The University Of Sheffield.
Department Of Economics

Mental health and productivity: evidence for the UK

Dacheng Huo, Nigel Rice, Jennifer Roberts, Cristina Sechel

Sheffield Economic Research Paper Series

SERPS no. 2022023

ISSN 1749-8368

16 December 2022

Mental health and productivity: evidence for the UK

Dacheng Huo[a], Nigel Rice[a,b,*], Jennifer Roberts[c], Cristina Sechel[c]

[a] *Centre for Health Economics, University of York, UK*
[b] *Department of Economics and Related Studies, University of York, UK*
[c] *Department of Economics, University of Sheffield, UK*

December 16, 2022

Abstract

Understanding the drivers of productivity is fundamental to securing future well-being, but there are still large gaps in our knowledge concerning the relationship between productivity and the health of the labour force. We explore whether changes in mental health contribute to changes in labour market productivity. We exploit the COVID-19 modules of the UK Household Longitudinal Study, which include a direct (self-reported) measure of productivity change relative to pre-COVID levels, as well as a clinically validated measure of mental health. To overcome endogeneity problems we use an instrumental variable approach implemented in an ordered probit model using two-stage residual inclusion. Our results show a strong positive relationship between mental health and productivity. At an individual level a unit decrease in mental health leads to an expected loss in productivity of approximately 4 minutes per working day. In our sample the average decrease in mental health over the period we study is -1.675, which predicts a reduction in productivity of 2,531 minutes for each hour that the sample works. Scaled up to the entire population of workers in June 2020, then total productivity losses would have been substantial.

Keywords: mental health, productivity, work from home, COVID-19, UKHLS.
JEL: I12, J14, J24.

Acknowledgements

The UK Household Longitudinal Study (UKHLS) is an initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research. The data were accessed via the UK Data Service.

*Corresponding author
Email addresses: `dacheng.huo@york.ac.uk` (Dacheng Huo), `nigel.rice@york.ac.uk` (Nigel Rice), `j.r.roberts@sheffield.ac.uk` (Jennifer Roberts), `c.sechel@sheffield.ac.uk` (Cristina Sechel)

1. Introduction

"Productivity isn't everything, but, in the long run, it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker". Paul Krugman in The Age of Diminished Expectations (Krugman 1994).

Productivity growth is recognised as the most important contributory factor to sustainable gains in living standards. Understanding the drivers of productivity, and in particular, reasons underlying recent global trends towards a slowdown in productivity growth is fundamental to securing future well-being. The relationship between well-being and productivity is however complex, especially when considering the potential impact of health (see Sharpe & Mobasher Fard (2022) for an overview). Health is a key component of human capital and an important factor of production (Layard 2013). We know, for example, that healthier workers are, on average, more productive (Burton et al. 2005) and this is particularly true in the case of mental health (MH) (Lerner & Henke 2008). MH problems, especially anxiety and depression, have a greater impact on ability to work than any other group of disorders.[1] As well as a large gap in employment rates between those with and without these problems, workers with MH problems also earn lower wages than those without (Contoyannis & Rice 2001). Increasing MH problems are therefore likely to exacerbate economic inequality, and may also be partly responsible for the persistently low productivity levels that characterise the UK economy. The declining MH of the population may also contribute to the post-recession 'productivity puzzle' in the UK (Pessoa & Van Reenen 2014), characterised by falling real wages alongside static (or declining) output per worker and rising employment. Changes in the composition of the workforce have been suggested as a possible explanation for this puzzle, but this debate has neglected health, focusing instead on education, skills and job type (Emmerson et al. 2013). However, the prevalence of chronic MH problems is increasing and welfare-to-work policies have increased the incentives for workers with MH problems to participate in work.[2]

[1] MH problems account for over 40% of Employment Support Allowance claims in the UK, compared to 16% for musculoskeletal problems (McInnes 2012)

[2] The employment rate of people with MH problems has been increasing steadily in recent years;

Empirical validation of the relationship between MH and productivity, particularly at an individual level, is hampered by limited availability of direct measures of productivity in large secondary data sets. Given the fundamental issues we outline above, this is an important evidence gap. In a rare empirical study in this area Oswald et al. (2015) use an experimental design to explore the relationship between happiness and productivity, and find that happier individuals have approximately 12% greater productivity in a piece rate setting.[3] In this study we employ a different approach, and attempt to address the evidence gap by exploring whether changes in MH contribute to changes in productivity in the UK. We exploit the COVID-19 modules of the UK Household Longitudinal Study which include a direct (self-reported) measure of productivity change relative to pre-COVID levels. This type of measure is rarely available and, as far as we are aware, has never been used to study the relationship between MH and productivity. As such, our analysis contributes a unique perspective that complements existing work on this topic, which relies mainly on proxy measures of productivity.

We measure MH via the General Health Questionnaire (GHQ) both before and during the pandemic. This allows us to identify a continuum of MH states that are not limited to having (or not having) diagnosed MH conditions. To overcome endogeneity problems we use an instrument based on changes in 'feelings of loneliness'. We argue that loneliness was more prevalent during the pandemic and operates on productivity via its impact on MH. Importantly our key variables are measured as changes from baseline (pre-COVID) which aids identification by ameliorating concerns about individual-specific and time-invariant confounders, as well as mis-reporting. This instrumental variable approach is implemented in ordered probit models using two-stage residual inclusion. We carry out a number of tests, which support our claims for the validity of the instrument. We also adapt the 'plausibly exogenous' approach of Conley et al. (2012) to subject 2SRI to the possibility that the the exclusion restriction is mildly violated. Our main result is robust to this approach.

Our results show a strong positive relationship between MH and productivity.

from around 37% at the start of 2018 to around 45% in early 2020 (Roberts et al. 2021).

[3]Happiness is not generally considered to be equivalent to mental health; the former is a measure of an emotional affective state, while the latter is more evaluative. Nevertheless Oswald et al. (2015) is one of the few studies that directly explores the relationship between either of these concepts and productivity.

Although women experienced larger reductions (worsening) in MH as well as larger falls in productivity compared to men, we find no statistically significant gender differences in the relationship between MH and productivity. These results are robust to various specifications and across COVID-19 module waves. Our results also suggest that while individual productivity changes as a result of MH deterioration are relatively small, these aggregate to substantial productivity losses for the economy as a whole. This link between MH and productivity strengthens the case for public policy to invest in MH prevention and treatment programmes. Not only will such measures improve population well-being directly through better health, but they also have the potential to enhance productivity, and hence, indirectly lead to further well-being gains through increased living standards.

2. Data

We use data from the Understanding Society COVID-19 Study, which consists of a series of 9 surveys conducted between April 2020 and October 2021 (ISER 2021). The eligible sample consists of all individuals aged 16 years and above in April 2020 from households who participated in waves 8 and 9 of the main UK Household Longitudinal Study (UKHLS)[4]. The surveys were administered online with some telephone interviews for households without internet access. We focus on modules 3 and 5, which were collected at the end of June and September 2020 respectively.[5] These modules cover the first and most intense phase of the pandemic where policies on social distancing, lockdown, and working from home were most prevalent. In mid-March 2020 the UK Prime Minister advised "now is the time for everyone to stop non-essential contact and travel." This was followed on 23 March with the announcement of the first lockdown in the UK, ordering people to stay at home (becoming legally enforceable on 26 March) and including the closure of pubs, restaurants, gyms

[4]Excluding individuals who refused to take part in the main UKHLS questionnaire or where mentally or physically unable to make informed decisions about participating in the survey. Individuals with unknown or foreign postal addresses were also excluded from the COVID-19 study.

[5]Productivity data was only collected in modules 3, 5, 7 and 9. Module 3 interviews took place between 25 June and 1 July, and module 5 interviews between 24 September and 1 October. Later surveys took place in January and September 2021. It is possible that other factors (beyond the initial shock of the pandemic) came into play during the later months, which we are unable to control for, but that could affect both MH and productivity. We focus on the first two modules in order to minimise bias arising from confounding factors and recall error.

and other social venues.[6] Following a peak of daily confirmed cases in April 2020, lockdown restrictions began to be eased with announcements on 10 May (plan for easing restrictions), 1 June (phased re-opening of schools in England), 15 June (non-essential shops reopened), and 23 June (relaxation of restrictions and introduction of a 2m social distance rule). Further easing of restrictions took place at the beginning of July when hospitality venues were reopened, followed by the opening of indoor theatres, bowling alleys and soft play venues in mid-August. The reintroduction of restrictions was ordered on 14 September (socialising in groups of up to 6), and a return to working from home on 22 September 2020. In total 31,964 people were invited to take part in the June COVID-19 module and 19,372 invited for the September module. Response rates were 44.2% in June and 66.5% in September. Importantly, the COVID-19 modules can be linked to data from past (and future) UKHLS waves, which allows us to link baseline information about respondents prior to the start of the pandemic. Cross-sectional and longitudinal survey weights are provided to increase sample representativeness to the national population at baseline.

Our outcome measure of productivity is based on responses to the question "Please think about how much work you get done *per hour* these days. How does that compare to how much you would have got done *per hour* in January/February 2020?". The response categories available are: 1. "I get much more done", 2. "I get a little more done", 3. "I get about the same done", 4. "I get a little less done", 5. "I get much less done" (see Table 1). As a follow-up question in the September wave respondents are asked to quantify how long it previously took to get done what currently (at the time of the module questionnaire) takes an hour. The responses available and the frequency of answers are provided in Table A1. For each set of respondents, i.e. those who report getting much or a little more done, and those who report getting much or a little less done, their is considerable overlap in the categories used to quantify how long it would have previously taken to complete one hours work. For example, of the 44% of respondents claiming they get much more done in a hour now than previously, 31% say that one hours work now would have taken up to 75 minutes previously, and 44% say it would have taken between 75 and 90 minutes. However, of respondents who say they get a little more done now, 56% reported it would have taken up to 75

[6]Exceptions were made for essential workers, people who were unable to work remotely, shopping for essential goods, accessing medical care, and undertaking exercise outdoors.

minutes previously and 34% between 75 and 90 minutes. This illustrates the likely existence of non-negligible measurement error in the original 5-point scale used to categorise productivity changes.[7] To reduce concerns over measurement errors we construct the following three point categorical variable: 1. "much less or a little less done", 2. "about the same", 3. "much more or a little more done". In the June module, the productivity question is only asked of respondents who work from home at least some of the time. This restriction was dropped in the September module, so all working respondents answered the question. Importantly, this question asks about the amount of work achieved per hour and represents a change in productivity benchmarked against the two months immediately preceding the first lockdown in March 2020.

MH is measured using the 12-item General Health Questionnaire (GHQ). GHQ is a widely used screening instrument for common mental disorders and a general measure of psychiatric well-being in the population (Goldberg et al. 1998). It has been validated in a number of international studies (Sartorius & Ustün 1995, Goldberg et al. 1997, Schmitz et al. 1999), and has been shown to be predictive of face-to-face clinical diagnosis of MH problems (Anjara et al. 2020). It has gained much attention as a measure of psychological health in studies of the relationship between MH and labour market outcomes (for example, see Garcia-Gomez et al. 2010, Mavridis 2015, Lagomarsino & Spiganti 2020, Bryan et al. 2022). The GHQ consists of 12 items intended to assess the severity of mental problems in the last few weeks; as such, this is a measure of current MH. Each item is scored using a 4-point Likert scale. These are then aggregated to generate a total score ranging from 0 to 36. For ease of interpretation, we reverse code this score such that higher values indicate better MH. GHQ scores are collected in the COVID-19 modules as well as in the main UKHLS waves. This allows us to measure the changes in MH relative to a pre-COVID baseline. Accordingly, our measure of MH, *GHQdiff*, is the difference between the GHQ score in June (or September) and the baseline GHQ score for the same individual in the final UKHLS interview taken before 2020[8]. Positive values denote improvement in

[7] It is possible that measurement error exists only in the quantification of amount of time gained or lost now compared to January/February However, it is unlikely that this is the sole source of mismeasurement.

