Supplementary Material for "Identifying employee, workplace, and population characteristics associated with COVID-19 outbreak clusters in the workplace: a population-based study"

Christopher E. Overton*[1,2,3], Rachel Abbey[4], Tarrion Baird[2,5], Rachel Christie[2,4], Owen Daniel[6], Julie Day[2,4], Matthew Gittins[7], Owen Jones[2], Robert Paton[2], Maria Tang[2], Tom Ward[2], Jack Wilkinson[7], Camilla Woodrow-Hill[2,8], Tim Aldridge[9], Yiqun Chen[9]

**corresponding author: c.overton@liverpool.ac.uk.*
[1]Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK
[2]All Hazards Intelligence, Data, Analytics and Surveillance, UK Health Security Agency, London, SW1P 3JR, UK
[3]Department of Mathematics, University of Manchester, Manchester, M13 9PL, UK
[4]Analytics and Data Science, Data, Analytics and Surveillance, UK Health Security Agency, London, SW1P 3JR, UK
[5]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK
[6]Data Science Campus, Office for National Statistics, UK
[7]Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, M13 9NT, UK
[8]Division of Psychology Communication and Human Neuroscience, University of Manchester, Manchester, M13 9NT, UK
[9]Science Division, Health and Safety Executive, Buxton, SK17 9JN, UK

S.1. Supplementary methods

S.1.1. Data preparation

S.1.1.1. Flow to work

Many of the variables in the datasets describe the residential characteristics of workers. To obtain employee (that is, a worker, who travels from home to work) demographics, we need to understand where these individuals work. To do this, we use the flow to work data from the 2011 census. For each residential MSOA of where the workers lived, we calculate the proportion of workers going to each workzone in England. This provides a network describing where individuals are expected to work.

S.1.1.2. Industry distributions

Analysis was performed with MSOA as the unit of analysis. Therefore, we aggregated the employee data from workzone level to MSOA level. To approximate the distribution of employees across industries in each MSOA, we weighted the workzone level variables by the industry distribution across that workzone. This led to the assumption that within an individual workzone, employees are uniformly distributed across the industries, based on the proportion of each industry present. At MSOA level, this led to the employee characteristics being weighted by the industry distributions in the workzones to which these individuals commute.

S.1.1.3. COVID-19 employee case rates

To understand background risk, we quantified the incoming infection risk for each workplace, which is a function of community prevalence. We did this by converting the community COVID-19 case rates into a employee COVID-19 case rate. We began by calculating the case rate for each MSOA through time; we geographically matched Pillar 2 SARS-CoV-2 PCR tests to MSOA by postcode, removing travel associated tests. Only symptomatic tests were modelled, to avoid the confounder of different asymptomatic screening policies of different workplaces. These tests were relatively sparse at the MSOA level, so we smoothed case rates using a generalised additive model (GAM) with a quasi-poisson error structure and log-link. The number of positive tests, n, is smoothed for each MSOA, using a separate cubic regressions spline for each MSOA, $f(t,m)$. We expect that the case rates of neighbouring MSOAs would be more closely correlated than those that do not share a border. We therefore fit a Markov random field term that adjusts MSOA intercepts ($\propto_m$) based on its neighbours. We label this term $g(m)$. In full, this gives the expression:

$$\log(\lambda_{t,m}) = \propto_m + \beta_d + g(m) + f(t,m) + \log(P)$$

We used a log-offset term in the model to scale the number of tests by the MSOA population size (P), to give a per-capita rate. A day of week effect controlled for weekly cycles in test reporting and was modelled as a random effect. By modelling the day of week effect in this way, we could omit it from the modelled case rate giving only the central trend. This allows us to better describe the true rate, as opposed to weekly anomalies in reporting and testing. Only people aged 18-64 were included in the population offset term and testing data to better reflect the working age population per MSOA. To save on computational time, we analysed MSOA case rates at the local authority level; the data were subset to specific lower-tier local authority (LTLA) and the model described above fitted for all time points within the geography.

To calculate an incoming force of infection into each sector in each workzone MSOA, we began by averaging MSOA case rate across each week of the study period ($\bar{\lambda}_m$). For each workzone, there was a measure of how many people commuted from their resident MSOA into that workzone. For each workzone with $N_{m,w}$ employees from M_w resident MSOAs to the workzone:

$$C_w = \frac{\sum_{m=1}^{M_w} \bar{\lambda}_m \times N_{m,w}}{\sum_{m=1}^{M_w} N_{m,w}}.$$

Here, C_w is the estimated number of infected employees out of the total number of employees coming into the workzone. For each workzone, we knew the proportion of businesses in workzone belonging to each "industry" ($\rho_{w,i}$). We assume that these employees uniformly attended each industry, so to calculate a per-industry, per-workzone per-capita case rate we scaled C_w by $\rho_{w,i}$.

S.1.1.4. COVID-19 vaccination

To provide a measure of COVID-19 vaccination levels, we counted the expected proportion of employees into each workplace who have had two or more vaccination doses. We also considered the counts of the number of residents in each workplace MSOA who have had two or more vaccination doses.

