (-0.1, 0.01)	-0.01 (-0.07, 0.05)
COVID infection (ref: no)			
yes	**0.07 (0.02, 0.12)**	**0.20 (0.16, 0.24)**	**0.11 (0.07, 0.15)**
Self-rated health (ref: Excellent/Very Good)			
Good-Poor	**0.30 (0.25, 0.35)**	**0.31 (0.28, 0.34)**	**0.33 (0.30, 0.37)**
Finances (ref: much worse off)			
a little worse off	**-0.15 (-0.23, -0.07)**	**-0.21 (-0.28, -0.14)**	**-0.22 (-0.29, -0.16)**
about the same	**-0.19 (-0.27, -0.11)**	**-0.34 (-0.40, -0.28)**	**-0.33 (-0.39, -0.28)**
a little better off	**-0.19 (-0.27, -0.11)**	**-0.35 (-0.41, -0.29)**	**-0.35 (-0.41, -0.29)**
much better off	-0.11 (-0.24, 0.02)	**-0.28 (-0.37, -0.19)**	**-0.34 (-0.40, -0.27)**
Household size (ref: more than 1 person)			
Single person household	**-0.15 (-0.27, -0.02)**	0.01 (-0.03, 0.05)	0.04 (-0.02, 0.11)
Intercept	**3.59 (0.86, 6.31)**	**1.56 (1.15, 1.96)**	**0.88 (0.11, 1.65)**
p of F test of excluded instruments	>0.10	<0.001	<0.001
N individuals	9,781	13,448	14,979

Bold coefficients indicate statistical significance at $p<0.01$