[8] The vast majority of baseline interviews (approximately 95%) occurred in 2019, the rest took place in early 2020.

MH, while negative values signal deterioration.

The UKHLS COVID questionnaires also contain a module on loneliness. Of particular relevance for this study is the question "In the last 4 weeks, how often did you feel lonely?" Response categories are "Hardly ever or never", "Some of the time", and "Often". Loneliness is synonymous with perceived rather than objective social isolation (Hawkley & Cacioppo 2010), and this question is the same as that included in the Office for National Statistics Opinions and Lifestyles Survey[9] and is part of the Government Statistical Service harmonised principle of loneliness.[10] A similar question was asked in the main UKHLS survey enabling us to observe the change in loneliness of respondents from just before the pandemic to the June, and September, Covid-19 waves. Accordingly, we construct dummy variables at each of the two Covid waves that represent reporting "More lonely" and "Less lonely", contrasted against "No change in reported loneliness" compared to the baseline pre-Covid-19 interview. As set out in Section 3 we use this information as an identifying instrument, excluded from the productivity outcome equation. It is worth noting here that the 12 questions included in the GHQ measure of MH do not include any questions on loneliness or social interaction so we can be confident that empirically our variables are measuring different constructs. We provide further justification for our choice of instrument in Section 3.

We include a number of personal and household characteristics as control variables, based on a standard Mincerian wage equation, where wages are assumed to reflect productivity (Mincer 1958). Detailed definitions of all variables are given in Table 1. For the analysis of the June 2020 data we also include a set of industry indicators, which capture the main sector in which the individual currently works.[11] We find that controlling for industry does not substantially change the main results.

Our analysis consists of 18-64 year olds who are employed or self-employed, excluding workers who were furloughed at the time of the interview, or before. After dropping individuals for whom data was missing on the set of key explanatory variables, the June sample consists of $N = 2,902$ individuals, of which $1,201$ were male

[9] www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/coronavirusandlonelinessgreatbritain/3aprilto3may2020.

[10] A number of questions that are intended for use across surveys so that data from different sources are comparable.

[11] Industry information is not available in the September 2020 module due to a routing error.

Table 1: Variable definitions

Variable	Definition
Dependent variable	
Productivity change	"Please think about how much work you get done per hour these days. How does that compare to how much you would have got done per hour back in January/February 2020?" Possible answers: "much less", "a little less", "same", "a little more", "much more". We group into three categories: 1 - much or a little less, 2 - same, 3 - much or a little more.
Explanatory variables	
GHQdiff	Change in GHQ score between relevant COVID module and pre-COVID baseline. Original GHQ score is measured on a 36-point scale (ranging from 0 to 35) where high values denote worse health. We reverse code so that higher values denote better health.
Health condition	Dummy variable = 1 if individual has a long term health condition at the time of interview.
Female	Dummy variable = 1 if individual is female.
Age	Age of respondent in years at the time of interview.
BAME	Dummy variable = 1 if ethnicity of individual is non-white.
Degree	Dummy variable = 1 if highest qualification attained is degree or equivalent.
Employment status	Three categories: employed (omitted), self-employed, both employed and self-employed.
Couple	Dummy variable = 1 if living as a couple in one household.
Kids 0-4	Dummy variable = 1 if there are kids aged 0-4 living in the household.
Kids 5-15	Dummy variable = 1 if there are kids aged 5-15 living in the household.
Kids 16-18	Dummy variable = 1 if there are kids aged 16-18 living in the household.
Industry	Industry dummies UK Standard Industrial Classification 2007 (omitted category Education).
Instrumental variables	
change in loneliness	Based on answers to the question: "In the last 4 weeks, how often did you feel lonely?" Possible answers: hardly ever, sometimes, often. Change relative to last pre-COVID interview. Three categories: no change (baseline), less lonely, more lonely.

and 1,701 were female[12]. Table 2 reports summary statistics for the main variables in our June sample for men and women separately, excluding the industry categories which are separately reported in Table 3. All statistics are weighted by the cross-sectional weights provided with the data to increase sample representativeness to the national population at baseline.

Approximately 43% of the sample reported no change in productivity relative to Jan/Feb 2020; 25.5% reported an increase (with 13.9% saying they got a little more done and 11.6% much more, not shown in table); 31.8% reported a decrease (with 21.5% getting a little less done and 10.3% getting much less done). Female respondents were more polarised in reporting changes in productivity compared to males, particularly in terms of getting less done (28.6% of men compared to 34.6% of women).

On average there was a decrease in MH during the pandemic; a result that has

[12] This constitutes approximately two thirds of respondents who were employed or self-employed in June 2020, not furloughed, and were asked about their change in productivity (individuals who reported never working from home in that month were not asked this question).

Table 2: Summary statistics: June 2020

	Full sample		Males		Females	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Productivity change						
(Much or a little) less done	0.318	0.013	0.286	0.018	0.346	0.017
Same	0.426	0.013	0.463	0.019	0.395	0.016
(Much or a little) more done	0.255	0.011	0.251	0.016	0.259	0.014
Health						
GHQdiff	-1.676	0.175	-1.477	0.234	-1.845	0.256
Health condition	0.409	0.012	0.408	0.019	0.410	0.017
Socio-demographic						
Female	0.539	0.013				
Age	44.81	0.346	46.08	0.511	43.72	0.466
BAME	0.077	0.007	0.091	0.011	0.064	0.007
Degree	0.698	0.013	0.723	0.017	0.676	0.018
Employment						
Employed	0.808	0.010	0.777	0.015	0.834	0.012
Self-employed	0.149	0.009	0.179	0.014	0.122	0.010
Both self & employed	0.044	0.005	0.044	0.008	0.043	0.006
Household						
Couple	0.731	0.013	0.779	0.019	0.690	0.017
Kids04	0.116	0.008	0.125	0.013	0.108	0.010
Kids515	0.300	0.011	0.305	0.017	0.295	0.015
Kids1618	0.116	0.007	0.118	0.011	0.114	0.010
Loneliness						
No change	0.687	0.012	0.751	0.017	0.632	0.017
Less lonely	0.147	0.009	0.123	0.012	0.167	0.012
More lonely	0.167	0.011	0.126	0.013	0.201	0.017
N	2,902		1,201		1,701	

Sample summary statistics weighted using cross-sectional weights provided with the data.

been documented elsewhere, for example Banks & Xu (2020) and Daly et al. (2020). The average decrease across the full sample was -1.675. However, the decrease was larger for females (-1.845) than for males (-1.477). This finding is in line with those of Orefice & Quintana-Domeque (2021) who show that gender differences in MH effects were associated with increased childcare and housework responsibilities for women, as well as the difference in COVID-19 related health concerns between men and women. Similarly Cheng et al. (2021) show that COVID-19 disrupted work-life balance in the household through increases in childcare, homeschooling and financial insecurity, and that the burden of these fell disproportionately on women, resulting in larger deterioration in their MH. A similar proportion, around 41%, of men and women reported having a health condition.

The majority, 81%, of the sample were employed. The proportion was higher for women (83%) than men (78%), and conversely a greater proportion of men than women were self-employed (18% versus 12%). Comparing loneliness in June 2020 with

Table 3: Industry summary statistics (June 2020)

	Full sample		Males		Females	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Industry						
Agriculture/forestry/fishing/mining/quarrying	0.010	0.002	0.011	0.004	0.009	0.002
Manufacturing	0.032	0.003	0.051	0.007	0.016	0.003
Utilities	0.025	0.004	0.030	0.006	0.021	0.005
Construction	0.039	0.008	0.055	0.008	0.025	0.012
Wholesale/retail	0.041	0.006	0.040	0.008	0.041	0.009
Repair of motor vehicles/transportation/storage	0.020	0.003	0.022	0.005	0.018	0.004
Accommodation/Food/Other services/households as employers	0.115	0.008	0.123	0.013	0.108	0.010
Information & communication	0.081	0.008	0.121	0.014	0.047	0.007
Financial & insurance	0.091	0.007	0.101	0.010	0.082	0.009
Real estate	0.014	0.002	0.019	0.004	0.010	0.002
Professional/scientific/technical	0.080	0.006	0.101	0.011	0.061	0.007
Admin/support services	0.048	0.005	0.021	0.005	0.070	0.009
Public administration & defense	0.063	0.007	0.082	0.015	0.046	0.005
Education	0.203	0.010	0.124	0.012	0.270	0.014
Human health/social work	0.097	0.007	0.060	0.009	0.129	0.009
Arts/entertainment/recreation	0.044	0.006	0.040	0.008	0.047	0.008
N	2,902		1,201		1,701	

Sample summary statistics weighted using cross-sectional weights provided with the data.

loneliness at the last pre-COVID interview, 14.7% of of respondents reported feeling less lonely and 16.7% reported feeling more lonely, with the rest reporting no change. The prevalence of loneliness in our data (not shown in table) [13] is very similar to that recorded in the Opinions and Lifestyle Survey, where 5% of people in Great Britain reported that they felt lonely 'often' between 3 April and 3 May 2020, and working age adults living alone were the most likely to report these feelings (Rees & Large 2020), although responses varied substantially across gender, with more women than men reporting feeling lonely "Some of the time", and twice the proportion of women reporting feeling lonely "Often". A substantial gender difference is also apparent when looking at changes in loneliness relative to the baseline interview. On one hand, women were more likely to feel less lonely (16.7%) compared to men (12.3%), but they were also much more likely to feel more lonely (20.1% compared to 12.6%).[14]

[13] In June 2020, 60.9% of our sample responded feeling lonely "hardly ever or never", 32.5% reported feeling lonely "Some of the time", and 6.7% reported feeling lonely "Often"

[14] See Lepinteur et al. (2022) for similar evidence on gender differences in loneliness during the COVID-19 pandemic, changes from pre-pandemic levels, and the relationship with life satisfaction.

3. Empirical approach

Our measure of productivity consists of responses on a categorical (*Likert*) scale, which we model using an ordered probit (OP) (Greene & Hensher 2010). Underlying this model is a latent variable, y_i^*, which is assumed to be a linear (in unknown parameters, λ and $\boldsymbol{\beta}_x$) function of the observed endogenous MH variable (MH_i), a vector of exogenous characteristics $\boldsymbol{x}_i$ (with no constant term), and a standard normal disturbance term, ε_i, such that

$$y_i^* = \lambda MH_i + \boldsymbol{x}_i'\boldsymbol{\beta}_x + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{1}$$

where λ is the parameter of interest. y_i^* is mapped onto observed $j = 0, \ldots, J-1$ outcomes as follows

$$y = j \text{ if } \mu_{j-1} \leq y_i^* < \mu_j \qquad \text{for } j = 0, \ldots, J-1,$$

where $\mu_{-1} = -\infty$ and $\mu_{J-1} = +\infty$ (to ensure well-defined probabilities, we also assume $\mu_{j-1} < \mu_j, \forall j$.) The expressions for the resulting probabilities and likelihood functions are well-known (for example, see Greene & Hensher (2010)). We estimate Equation (1) separately for the June and September modules, and for men and women. Although we use cross-sectional data, our model controls for potential time-invariant measurement error and individual unobserved effects in both productivity and MH as both are measured as changes from the baseline (pre-COVID) wave.

We augment Equation (1) with the following linear model for MH_i,

$$MH_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\gamma}_x + \gamma_z z_i + \epsilon_i, \qquad i = 1, \ldots, N, \tag{2}$$

where z represents an 'instrument' predictive of MH, but only associated with productivity, y_i^*, through its relationship with MH_i. Hence z is excluded from the outcome Equation (1).

Our estimation procedure is as follows. We assume MH is endogenous and that an instrument, z_i, exists. We apply two-stage residual inclusion, 2SRI, (see Terza et al. 2008) as an alternative to standard instrumental variables for non-linear models. 2SRI consists of firstly estimating Equation (2), by OLS and calculating residuals, $\hat{\epsilon}_i$. In the second-stage, the OP model, Equation (1), is estimated via maximum likelihood with the addition of the estimated residual from the first-stage included as a regressor.

Cross-sectional weights are applied to all estimations. Note that the endogenous regressor, MH_i, is maintained in the second-stage regression. Inclusion of $\hat{\epsilon}_i$ acts as a substitute for omitted variables in the second-stage. This approach provides a consistent estimator of the parameter of interest, λ. Standard errors are estimated via bootstrap with 1000 iterations.[15] Estimated probabilities of membership for each of the categories of productivity can then be obtained.

2SRI is a special case of the more general control function (CF) approach to dealing with endogeneity. Newey et al. (1999) shows that there exists a function of the residuals obtained from a first-stage that can be used as a correct adjustment for the endogeneity that exists in the second-stage equation. We consider more flexible non-linear forms for the function of residuals by separately including quadratic and cubic terms in the second-stage (for a fuller discussion, see Garrido et al. 2012). We test for the significance of these terms to determine which functional form of the residuals is preferred. We further estimate Equations (1) and (2) jointly by placing additional parametric structure on Equation (2) by assuming ϵ_i is normally distributed as $N(\nu, \sigma_\nu)$.

Our approach to estimation requires a suitable instrument excluded from the outcome equation and included in the first-stage regression of MH. As outlined above we construct a variable representing the change in loneliness between pre-COVID-19 and either the June or September wave by constructing dummy variables representing “More lonely” and “Less lonely”, contrasted against “No change in reported loneliness”. As Equation (2) is linear, we are implicitly controlling for individual time-invariant effects since both MH and our instrument are constructed as changes from baseline. The June questionnaire was fielded during a time of uncertainty and following a period of lockdown, social distancing and restrictions on mixing with family and friends. This was likely to contribute to what Tiwari (2013) terms ‘situational loneliness’, which arises through environmental factors and/or what has been termed ‘social loneliness’ characterised by the absence of a social network (Weiss 1973). Accordingly, our measure of the change in loneliness is likely to be driven by feelings

[15] Analytical standard errors can be computed for 2SRI (see Terza 2017). However, we use the bootstrapped standard errors as these are easier to obtain when considering non-linear functions of residuals to mimic a control function approach. They also allow consistency in estimated standard errors across our models.

associated with a sense of isolation and a general lack of meaningful social interaction and connectedness; feelings which were exacerbated by the restrictions that were in place (Pai & Vella 2021, Groarke et al. 2020). Loneliness is known to be associated with poorer MH (Hajek et al. 2020, Mushtaq et al. 2014), and while there is some evidence for a reciprocal relationship between loneliness and depressive symptoms (Cacioppo & Hughes 2006), longitudinal analysis by Cacioppo et al. (2010) indicates that loneliness is predictive of subsequent changes in depressive symptoms but not vice versa over a one year period.[16] Thus our conjecture is that across the period of our analysis[17] change in loneliness predicts changes in respondents' mental wellbeing, but has no direct impact on productivity changes, and as such, is a valid and relevant instrument appropriately excluded from our outcome equation.[18] We provide empirical support for the use of the instrument in Sections 4.3 and 4.4.

4. Results

4.1. Model specifications

Table 4 presents results from estimation using data from the June wave. The first row represents OP estimates of Equation (1), ignoring potential endogeneity of MH. This is then followed by 2SRI, and a more flexible CF approach utilising estimated residuals ($\hat{\epsilon}_i$) (2SRI), and power terms (flexible CF) from the first-stage regression of Equation (2) as additional regressors in the outcome equation. The final set of results

[16]Using a cross-lagged model to simultaneously estimate the relationship between loneliness and depressive symptoms, Cacioppo et al. (2010) find that loneliness in one year significantly predicts fluctuations in depressive symptoms in the subsequent year, but no significant effect of depressive symptomatology on loneliness from one year to the next.

[17]Approximately one year or less between the time of the last pre-COVID interview to the June/September interviews in 2020.

[18]We note that there is a related, albeit limited, literature that considers the impact of work-related loneliness on job performance (see Bai (2021)). However, this uses a specific measure of workplace loneliness arising from perceived social interactions and interpersonal relationships with work colleagues; the breakdown of which is hypothesised to lead to a sense of alienation, and a lower commitment to an organisation. These studies do not establish causal pathways between workplace loneliness and changes in job performance, which hypothetically could operate in part through impacts on MH. Our measure of loneliness was developed as a general measure relating to the previous four weeks, and given the restrictions in place at the time of the surveys, it is highly unlikely that respondents were thinking in specific terms about the quality of relationships at their workplace, but more general feelings of a lack of social support and interpersonal interactions. The question on loneliness in the COVID-19 module of UKHLS is asked of respondents prior to questions from the employment module.

assume joint normality of the errors, estimated via maximum likelihood. To conserve space results for the estimated coefficient on *GHQdiff* are reported together with the coefficients on the various residual terms obtained from the first-stage regressions, and the correlation between the error disturbances for joint estimation (ρ).

Table 4: Effect of mental health on productivity

	June 2020	
	Coef.	S.E.
Ordered probit		
GHQdiff	0.039 ***	(0.005)
2SRI		
GHQdiff	0.083 ***	(0.017)
$\hat{\epsilon}$	-0.048 ***	(0.018)
CF (i)		
GHQdiff	0.083 ***	(0.017)
$\hat{\epsilon}$	-0.048 ***	(0.018)
$\hat{\epsilon}^2$	0.000	(0.000)
CF (ii)		
GHQdiff	0.083 ***	(0.017)
$\hat{\epsilon}$	-0.038 **	(0.018)
$\hat{\epsilon}^2$	0.000	(0.000)
$\hat{\epsilon}^3$	0.000 ***	(0.000)
Joint estimation		
GHQdiff	0.080 ***	(0.014)
ρ	-0.268 ***	(0.090)
N	2,902	

* p<0.10, ** p<0.05, *** p<0.01. All regressions contain variables for sex, age, age-squared, educational attainment, ethnicity, employment, living as a couple, number of children in household, health conditions, and industry of employment. Regressions are weighted using cross-sectional weights provided in the June 2020 Covid module of UKHLS. Bootstrapped standard errors are reported using 1000 replications. $\rho = corr\,(\varepsilon_i, \epsilon_i)$.

Ignoring potential endogeneity of MH results in a coefficient estimate for *GHQdiff*, on a latent scale, of 0.039. This is positive, indicating that increases (decreases) in MH are associated with reporting of increases (decreases) in work productivity. The result is statistically significant at conventional levels. 2SRI estimation using loneliness as an instrument in the first-stage leads to a doubling of the coefficient on *GHQdiff* to 0.083, which remains statistically significant at the 1% level. The corresponding coefficient on the residual term included in the outcome equation is negative and significant. This provides supportive evidence for the endogeneity of MH in a productivity equation. When including additional power terms of the residual to add flexibility and non-linearity to the CF approach, the coefficient on MH does not vary substantively, and the higher powers of the residuals tend not to be significant. There appears little to be gained by conditioning on higher powers compared to the standard 2SRI approach.[19] Estimating Equations (1) and (2) jointly via maximum likelihood also results in a similar sized coefficient estimate on MH, which again is statistically significant. This approach, however, imposes additional parametric structure on the model. From the set of results reported in Table 4 our preferred estimator is 2SRI, and we implement this approach in all subsequent model specifications and sub-sample analyses.

4.2. Main estimation results

Table 5 reports estimation results for the June sample, and for males and females separately. The top panel presents the main estimation sample for individuals who reported working from home *at least some of the time*; the bottom panel focuses on the subsample of respondents who reported *always* working from home. For each sample, results from an OP and 2SRI are reported. Throughout we report only the coefficient estimate and standard error for *GHQdiff* and the residual for the outcome equation. Full results including coefficient estimates for the set of control variables for the full sample (top panel) are reported in the Appendix, Table A2.[20] The first-stage F-statistic for joint significance of the loneliness dummy variables is also reported together with the sample size. Bootstrapped standard errors using 1000 replications are reported throughout.[21]

Improvement in MH has a positive and statistically significant effect on productivity changes.[22] The effect is larger for males than females who reported working

[19] See Garrido et al. (2012), and O'Malley et al. (2011) for a more general discussion of these issues in situations where there is an endogenous binary treatment variable.

[20] Full estimation results on other models are available from the authors upon request.

[21] Occasionally, a bootstrap replicate failed to estimate, due to empty cell frequencies for certain categories of the set of conditioning variables. This was more common for analyses using small subsamples. Where this occurred, these are reported in the table footnotes.

[22] This result also holds when restricting the sample to employed respondents only (i.e. excluding individuals who are wholly or partially self-employed)

Table 5: Effect of MH on productivity in June 2020, by gender

	June 2020		
	Pooled Coef (S.E.)	Males Coef (S.E.)	Females Coef (S.E.)
At least sometimes worked from home (full sample)			
Ordered Probit			
GHQdiff	0.039 ***	0.042 ***	0.037 ***
	(0.005)	(0.008)	(0.006)
2SRI			
GHQdiff	0.083 ***	0.100 ***	0.071 ***
	(0.017)	(0.030)	(0.022)
$\hat{\epsilon}$	-0.048 ***	-0.062 **	-0.039 *
	(0.018)	(0.032)	(0.023)
1st-stage F-stat	61.74	27.44	38.80
N	2,902	1,201	1,701
Always worked from home			
Ordered Probit			
GHQdiff	0.045 ***	0.052 ***	0.040 ***
	(0.006)	(0.010)	(0.008)
2SRI			
GHQdiff	0.083 ***	0.073 **	0.086 ***
	(0.020)	(0.033)	(0.027)
$\hat{\epsilon}$	-0.042 *	-0.023	-0.050 *
	(0.021)	(0.035)	(0.028)
1st-stage F-stat	47.56	23.26	30.40
N	1,925	811	1,114

* p<0.10, ** p<0.05, *** p<0.01. All regressions contain variables for age, age-squared, educational attainment, ethnicity, employment, living as a couple, number of children in household, health conditions, and industry of employment. Regressions are weighted using cross-sectional weights provided with the data. Bootstrapped standard errors are reported using 1000 replications.

from home at least some of the time (top panel), particularly once the endogeneity of MH is controlled using 2SRI. While females are more likely to report a larger decrease in MH during the pandemic, and also more likely to report a decrease in productivity than men, there is some indication that the relationship between changes in MH and changes in productivity is potentially stronger for men than for women. The gender difference is reversed for individuals who reported working from home all of the time in June 2020 when we control for endogeneity.[23] However, as we note below, the

[23] For the full June sample, the average change in reported *GHQdiff* among individuals who worked from home exclusively was -1.615, while those that work from home occasionally had an average reported change in *GHQdiff* of -1.675.

marginal effects are not statistically significantly different between men and women. For all estimates, the F-statistic for the joint significance of the loneliness variables in the first-stage is large, and would appear to be of acceptable size. The regression coefficients on loneliness also display expected signs. For example, in the full sample (column 1 in the top panel of Table 5), the coefficients from the first-stage are 1.76 and -4.66 for “less lonely”and “more lonely”, indicating that feeling less lonely relative to the baseline interview is associated with an improvement in MH and feeling more lonely is associated with a worsening of MH.[24]

The coefficient estimates provided in Table 5 are difficult to interpret beyond the general direction of effect, and strictly are not comparable across sub-samples (e.g. gender) due to different scaling of the estimates.[25] Table 6 reports average marginal effects from the 2SRI model for the various June samples presented in Table 5. These provide the effect of a unit change in *GHQdiff* on the probability of reporting each of the three ordered outcomes representing the change in productivity. As expected, increases (decreases) in *GHQdiff* increase (decrease) the probability of reporting getting more done and decrease (increase) the probability of reporting getting less done.

In the pooled specification using the full sample (top panel, first column of Table 6), a unit increase (improvement) in *GHQdiff* leads to a 2.5 percentage point (ppt) increase in getting more done and a 2.7 ppt decrease in getting less done in June 2020 relative to January/February 2020. These effects are significant at the 1% level and are similar regardless of home working arrangements in the pooled specification. However, there are interesting gender differences. The marginal effects of changes in MH are diminished when looking at the subsample of men who worked exclusively from home in June 2020 compared to the full June sample, while the opposite is true for women. As previously indicated by the coefficient estimates, the marginal effects are larger for men compared to women in the full sample, but smaller when looking only at individuals who always worked from home. However, these gender differences

[24]Full estimation results for the 1st stage regression using the June 2020 sample are provided in Table A3.

[25]For an ordered probit model location and scale are not separately identified Greene & Hensher (2010).

are not statistically significant at standard levels.[26]

While these effects appear modest, they represent substantial proportions of the overall sample means of reporting lower (higher) productivity. Overall, 32% of respondents reported getting less done and 26% reported getting more done in June 2020 (Table 2). The corresponding 2SRI marginal effects presented in Table 6 (column 1, top panel) represent a change of just under 10% of these means.

As mentioned previously, the estimates presented in Table 6 represent the average effect on productivity due to a one unit change in *GHQdiff*. However, one unit on the GHQ scale is relatively small and does not typically reflect substantial changes in the underlying MH of the individual. The GHQ scale ranges from 0 to 36 and for the individuals in our June sample the average GHQ score pre-COVID is 25 with a standard deviation of 5. It is therefore relevant to consider the effects of larger changes in MH. For example, we can compare someone who experienced no change in MH with someone who experienced a drop of 5 points on the GHQ scale. For accuracy, we compare predicted probabilities rather than a multiplier of the average marginal effects. For someone with no change in MH, the predicted probability of reporting getting less done is on average 27.5% for our sample, while the predicted probability of getting more done is 29.8%. The respective probabilities for someone reporting a 5 point drop in MH are 41.6% and 18%. These estimates suggest very large differences in productivity for these two types of individuals.

The 2SRI estimates of the effect of mental health on productivity changes are larger than the corresponding OLS estimates. This may indicate that our estimate is a LATE identified off a specific group of individuals who experienced a different response to changes in mental health than the overall sample of individuals. To explore this possibility, we characterise the sample of 'compliers' by estimating the first stage coefficients on changes in loneliness across subsamples of individuals.[27] This will allow us to identify heterogeneity in the mental health response to loneliness. We do this by creating interaction terms between a given characteristic, e.g. being a male, and

[26]We test this by including a full set of gender interaction terms in our pooled 2SRI model, both in the first and second stage regressions, and testing the difference in marginal effects between men and women.

[27]This follows Nicoletti et al. (2022) and von Hinke et al. (2022) who use this approach in the absence of a binary endogenous treatment variable where the complier analysis suggested by Angrist & Pischke (2008) is not applicable.

Table 6: Marginal effect by gender

	June 2020		
	Pooled	Males	Females
	Coef	Coef	Coef
	(S.E.)	(S.E.)	(S.E.)
At least sometimes worked from home (full sample)			
Change in productivity			
Much or little less done	-0.027 ***	-0.031 ***	-0.024 ***
	(0.006)	(0.009)	(0.007)
Same	0.003 ***	0.002	0.003 **
	(0.001)	(0.002)	(0.001)
Much or little more done	0.025 ***	0.030 ***	0.021 ***
	(0.005)	(0.009)	(0.006)
N	2,902	1,201	1,701
Always worked from home			
Change in productivity			
Much or little less done	-0.026 ***	-0.022 **	-0.028 ***
	(0.006)	(0.009)	(0.008)
Same	0.001	0.000	0.002
	(0.001)	(0.002)	(0.001)
Much or little more done	0.026 ***	0.022 **	0.026 ***
	(0.006)	(0.010)	(0.008)
N	1,925	811	1,114

* p<0.10, ** p<0.05, *** p<0.01. Marginal effects calculated at the observed values of covariates for each individual, i, and averaged over all i. Bootstrapped standard errors are reported using 1000 replications.

the two dummy variables representing 'less lonely' and 'more lonely'. We test the significance of these interaction terms to assess whether heterogeneity exists in the first stage. We repeat this for each observed characteristic.

Table A4 reports the first stage coefficients on changes in loneliness. The only significant (at the 5% level) interaction term is for holding a degree.[28] This indicates that individuals without a degree who reported becoming 'more lonely' experienced a significantly greater negative change to their mental health than their more educated counterparts. Of those that reported feeling 'less lonely' the positive impact on mental health was more subdued than for their more educated counterparts. While a similar proportion of individuals with a degree and without a degree reported becoming 'less lonely' (14.82% and 14.30% respectively), a higher proportion of respondents without

[28] We note that the interaction terms with having children aged between 16 and 18 years old might also be considered significant. However, when the full set of interactions with the set of children dummies (i.e. for all ages) are included and modelled jointly, they do not appear as significant (F-stat (6,2866) = 1.61; p = 0.141).

a degree reported feeling 'more lonely' (18.50% versus 15.86% for respondents with a degree). The average change in mental health from pre-pandemic to June 2020 was more negative for individuals without (-1.98), than with (-1.54) a degree. The results suggest that less educated individuals are at a greater risk of mental health problems when experiencing feelings of loneliness. Combined with treatment effect heterogeneity, the IV estimates are likely higher than the OLS because they estimate a LATE with a relatively high weight for lower educated workers.

4.3. Instrument validity

We expect the instrument to have a strong relationship with MH since loneliness has been shown to be predictive of changes in mental health in the short-term but not vice-versa (Cacioppo et al. 2010), and given the question is framed in terms of one's recent feelings (i.e. in the last four weeks). We have no reason to believe a priori that changes in loneliness independently predict changes in productivity, other than through changes in MH.

A potential threat to the validity of the instrument is that the change in loneliness is not randomly distributed across individuals in the sample and might be confounded with unobservable characteristics. We rule out that loneliness and MH may be correlated in levels due to unobserved individual-specific time-invariant characteristics (such as personality traits) since the instrument, MH and outcomes are measured as changes from baseline.[29] While we cannot test directly whether other unobservables are correlated with changes in loneliness, we can offer reassurance that this is unlikely to be problematic in these data. Table A5 explores the extent of correlation between changes in loneliness and observed characteristics. The first column reports coefficients and standard errors from an ordered probit regression of the instrument on the set of covariates. Gender, having a health condition, and living as a couple are the only coefficients that are significant. As can be seen in columns three and four these affects are driven by the sample of males. The results suggest that changes in loneliness are not systematically distributed across females conditional on covariates. While there is some indication of imbalance for males when the sample is split by

[29] Personality traits are generally considered as largely fixed after a specific age, particularly over relatively short periods of time such as those considered in this analysis (Borghans et al. 2008, McCrae & Costa 2008, Cobb-Clark & Schurer 2012).

gender, our main specification fails to produce statistically significant differences in 2SRI effects of MH on productivity for males and females. Identification relies on compliers for whom MH changes are induced via changes in loneliness. The final column of Table A5 reports the results from a probit regression of reporting 'more lonely' or 'less lonely' on the set of covariates. This conditional regression shows that covariates are closely balanced across individuals reporting increased or decreased loneliness.

As an additional check on the influence of socio-economic controls on the instrument, Table A6 reports the coefficients on the instrument from the first-stage regression for the benchmark model that includes the set of controls and for a separate model excluding controls. The final column in the table reports t-statistics for tests of equivalence of coefficients across the two specifications. The results show that the first-stage coefficients on the change in loneliness across the two models are very similar and not statistically different. Taken as a whole, the above two descriptive exercises provides some reassurance that changes in loneliness are unlikely to be systematically distributed across individuals conditional on observed covariates in ways to invalidate the use of the instrument. We explore this conjecture further below.

In addition to descriptive tests of instrument validity we also investigate the conjecture that the instrument is properly excluded from the outcome equation by a number of informal tests of validity. First, assuming linearity, a Sargan test of over-identifying restrictions using two-stage least squares on the full sample of June respondents supports the exclusion restriction ($J- = 0.89, p = 0.34$). Secondly, as an additional sense check we run an analogous regression to what is often termed a zero first-stage test of instrument validity in the IV literature (see Lal et al. 2021 for a discussion). This consists of locating a sub-sample of the data for which the instrument is not expected to influence the endogenous regressor in the first-stage (see Bound & Jaeger 2000 in the context of treatment effects).[30] Where such a first-stage can

[30] For example, Altonji et al. (2005) consider the effect that Catholic high school attendance has on later life outcomes, and use whether an individual is a Catholic as an instrument.[31] For the instrument to be valid, it should be appropriately excluded from the outcome equation, and instead operate fully through attendance at a Catholic high school. A zero first-stage can be found among the sub-sample of individuals for whom practically no one attended a Catholic high school. The exclusion restriction can be tested through a reduced form regression of outcomes on controls and the instrument, but excluding high school type, in the zero first-stage sub-sample.

be located, the sub-sample provides a useful placebo test of the exclusion restriction. We apply an analogous approach to the two-stage residual inclusion estimation used in our analysis. This is obtained by estimating the reduced form of the outcome equation on the zero first-stage sub-sample (with $M \in N$) as follows

$$y_i^* = \delta z_i + \boldsymbol{x}_i' \boldsymbol{\beta}_x + \nu_i, \qquad i = 1, \ldots, M. \tag{3}$$

Failure to reject the null hypotheses, $H_0 : \delta = 0$, in Equation (3) provides support for the validity of the exclusion restriction. Underlying this approach is an assumption of homogeneity of the direct effect of δ in the zero-first stage and in the full sample.[32] Accordingly, the zero first-stage sub-sample should not be based on the instrument nor the outcome.

We define the zero first-stage in two ways. First as the sub-sample of individuals who report no change in MH across the period from prior to the pandemic to June 2020 ($n = 385$, 194 women and 191 men).[33] Secondly, we relax the zero first-stage definition to include respondents who reported a minimal change in MH (at most a one point change in the GHQ score, ($n = 929$). We include the second definition, as the increase in sample size over the first definition ensures greater precision in parameter estimates.[34]

For the first sub-sample, for whom no first-stage exists, we fail to reject the null of no effect of the instrument on productivity in the reduced form regression of Equation (3) ($F_2 = 1.03$; $p = 0.36$). For the second sub-sample, a first-stage exists. However, in the first-stage we fail to reject the null that the coefficient estimates on loneliness are jointly zero ($F_2 = 0.56$; $p = 0.57$) lending support to the zero-first stage approach. We also fail to reject the null of no effect of the instrument in the reduced form regression of Equation (3) ($F_2 = 0.40$; $p = 0.67$). Both tests provide supporting evidence that the exclusion restriction in Equation (1) holds.

[32] Although this may be seen as a weaker assumption that the usual IV assumption of a zero direct affect of the instrument on outcomes.

[33] The first-stage regression does not exist for this sub-sample of individuals because there is no variation in MH.

[34] We note that the lack of a first-stage in these examples are a construct of the data rather than a naturally occurring phenomenon.

4.4. Sensitivity to the exclusion restriction

We argue that loneliness directly affects MH. Literature supports these claims (Hajek et al. 2020, Mushtaq et al. 2014, Cacioppo et al. 2010). As our three main variables capture changes from a period just prior to the pandemic to during it, these reduce concerns over individual-specific unobservable components, for example personality traits, that might otherwise induce correlation between loneliness, MH and productivity. Empirical support for these claims is provided in the section above. However, it might be argued that time varying unobservable characteristics might lead to plausible violations of the exclusion restriction, or that the test statistics reported above are insufficiently powered to reject the null hypothesis. We explore the likely consequences of violations of the exclusion restriction in this section.

Conley et al. (2012) describe an approach to causal inference in linear models using instrumental variables where the instrument is described as "plausibly exogenous". Plausibility of the instrument relates to situations where the exclusion restriction is relaxed and can be thought of as approximating exogeneity. We extend this approach to our 2SRI estimator. We relax the exclusion restriction in Equation (1) as follows:

$$y_i^* = \lambda MH_i + \boldsymbol{x}_i'\boldsymbol{\beta}_x + \pi z_i + \varepsilon_i, \qquad i = 1, \ldots, N. \tag{4}$$

The exclusion restriction is equivalent to assuming $\pi = 0$. Plausible exogeneity incorporates information that π is near 0, but not exactly 0 (Conley et al. 2012). In the presence of endogenous MH_i, the parameters, λ and π are not jointly identified. Instead, prior information or assumptions may be imposed on π to obtain estimates of the key parameter of interest, λ. We follow a similar approach for our 2SRI estimator. This is possible in our application due to the first-stage relationship between MH, x, z and ϵ in Equation (2) being linear. Conley et al. (2012) propose a number of approaches to specifying prior information about the parameter, π. These range from specifying a set of plausible values for π across its support, to a full Bayesian approach with prior distributions specified for all parameters of the model and the error distribution. However, these approaches rely on forming opinions on a suitable range of values or distribution for π. A more recent addition to this literature is helpful in this regard. van Kippersluis & Rietveld (2018) suggest using information from a zero first-stage to obtain a suitable parameter for π. In principle, the idea rests on the assumption that the zero first-stage suggests a violation of the exclusion restriction. While this is not the case in our example, augmenting the plausible exogenous approach with a zero first-stage parameter is potentially useful as an additional sensitivity check on the robustness of our main findings.

We implement the above approach by first restricting the parameter π in Equation (4) to the estimate obtained in the zero first-stage. This provides an estimate of λ conditional on a potential violation of the exclusion restriction. We take the coefficient estimated from the zero first-stage sub-sample of individuals for whom *GHQdiff* changed at most one point. We take this sub-sample, rather than the sub-sample for which $GHQdiff = 0$, since the former has a much larger sample and we can place greater confidence in the precision of the parameters obtained. We fix π to the values from the sub-sample analysis and estimate Equation (4) together with Equation (2) using 2SRI. For inference, we then perturb π by allowing it to take values $+/-1$ standard deviations away from its estimated value.[35] As our instrument contains two dummy variables, this leads to a set of four permutations. As suggested by Conley et al. (2012) we take the union of confidence intervals across these estimates to obtain a measure of uncertainty in our estimate of λ. This results in $\hat{\lambda} = 0.076$ with a union of 95% confidence intervals ranging from 0.018 to 0.135. While uncertainty around the point estimate is large, the confidence interval excludes 0, and the point estimate remains close to that reported as our main result (0.083). This provides a reassuring sensitivity check of our main findings, and supports claims for the exclusion restriction in Equation (1).

4.5. Robustness checks

Table A7 presents results of robustness checks. It might be hypothesised that the relationship between MH and productivity is confounded by working patterns and contractual arrangements for different groups of workers. For example, the self-employed may have more flexibility to factor in leave or work less demanding hours when faced with a MH issue. Alternatively, they might not feel able to take sickness leave for fear of losing clients and consequently continue to work at a less productive rate. Individuals who are salaried, rather than paid by the hour, may experience

[35] Allowing the parameter to vary uniformly by this amount is extreme, and instead imposing a distribution on π would reflect a high probability of values near to $\hat{\lambda}$ and diminishing probability as values moves away.

stronger labour market attachment and greater job security manifesting in a different relationship between MH and productivity than hourly paid workers. We estimate our preferred model on sub-samples of employed workers and separately salaried workers. Results are presented in the first two columns of Table A7. 2SRI coefficient estimates on *GHQDiff* for salaried only workers (0.084) are indistinguishable from our main result (coefficient of 0.083). However, the coefficient on the employed subsample is lower at 0.070 and remains statistically significant at the 1% level, indicating a potentially stronger relationship between MH and productivity for the self-employed.

In a perfectly competitive market without frictions, worker productivity would be compensated through wage rates. Higher wage earners might also be more susceptible to productivity loses due to MH problems than lower wage earners due to being more productive prior to the onset of a period of ill health. The third column of Table A7 conditions estimation on pre-pandemic levels of wages. The estimated coefficient on wages in the outcome equation is 0.0003 (se 0.000091) implying that respondents with higher (lower) levels of wages were more (less) productive. However, the impact of a change in MH on productivity changes remains similar, albeit slightly smaller in magnitude, to the results reported for the main model (0.077 versus 0.083).

We attempt to leave out individuals who may have misinterpreted the productivity question. Participants who reported a rise or fall in productivity were asked the reason for this change in productivity in the September wave. Responses include items such as 'lack of motivation', 'childcare/home-schooling', 'interrupted less'. We exclude individuals who reported a fall in productivity due to 'having had less work to do' and individuals who reported a rise in productivity due to 'having had more work to do' or 'not needed to commute/travel to work' because these responses may indicate changes in the number of hours worked rather than productivity per hour. In the full September module with no work from home restrictions, these categories make up 11.1% of respondents who report a fall in productivity, and 22.2% (more work to do) and 17.7% (no commute/travel) of respondents reporting a rise in productivity. Excluding these individuals has little effect on the MH coefficient in the main 2SRI model using the sample of respondents from the September module. (0.053 compared to 0.054 for the full September sample shown in Table 7).

4.6. Heterogeneity analyses

Table A8 considers heterogeneity in the relationship between changes in MH and changes in productivity by considering different ages of individuals and by MH status at baseline. These are compared for the full sample of respondents who reported working from home at least sometimes in the June wave. A consistent finding across the pooled sample of men and women and separately by gender is that older workers (50 years and above) display a stronger relationship between MH and productivity compared to younger workers (less than 50 years). The magnitudes of the differentials are notable[36], although given the reduced sample sizes and accordingly precision upon which these estimates are based, they are unlikely to be statistically significantly different.

The final two columns of Table A8 present results separately for individuals with good and poor MH at baseline. These categories are defined by the caseness score for GHQ. Each of the 12 items for the GHQ is scored on a four point scale (0-3), such that the overall measure ranges from 0 to 36, with 36 (on the original scale) being most distressed. The caseness scale recodes values of 0 and 1 on each of the 12 items to 0 and values of 2 and 3 to 1. Hence, the caseness scale is from 0 (least distressed) to 12 (most distressed). We use a cut-off of 4 and above to represent poor initial MH, and less than 4 as good initial MH.[37] Using the caseness threshold, the estimated relationship between MH and productivity is slightly larger for individuals who report good MH at baseline than for those reporting poor MH, but differences are small.

4.7. Longer-run effects on productivity changes

The September wave of the COVID questionnaire also contained the productivity question. To be consistent with the June wave, the benchmark used to report changes in productivity was also January/February of 2020. This was asked of all respondents as long as they were employed, self-employed or both. There was no restriction on

[36] Although the scaling between models will differ, the coefficients for older individuals are between 41% and 46% higher than for the younger individuals.

[37] The National Health Service in England uses a cutoff of 4 as the threshold to monitor the percentage of people suffering from poor MH and an indicator of possible psychiatric disorder in the general population, see https://files.digital.nhs.uk/BA/46AF8E/Spec_03J_321VSP2_10_V1.pdf.

respondents working from home.[38] We further restrict the sample to individuals who had not been furloughed. The September wave omitted to collect information on industry sector. Accordingly, we include the set of conditioning variables used in the June analysis with the exception of the industry dummies. Table 7 reports these results. Column (1) presents June 2020 estimates analogous to the first column of Table 5 but omitting the industries variables. The coefficient estimate of *GHQdiff* is very similar for the two specifications.[39] Column (2) restricts the sample to only those individuals who reported working form home at least some of the time. This is to ensure comparability with the June wave (column (1)). The estimates in column (3) restricts the sample further by considering only individuals who worked from home always. Column (4) presents results without imposing any work from home restrictions.

The effect of MH on productivity is lower in September (column 2) compared to June (columns 1) for respondents who reported working from home at least some of the time. While the first-stage F-statistic is lower than for the June sample its value suggests the instrument is relevant. Restricting the September sample to only those respondents always working from home (column (3)) results in an increase in the coefficient estimate of *GHQdiff*. The increase is similar to the corresponding increase observed in the June wave (Table 5). Placing no restrictions on working from home (column (4)) results in a smaller coefficient on *GHQdiff* (0.054). Again, the first-stage F-statistic appears sufficiently large to have confidence in the relevance of the instrument. Accordingly, while the pandemic restrictions were more relaxed in September compared to June 2020, the relationship between changes in MH and changes in productivity persist and are of similar magnitude. Analysis of marginal effects by gender (not reported) reveals that the difference between men and women remains statistically non-significant in September 2020.

[38] By September Government advice on working-from-home had relaxed since the June wave and accordingly it would have been less relevant to only question respondents who were at least working form home some of the time.

[39] Omitting industry information increases the sample size in column (1) Table 7 by 27 compared to the respective model of Table 5.

Table 7: Effect of MH on productivity

	June 2020	September 2020		
	WFH At least Sometimes	WFH At least Sometimes	WFH Always	WFH No Restriction
	(1)	(2)	(3)	(4)
	Coef (S.E.)	Coef (S.E.)	Coef (S.E.)	Coef (S.E.)
Ordered Probit				
GHQdiff	0.039 ***	0.022 ***	0.029 ***	0.028 ***
	(0.005)	(0.007)	(0.010)	(0.006)
2SRI				
GHQdiff	0.083 ***	0.068 **	0.082 *	0.054 ***
	(0.017)	(0.033)	(0.047)	(0.019)
$\hat{\epsilon}$	-0.048 ***	-0.049	-0.056	-0.027
	(0.018)	(0.033)	(0.047)	(0.020)
1st-stage F-stat	61.74	20.00	11.50	39.65
N	2,902	1,938	1,016	3,600

* p<0.10, ** p<0.05, *** p<0.01. All regressions contain variables for sex, age, age-squared, educational attainment, ethnicity, employment, living as a couple, number of children in household, health conditions. Note industry of employment not included. Regressions are weighted using cross-sectional weights provided with the data. Bootstrapped standard errors are reported using 1000 replications. Exceptions: column (2) 998; column (3) 982 replications.

4.8. Quantification of productivity changes

A follow-up question to the change in productivity aimed at quantifying the gains or losses reported was asked of respondents. This was in the form of the question "Thinking about how much less (more) you get done these days, would you say that what you can do in an hour now would previously have taken you:".[40] The question was included in the September wave (but not the June wave) and relates to the sample of respondents in column (4) of Table 7. Table 8 summarises the responses on the 3-point scale we use.[41]. The sample consists of 3,601 respondents, of which 2,180 report no change in productivity. Of the remaining 1,421 respondents, 512 report a fall in productivity (367 "A little less" and 145 "Much less"), and 909 report a rise in productivity (477 "A little more" and 432 "Much more").

The first column of the table summarises the fall and rise in productivity in minutes per hour by taking the midpoint of the reported category. For example, if

[40] We have included "(more)" to differentiate the question asked of respondents reporting a productivity fall from that asked to respondents reporting a productivity gain.

[41] Note Table A1 provides the responses on the original 5-point scale

prior to the pandemic it took between 45 and 60 minutes to complete what could be completed at the September wave in one hour, then the estimated productivity loss is 8 minutes (approximate midpoint between a loss of 15 minutes and 0 minutes). We do likewise for all other categories including responses to a rise in productivity.[42] Taking the midpoint enables us to compute the expected loss or gain in productivity per hour. For example, for respondents reporting getting much or a little less done, the expected loss in minutes is $[(230 \times 8) + (183 \times 23) + (89 \times 45)]/502 = 20.03$ minutes per hour. The expected gain for those reporting much or a little more done is 19.84 minutes per hour.

A unit decrease (increase) in GHQ leads to an increase in the probability of reporting getting much or a little less (more) done of 0.027 (0.025) (Table 6). This equates to an expected change in productivity of approximately half a minute per hour ($20mins \times 0.026$), or approximately 4 minutes a day assuming 7.5 hours in a work day. This figure appears small, but considering that the average change in the GHQ score observed in the sample is -1.675, then across the full sample this equates to an aggregate expected loss of productivity of 2,531 minutes for every hour worked[43] This computes to 1,582 hours[44] (211 days) per week in lost productivity across the full sample. If this sample were representative of the population of workers in June 2020, then total productivity losses would have been substantial.

Table 8: September wave: Quantification of changes in productivity

Fall in productivity	Mins	Little or much less Freq.	%	*Rise in productivity*	Mins	Little or much more Freq.	%
Don't know		10	1.95	Don't know/refusal		12	1.32
Between 45 and 60 minutes	8	230	44.92	Up to 75 minutes	8	400	44.00
Between 30 and 45 minutes	23	183	35.74	Between 75 and 90 minutes	23	353	38.83
Less than 30 minutes	45	89	17.38	More than 90 minutes	45	144	15.84
N		512	100	N		909	100

Responses to the question "Thinking about how much less (more) you get done these days, would you say that what you can do in an hour now would previously have taken you:"

[42] The final category for a rise in productivity where the response is "More than an hour and a half", is truncated at the upper limit at an hour and three-quarters when estimating the midpoint.

[43] $-1.675 \times 0.026 \times 20.03 \times 2902 = 2531$ minutes.

[44] $\frac{2531 \times 37.5}{60} = 1582$.

5. Discussion and conclusion

The World Health Organisation estimates an annual cost of depression and anxiety to the global economy of around US$1 trillion in lost productivity.[45] MH problems can make it difficult to carry out day-to-day work activities, leading to decreased efficiency and higher levels of presenteeism. For example, depression has been found to limit physical job demands 20% of the time, and to impair cognitive performance at least 35% of the time (Lerner & Henke 2008). The prevalence of common mental disorders is high in the UK, with one in six people reporting these in 2014, and has been increasing since the early 1990s (Baker 2021). If this trend continues, we anticipate MH will be an increasing factor in explaining worker productivity. Given the well-documented low levels of productivity in the UK, understanding the link between MH and productivity is a particularly important aspect of managing the UK economy and for future labour market policy. Our analysis provides much needed evidence that quantifies the effect of MH on productivity.

Unlike most studies that rely on wages to proxy productivity, we exploit unique data, which includes a direct (self-reported) measure of productivity change. We find that a change in MH has a statistically significant (albeit modest) effect on a change in individual productivity. Deteriorating MH leads to a higher probability of getting less done at work relative to baseline. The opposite is true for improvements in MH. Despite women reporting a greater reduction in MH and productivity relative to the pre-Covid period, we find no evidence that the effect differs statistically by gender.

Using additional self-reported data on how much more or less each individual got done, we estimate that a deterioration in MH of 1.675 GHQ units could lead to an aggregate expected loss of 2,531 minutes for every hour worked across the sample of respondents in this survey. If these estimates are applicable to the UK population of workers, the implied productivity losses are substantial. We expect the effects to be much larger if one considers changes in MH that lead to diagnosed conditions. This link between MH and productivity strengthens the case for public policy to invest in MH prevention and treatment programmes. Not only will such measures improve population well-being directly through improved health, but will also enhance

[45] www.who.int/teams/mental-health-and-substance-use/promotion-prevention/mental-health-in-the-workplace

productivity, and hence, indirectly lead to further well-being gains through increased living standards.

References

Altonji, J., Elder, T. & Taber, C. (2005), 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools', *Journal of Political Economy* **113**(1), 151–184.

Angrist, J. & Pischke, J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Anjara, S., Bonetto, C., Van Bartel, T. & Brayne, C. (2020), 'Using the GHQ-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations', *International Journal of Mental Health Systems* **15**, 62. **URL:** *https://doi.org/10.1186/s13033-020-00397-0*

Bai, X. (2021), 'A review of recent studies on workplace loneliness', *Academic Journal of Business & Management* **3**(4), 78–83.

Baker, C. (2021), 'Mental health statistics (England)', *UK House of Commons Library* **CBP-06988**.

Banks, J. & Xu, X. (2020), 'The mental health effects of the first two months of lockdown during the COVID-19 pandemic in the UK', *Fiscal Studies* **41**(3), 685–708.

Borghans, L., Lee Duckworth, A., Heckman, J. & Weel, B. (2008), 'The economics and psychology of personality traits', *Journal of Human Resources* **43**(4), 972–1059.

Bound, J. & Jaeger, D. (2000), 'Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?', *Research in Labor Economics* **19**(4), 83–108.

Bryan, M., Rice, N., Roberts, J. & Sechel, C. (2022), 'Mental health and employment: A bounding approach using panel data', *Oxford Bulletin of Economics and Statistics* **84**(5), 1018–1051.

Burton, W. N., Chen, C. Y., Conti, D. J., Schultz, A. B., Pransky, G. & Edington, D. W. (2005), 'The association of health risks with on-the-job productivity', *Journal of Occupational and Environmental Medicine* **47**(8), 769–777.

Cacioppo, J. T., Hawkley, L. C. & Thisted, R. A. (2010), 'Perceived social isolation makes me sad: five year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study', *Psychology and Aging* **25**(2), 453–463.

Cacioppo, J. T. & Hughes, M. E. (2006), 'Loneliness as a specific risk factor for depressive symptoms: Cross-sectional and longitudinal analyses', *Psychology and Aging* **21**(1), 140–151.

Cheng, Z., Mendolia, S., Paloyo, A. R., Savage, D. A. & Tani, M. (2021), 'Workig parents, financial insecurity, and childcare: mental health in the time of COVID-19 in the UK', *Review of Economics of the Household* **19**, 123–144.

Cobb-Clark, D. & Schurer, S. (2012), 'The stability of big-five personality traits', *Economics Letters* **115**, 11–15.

Conley, T., Hansen, C. & Rossi, P. (2012), 'Plausibly exogenous', *The Review of Economics and Statistics* **94**(1), 260–272.

Contoyannis, P. & Rice, N. (2001), 'The impact of health on wages: Evidence from the british household panel survey', *Empirical Economics* **26**(4), 599–622.

Daly, M., Sutin, A. R. & Robinson, E. (2020), 'Longitudinal changes in mental health and the COVID-19 pandemic: evidence from the UK Household Longitudinal Study', *Psychological Medicine* pp. 1–10. https:doi:10.1017/S0033291720004432.

Emmerson, C., Johnson, P. & Miller, H. (2013), 'The IFS Green Budget'. https://ifs.org.uk/green-budget/2013.

Garcia-Gomez, P., Jones, A. M. & Rice, N. (2010), 'Health effects on labour market exits and entries', *Labour Economics* **17**, 62–76.

Garrido, M. M., Deb, P., Burgess, J. F. & Penrod, J. D. (2012), 'Choosing models for health care cost analysis: Issues of nonlinearity and endogeneity', *Health Services Research* **47**(6), 2377–2397.

Goldberg, D. P., Gater, R., Sartorius, N., Ust̀un, T. B., Piccinelli, M., Gureje, O. & Rutter, C. (1997), 'The validity of two versions of the GHQ in the WHO study of mental illness in general health care', *Psychological Medicine* **27**(1), 191–197.

Goldberg, D. P., Oldehinkel, T. & Ormel, J. (1998), 'Why GHQ threshold varies from one place to another', *Psychological Medicine* **28**(4), 915–921.

Greene, W. & Hensher, D. (2010), *Modeling ordered choices*, Cambridge University Press.

Groarke, J., Berry, E., Graham-Wisener, L., McKenna-Plumley, P., McGlinchey, W. & Armour, C. (2020), 'Loneliness in the UK during the COVID-19 pandemic: Cross-sectional results from the COVID-19 Psychological Wellbeing Study', *PLoS ONE* **19**(9), 1–18.

Guiso, L., Sapienza, P. & Zingales, L. (2016), 'Long-term persistence', *Journal of the European Economic Association* **14**(6), 1401–1436.

Hajek, A., Kretzler, B. & König, H. H. (2020), 'Multimorbidity, loneliness, and social isolation. a systematic review', *International journal of environmental research and public health* **17**(22), 8688.

Hawkley, L. C. & Cacioppo, J. T. (2010), 'Loneliness matters: A theoretical and empirical review of consequences and mechanisms', *Annals of Behavioral Medicine* **40**(2), 218–277.

ISER (2021), 'Understanding Society: COVID-19 Study, 2020-2021. [data collection]. 11th Edition. UK Data Service'. SN: 8644, 10.5255/UKDA-SN-8644-11; Accessed May 16, 2022 4:10 PM.

Krugman, P. (1994), *The Age of Diminished Expectations*, The MIT Press, Cambridge, Massachusetts.

Lagomarsino, E. & Spiganti, A. (2020), 'No gain in pain: psychological well-being, participation, and wages in the BHPS', *European Journal of Health Economics* **21**, 1375–1389.

Lal, A., Lockhart, M., Xu, Y. & Zu, Z. (2021), How much should we trust instrumental variable estimates in political science? practical advice based on over 60 replicated studies, Technical report, Standford University and UCSD.
URL: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=* 3905329

Layard, R. (2013), 'Mental health: the new frontier for labour economics', *Journal of Labour Policy* **2**(2), 1–17.

Lepinteur, A., Clarke, A., Ferrer-I-Carbonell, A., Piper, A., Schröder, C. & D'Ambrosio, C. (2022), Gender, Loneliness and Happiness during COVID-19, Technical Report No 2210, Centre Pour La Recherche Economique Et Ses Applications (CEPREMAP).
URL: *https://www.cepremap.fr/depot/docweb/docweb2210.pdf*

Lerner, D. & Henke, R. (2008), 'What does research tell us about depression, job performance, and work productivity?', *Occupational and Environmental Medicine* **50**(4), 401–410.

Mavridis, D. (2015), 'The unhappily unemployed return to work faster', *Journal of Labor Economics* **4**(1), 1–22.

McCrae, R. & Costa, P. (2008), The five-factor theory of personality, *in* O. John, R. Robins & L. Pervin, eds, 'Handbook of personality: Theory and research (3rd ed.)', The Guildford Press, pp. 186–181.

McInnes, R. (2012), ESA and Incapacity Benefit statistics, Technical Report Standard Note SN01420, House of Commons Library.
URL: *https://commonslibrary.parliament.uk/research-briefings/sn01420/*

Mincer, J. (1958), 'Investment in human capital and personal income distribution', *Journal of Political Economy* **66**(4), 281–302.

Mushtaq, R., Shoib, S., Shah, T. & Mushtaq, S. (2014), 'Relationship between loneliness, psychiatric disorders and physical health? A review on the psychological aspects of loneliness', *Journal of Clinical and Diagnostic Research* **8**(9), WE01–WE04.

Newey, W. K., Powell, J. T. & Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**(3), 565–603.

Nicoletti, C., Salvanes, K. & Tominey, E. (2022), 'Mothers working during pre-school years and child skills. Does income compensate?', *Journal of Labour Economics* **In press**.

Nunn, N. (2008), 'The long-term effects of africa's slave trades', *The Quarterly Journal of Economics* **123**(1), 139–176.

Nunn, N. & Wantchekon, L. (2011), 'The slave trade and the origins of mistrust in africa', *American Economic Review* **101**(7), 3221–3252.

O'Malley, A., Frank, R. & Normand, S.-L. T. (2011), 'Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia', *Statistics in Medicine* **30**, 1971–1988.

Orefice, S. & Quintana-Domeque, C. (2021), 'Gender inequality in COVID-19 time: evidence from UK prolific participants', *Journal of Demographic Economics* **87**, 261–87.

Oswald, A. J., Proto, E. & Sgroi, D. (2015), 'Happiness and productivity', *Journal of Labor Economics* **33**(4), 789–822.

Pai, N. & Vella, S. L. (2021), 'COVID-19 and loneliness: A rapid systematic review', *Aust N Z J Psychiatry* **55**(12), 1144–1156.

Pessoa, J. & Van Reenen, J. (2014), 'The UK productivity and jobs puzzle: Does the answer lie in wage flexibility?', *Economic Journal* **124**(576), 433–452.

Rees, E. & Large, R. (2020), Coronavirus and loneliness, Great Britain: 3 April to 3 May 2020, Technical Report Statistical Bulletin, Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/coronavirusandlonelinessgreatbritain/3aprilto3may2020.

Roberts, J., Bryan, M., Bryce, A., Rice, N. & Sechel, C. (2021), 'Disability Employment Gap: Written evidence to the Work and Pensions Committee Inquiry 2020'. https://committees.parliament.uk/writtenevidence/19219/pdf/.

Sartorius, N. & Ustün, T. B. (1995), *Mental Illness in General Health Care: an International Study*, John Wiley & Sons, New York.

Schmitz, N., Kruse, J. & Tress, W. (1999), 'Psychometric properties of the general health questionnaire (GHQ-12) in a German primary care sample.', *Acta Psychiatrica Scandinavica* **100**(6), 462–468.

Sharpe, A. & Mobasher Fard, S. (2022), 'The current state of research on the two-way linkages between productivity and well-being', ILO Working Paper 56, (Geneva, ILO).

Terza, J. V. (2017), 'Two-stage residual inclusion estimation: A practitioners guide to stata implementation', *The Stata Journal* **17**(4), 916–938.

Terza, J. V., Basu, A. & Rathouz, P. (2008), 'Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling', *Journal of Health Economics* **27**, 531–543.

Tiwari, S. C. (2013), 'Loneliness: a disease?', *Indian Journal of Psychiatry* **55**(4), 320–322.

van Kippersluis, H. & Rietveld, C. (2018), 'Beyond plausibly exogenous', *Econometrics Journal* **21**, 316–331.

von Hinke, S., Rice, N. & Tominey, E. (2022), 'Mental health around pregnancy and child development from early childhood to adolescence', *Labour Economics* **78**, 102245.

Weiss, R. (1973), *Loneliness: The experience of emotional and social isolation*, MIT Press, Cambridge, Massachusetts.

Appendix

Table A1: September wave: Quantification of changes in productivity

	September 2020						
Fall in productivity	Midpoint loss	A little less		Much less		Little or much less	
		Freq.	%	Freq.	%	Freq.	%
Don't know	-	3	0.82	7	4.83	10	1.95
Between 45 and 60 minutes	8	202	55.04	28	19.31	230	44.92
Between 30 and 45 minutes	23	139	37.87	44	30.34	183	35.74
Less than 30 minutes	45	23	6.27	66	45.52	89	17.38
N		367	100	145	100	512	100
Rise in productivity	Midpoint gain	A little more		Much more		Little or much more	
		Freq.	%	Freq.	%	Freq.	%
Refusal	-	1	0.21	0	0	1	0.11
Don't know	-	5	1.05	6	1.39	11	1.21
Up to 75 minutes	8	268	56.18	132	30.56	400	44.00
Between 75 and 90 minutes	23	163	34.17	190	43.98	353	38.83
More than 90 minutes	45	40	8.39	104	24.07	144	15.84
N		477	100	432	100	909	100

The overall sample size is 3,601. 2,180 (60.5%) individuals reported no change in productivity. The question asked of respondents who reported a change in productivity is; top (bottom) panel: "Thinking about how much less (more) you get done these days, would you say that what you can do in an hour now would previously have taken you:" Note that the responses were provided as hours for 60 minutes and above. For example, between 75 minutes and 90 minutes was presented as being "Between an hour and a quarter and an hour and a half". Likewise for other responses. These are presented in minutes in the above to conserve text space.

Table A2: Preferred model: June 2020

	Full sample Coef (s.e.)	Males Coef (s.e.)	Females Coef (s.e.)
GHQdiff	0.083 ***	0.100 ***	0.071 ***
	(0.017)	(0.030)	(0.022)
$\hat{\varepsilon}$	-0.048 ***	-0.062 **	-0.039 *
	(0.018)	(0.032)	(0.023)
Health condition	0.079	-0.004	0.127
	(0.064)	(0.098)	(0.088)
Socio-demographic			
Male	0.007		
	(0.064)		
Age	0.040 **	0.035	0.044 *
	(0.017)	(0.026)	(0.024)
Age^2	0.000 **	0.000	0.000 *
	(0.000)	(0.000)	(0.000)
Degree	0.010	-0.173	0.131
	(0.069)	(0.107)	(0.092)
BAME	0.091	0.026	0.164
	(0.121)	(0.194)	(0.137)
Employment			
Self-employed	-0.332 ***	-0.366 ***	-0.317 ***
	(0.085)	(0.129)	(0.107)
Both self & employed	-0.169	-0.057	-0.267
	(0.155)	(0.251)	(0.204)
Household			
Couple	0.091	0.067	0.074
	(0.078)	(0.137)	(0.093)
Kids 0-4 yo	0.028	0.121	-0.067
	(0.094)	(0.155)	(0.133)
Kids 5-15 yo	-0.182 ***	-0.137	-0.230 **
	(0.068)	(0.118)	(0.090)
Kids 16-18 yo	-0.055	0.079	-0.146
	(0.090)	(0.142)	(0.109)
Industry (Education omitted)			
Agriculture/Forestry/Fishing/Mining/Quarrying	0.227	-0.192	0.581 **
	(0.193)	(0.278)	(0.267)
Manufacturing	0.349 ***	0.223	0.419 *
	(0.134)	(0.199)	(0.226)
Utilities	0.301	0.355	0.166
	(0.193)	(0.253)	(0.316)
Construction	0.327	0.274	0.210
	(0.221)	(0.206)	(0.649)
Wholesale/Retail	0.670 ***	0.736 *	0.591 **
	(0.220)	(0.379)	(0.290)
Repair of Motor Vehicles/Transportation/Storage	0.505 **	0.204	0.691 **
	(0.206)	(0.312)	(0.315)
Accommodation/Food services/Other service activities/HH as employers	0.431 ***	0.336 *	0.438 ***
	(0.108)	(0.181)	(0.141)
Information & communication	0.459 ***	0.431 **	0.411 **
	(0.130)	(0.202)	(0.170)
Financial & Insurance	0.390 ***	0.333 *	0.362 **
	(0.114)	(0.188)	(0.145)
Real estate	0.243	0.247	0.094
	(0.180)	(0.264)	(0.289)
Professional/Scientific/Technical	0.288 **	0.178	0.335 **
	(0.120)	(0.179)	(0.150)
Admin/Support services	0.139	-0.140	0.226
	(0.130)	(0.285)	(0.147)
Public administration & defense	0.053	-0.190	0.268 *
	(0.155)	(0.252)	(0.147)
Human health/Social work	0.502 ***	0.396	0.548 ***
	(0.108)	(0.295)	(0.112)
Arts/Entertainment/Recreation	-0.085	-0.379	0.110
	(0.149)	(0.236)	(0.197)
N	2,902	1,201	1,701

* $p<0.10$, ** $p<0.05$, *** $p<0.01$. Regressions weighted using cross-sectional weights provided with the data.

Table A3: First stage OLS results: June 2020

	Full sample Coef (s.e.)	Males Coef (s.e.)	Females Coef (s.e.)
Health condition	-0.567 *	-0.162	-0.871 **
	(0.334)	(0.468)	(0.432)
Socio-demographic			
Male	0.024		
	(0.330)		
Age	-0.090	0.086	-0.297 **
	(0.082)	(0.100)	(0.133)
Age^2	0.001	-0.001	0.003 **
	(0.001)	(0.001)	(0.002)
Degree	0.296	0.851 **	-0.381
	(0.339)	(0.432)	(0.504)
BAME	0.727	0.840	0.669
	(0.512)	(0.594)	(0.827)
Employment			
Self-employed	-0.841 **	-1.037 **	-0.283
	(0.401)	(0.479)	(0.642)
Both self & employed	-0.532	-0.910	0.003
	(0.758)	(1.378)	(0.738)
Household			
Couple	0.537	0.881	0.475
	(0.436)	(0.609)	(0.570)
Kids 0-4 yo	-0.441	-0.523	-0.342
	(0.461)	(0.653)	(0.673)
Kids 5-15 yo	-0.362	-0.163	-0.611
	(0.383)	(0.472)	(0.552)
Kids 16-18 yo	0.466	-0.306	1.068 **
	(0.343)	(0.481)	(0.472)
loneliness			
less lonely	1.762 ***	2.037 ***	1.578 ***
	(0.385)	(0.536)	(0.505)
more lonely	-4.661 ***	-4.054 ***	-5.027 ***
	(0.505)	(0.711)	(0.679)
Industry (Education omitted)			
Agriculture/Forestry/Fishing/Mining/Quarrying	1.298	2.560	-0.905
	(1.354)	(1.832)	(1.427)
Manufacturing	0.165	-0.629	0.531
	(0.509)	(0.660)	(0.842)
Utilities	-0.282	-1.692	0.416
	(0.869)	(1.198)	(1.124)
Construction	-0.042	-0.029	-1.409
	(0.597)	(0.682)	(1.037)
Wholesale/Retail	-0.298	-1.848	0.443
	(1.620)	(1.636)	(2.503)
Repair of Motor Vehicles/Transportation/Storage	0.949	0.355	1.493
	(1.056)	(1.529)	(1.394)
Accommodation/Food services/Other service activities/HH as employers	-0.504	-1.297 **	-0.116
	(0.459)	(0.636)	(0.635)
Information & communication	-0.789	-1.695 **	-0.319
	(0.626)	(0.850)	(0.831)
Financial & Insurance	0.195	-0.493	0.205
	(0.512)	(0.735)	(0.736)
Real estate	-0.229	-1.032	0.567
	(0.758)	(1.032)	(1.087)
Professional/Scientific/Technical	-0.030	-0.747	0.333
	(0.457)	(0.603)	(0.658)
Admin/Support services	-0.805	0.917	-1.271
	(1.240)	(1.120)	(1.494)
Public administration & defense	-0.271	-0.885	0.097
	(0.521)	(0.716)	(0.710)
Human health/Social work	-0.691	-3.361 *	0.448
	(0.716)	(1.959)	(0.603)
Arts/Entertainment/Recreation	-0.114	-0.274	-0.813
	(0.975)	(1.291)	(1.291)
Constant	0.535	-3.491	5.393 **
	(1.803)	(2.336)	(2.674)
N	2,902	1,201	1,701

* p<0.10, ** p<0.05, *** p<0.01. Regressions weighted using cross-sectional weights provided with the data.

Table A4: Analysis of Local Average Treatment Effects

	Main effects		Interactions with main effects			
	less lonely	More lonely	less lonely	More lonely	F-stat (2, 2870)	p -value
Full model	1.76	-4.66				
Full mode with interaction terms						
Health condition	1.815	-4.395	-0.131	-0.618	0.16	0.854
Male	1.516	-5.096	0.569	1.132	0.72	0.487
Degree	1.133	-6.869	0.897	3.254	3.77	0.023
BAME	1.738	-4.965	0.454	3.583	2.32	0.099
self-employed	1.856	-4.819	-0.846	1.273	1.01	0.365
Both self & employed	1.744	-4.729	0.384	1.905	0.59	0.556
Couple	1.320	-6.204	0.581	2.349	1.68	0.186
Kids 0-4 yo	1.746	-4.753	0.142	0.714	0.16	0.854
Kids 5-15 yo	1.383	-4.778	1.267	0.382	1.21	0.298
Kids 16-18 yo	1.867	-4.858	-0.794	2.493	2.94	0.053

Table shows first-stage regression results of MH on the set of covariates and the instrument (change in loneliness), together with interaction terms for main effects with the instrument. The final two columns provide test statistics for the null of no effect of the interaction terms. Analyses undertaken on the June 2020 sample for individuals working from home at least some of the time.

Table A5: Ordered probit regression of the instrument on covariates

	Full sample Coef (s.e.)	Males Coef (s.e.)	Females Coef (s.e.)	Conditional sample Coef (s.e.)
Health condition	0.156**	0.246**	0.056	0.120
Socio-demographic	(0.069)	(0.107)	(0.088)	(0.114)
Male	-0.293***			0.016
	(0.072)			(0.120)
Age	0.023	0.036	0.027	-0.050
	(0.022)	(0.027)	(0.030)	(0.039)
Age^2	-0.0004*	-0.0005*	-0.0005	0.0004
	(0.0002)	(0.0003)	(0.0003)	(0.0004)
Degree	-0.045	-0.107	0.027	-0.012
	(0.083)	(0.123)	(0.104)	(0.137)
BAME	0.191*	0.230	0.124	-0.164
	(0.115)	(0.171)	(0.148)	(0.202)
Employment				
Self-employed	-0.128	-0.149	-0.108	0.095
	(0.102)	(0.153)	(0.137)	(0.188)
Both self & employed	0.037	0.151	0.017	-0.409
	(0.127)	(0.189)	(0.161)	(0.239)
Household				
Couple	-0.209**	-0.546***	-0.092	-0.157
	(0.087)	(0.147)	(0.104)	(0.131)
Kids 0-4 yo	-0.056	-0.081	0.051	0.141
	(0.109)	(0.167)	(0.135)	(0.180)
Kids 5-15 yo	-0.116	0.185	-0.332***	-0.009
	(0.078)	(0.120)	(0.102)	(0.136)
Kids 16-18 yo	-0.131	-0.092	-0.151	-0.306
	(0.093)	(0.137)	(0.122)	(0.192)
Industry (Education omitted)				
Agriculture/Forestry/Fishing/Mining/Quarrying	-0.376	-0.607	-0.230	0.130
	(0.320)	(0.461)	(0.430)	(0.547)
Manufacturing	-0.054	-0.121	0.211	-0.141
	(0.161)	(0.230)	(0.253)	(0.292)
Utilities	-0.128	0.227	-0.600*	-0.004
	(0.234)	(0.315)	(0.327)	(0.360)
Construction	0.269	0.101	0.697	0.467
	(0.323)	(0.249)	(0.559)	(0.356)
Wholesale/Retail	-0.274	-0.156	-0.345	0.040
	(0.224)	(0.327)	(0.292)	(0.350)
Repair of Motor Vehicles/Transportation/Storage	0.168	0.272	0.078	-0.122
	(0.211)	(0.311)	(0.294)	(0.408)
Accommodation/Food services/Other service activities/HH as employers	0.075	0.146	0.084	-0.393
	(0.115)	(0.206)	(0.140)	(0.209)
Information & communication	0.045	0.052	0.160	-0.174
	(0.149)	(0.226)	(0.192)	(0.239)
Financial & Insurance	-0.218*	0.062	-0.452***	-0.632
	(0.125)	(0.216)	(0.157)	(0.239)
Real estate	0.027	-0.273	0.378	-0.579
	(0.198)	(0.336)	(0.267)	(0.390)
Professional/Scientific/Technical	0.074	0.175	-0.028	-0.448
	(0.120)	(0.202)	(0.165)	(0.205)
Admin/Support services	0.154	0.469	0.052	0.139
	(0.174)	(0.322)	(0.208)	(0.272)
Public administration & defense	-0.108	0.016	-0.201	-0.331
	(0.157)	(0.264)	(0.174)	(0.263)
Human health/Social work	-0.116	-0.212	-0.093	-0.066
	(0.112)	(0.246)	(0.125)	(0.193)
Arts/Entertainment/Recreation	0.134	0.334	0.042	-0.174
	(0.193)	(0.331)	(0.232)	(0.297)
N	2,902	1,201	1,701	864

Table shows regression coefficients and standard errors of the change in loneliness regressed on the set of covariates. These are presented for the full sample, and separately for males and females. Analyses undertaken on the June 2020 sample for individuals working from home at least some of the time. The final column reports results conditional on reporting either 'less lonely' or 'more lonely'. Regressions weighted using cross-sectional weights provided with the data. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A6: First-stage coefficients for the benchmark model versus a model with no controls

	No controls Coef (s.e.)	Benchmark Coef (s.e.)	t-statistic
Full sample			
Less lonely	1.738	1.762	-0.044
	(0.393)	(0.385)	
More lonely	-4.786	-4.661	-0.167
	(0.554)	(0.505)	
Males			
Less lonely	2.040	2.037	0.004
	(0.543)	(0.536)	
More lonely	-4.204	-4.054	-0.142
	(0.777)	(0.711)	
Females			
Less lonely	1.522	1.578	-0.075
	(0.544)	(0.505)	
More lonely	-5.130	-5.027	-0.102
	(0.743)	(0.679)	

Table shows first-stage regression results of MH on the instrument (change in loneliness). The benchmark model conditions on the set of controls. This is provided for the full sample and separately for males and females. Analyses undertaken on the June 2020 sample for individuals working from home at least some of the time. Regressions weighted using cross-sectional weights provided with the data.

Table A7: Robustness checks (June 2020 sample for individuals working from home at least some of the time)

	Employed Coef (s.e.) (1)	Salaried Coef (s.e.) (2)	With baselines wages Coef (s.e.) (3)
Ordered Probit			
GHQdiff	0.035 ***	0.039 ***	0.039 ***
	(0.005)	(0.006)	(0.005)
2SRI			
GHQdiff	0.070 ***	0.084 ***	0.077 ***
	(0.018)	(0.019)	(0.017)
$\hat{\varepsilon}$	-0.039 **	-0.049 **	-0.042 **
	(0.018)	(0.020)	(0.018)
1st stage F-stat	53.04	50.02	63.00
N	2,314	2,096	2,646

* p<0.10, ** p<0.05, *** p<0.01. Bootstrapped standard errors in parentheses (1000 repetitions). Sample summary statistics are weighted using cross-sectional weights provided with the data. Coefficient on baseline wages in model (3) is positive and significant at the 1% level (0.0002956).

Table A8: Heterogeneity Analysis (June 2020 sample for individuals working from home at least some of the time)

	Under 50 Coef (s.e) (1)	50 plus Coef (s.e) (2)	men under 50 Coef (s.e) (3)	men 50 plus Coef (s.e) (4)	women under 50 Coef (s.e) (5)	women 50 plus Coef (s.e) (6)	good MH Coef (s.e) (7)	bad MH Coef (s.e) (8)
Ordered Probit								
GHQdiff	0.037 ***	0.044 ***	0.046 ***	0.037 ***	0.034 ***	0.049 ***	0.049 ***	0.041 ***
	(0.006)	(0.008)	(0.011)	(0.012)	(0.008)	(0.009)	(0.007)	(0.009)
2SRI								
GHQdiff	0.077 ***	0.113 ***	0.086 *	0.121 ***	0.068 ***	0.097 **	0.090 ***	0.085 *
	(0.020)	(0.030)	(0.045)	(0.044)	(0.026)	(0.042)	(0.020)	(0.047)
$\hat{\varepsilon}$	-0.045 **	-0.073 **	-0.043	-0.091 **	-0.039	-0.052	-0.046 **	-0.046
	(0.020)	(0.032)	(0.047)	(0.046)	(0.027)	(0.044)	(0.021)	(0.049)
1st stage F-stat	43.98	32.22	15.12	15.85	34.43	18.49	53.30	12.08
N	1,534	1,368	585	616	949	752	2,349	553
Bootstrap reps	1,000	1,000	983	1,000	942	991	1,000	976

* p<0.10, ** p<0.05, *** p<0.01. Bootstrapped standard errors in parentheses (1000 repetitions). Sample summary statistics are weighted using cross-sectional weights provided with the data.