S.1.1.5. COVID-19 outbreaks

We were interested in the number of active outbreaks in each MSOA, broken down by industry. To calculate this, we first define an "event" as an occasion where a COVID-19 case selfreported as being present at a location/potential site of infection, recorded via the backwards contact-tracing process. An outbreak "cluster" is then defined as a collection of events which occurred at a single location (common exposure event), defined by the Unique Property Reference Number (UPRN) as a spatial reference identifier associated with the event in the CTAS data; and where each event occurs within some episode period (in our case, defined as six full days) of the previous event. An outbreak cluster is "active" on a given date if that date lies between the dates of the chronologically first and last events which belong to the outbreak, inclusive of those dates.

As a persistent identifier, the UPRN was additionally used to attach industry information to each outbreak. For each UPRN, full address information was extracted from the subset of OS AddressBase Plus corresponding to places of work. This address information was then linked to the Workplaces layer of the National Population Database, using a combination of best-match organisation name matching, and direct postcode matching. The NPD Workplaces layer includes an industry code for each entry, which can be used to derive the broader industry which the workplace belongs.

S.1.1.6. Employee age, sex, and ethnicity

From the 2011 Census, we had data on the age, sex and ethnicity of employees by workplace zone. We grouped the ages into 4 age groups: 18-29; 30-44; 44-59; 60-64. Ages outside of this range were removed since this is the typical workforce age. Sex was grouped into Male and Female. Ethnicity was grouped into White, Asian, Black/African/Caribbean, and Mixed/Multiple/Other.

For each workzone, we calculated the number of workers in each of these demographic groups. We then calculated the expected proportion of workers in each industry using the NPD workplace layer. We calculated the expected number of workers in each demographic group in each industry by multiplying the number of workers in that demographic group by the proportion of workers in that industry. The data were then aggregated up to MSOA level by summing the number of workers in each demographic group and industry across all workzones. From this aggregated data, the proportion of workers within each demographic group in that industry and MSOA was calculated by dividing by the total number of workers in that industry and MSOA. This was then converted to a percentage.

S.1.1.7. Employee work commute mode

From the 2011 Census, we had data on the age of employees by residential MSOA. These data were grouped by mode of travel: Train; Taxi (or vehicle passenger); Single occupancy; Bus, metro or tram; other.

Since these data are at residential MSOA rather than workplace zone, we first needed to map these to workplace zone level, before following the procedure described above for age, sex, and ethnicity. To map these data to workplace zone, we used the travel to work data from the 2011 census. From this, we could calculate the expected number of employees leaving the focal MSOA to work in the target workplace zone. Multiplying this by the proportion of employees using each mode of transport gives the expected number of employees using each mode of transport who live in the residential MSOA and work in the workzone. Repeating this for all residential MSOAs that commute into the target workzone provides the total expected number of employees into the target workzone for each mode of transport. We then followed the procedure above to aggregate by industry and MSOA.

S.1.1.8. Employee IMD

Using the Department for Levelling Up, Housing and Communities indices of multiple deprivation (IMD) classifications (13), we had data on the age of employees by residential MSOA. Using this, we calculated the distribution across IMD deciles in each residential MSOA. We then followed the mapping procedure described above for the mode of transport data to calculate the expected IMD distribution of employees into each workplace MSOA. From this, we calculated the mean IMD of employees into each workplace MSOA.

S.1.1.9. Workplace mobility class

The workplace mobility dataset classifies each LSOA into one of eight levels: with level 1 being fully metropolitan, and 8 fully rural. We regroup these into 4 levels: Metropolitan (L1 and L2), Exurban (L3 and L4), Suburban (L5 and L6), and Rural (L7 and L8). To obtain an MSOA level metric, we defined each MSOA by the most common mobility class amongst LSOAs within that MSOA.

S.1.1.10. Workplace broad industry sector

From the NPD workplace layer, we had data on each workplace, providing the name, address, industry sector, and number of employees. For each workzone, we grouped the data by industry sector and counted the number of workplaces and number of employees in that sector within that workzone.

S.1.1.11. Workplace proportion of employees on permanent contracts

This data (22) provides the proportion of employees on permanent contracts (permanence) for each two-digit SIC division code (1-99). These SIC groupings are nested within the industry sector definitions used in this study. For each MSOA and industry, we calculated the number of employees within each SIC grouping. We then calculated the average permanence across all SIC groupings within that industry in that MSOA. This was done by multiplying the permanence by the number of employees in each SIC grouping, summing across all SIC groupings within that industry, and then dividing by the number of employees in that industry in that MSOA.

S.1.1.12. Workplace relative measure of physical proximity

This data (21) provides a relative measure of physical proximity between employees in the workplace (proximity) for each SIC division code (1-99). This data is derived from the Annual Population Survey, and uses the following proximity scores: 0 - I do not work near other people (beyond 100 ft.); 25 - I work with others but not closely (for example, private office); 50 - Slightly close (for example, shared office); 75 - Moderately close (at arm's length); 100 - Very close (near touching). The data takes the average score for each 2-digit SIC division code. These SIC groupings are nested within the industry sector definitions used in this study. For each MSOA and industry, we calculate the number of employees within each SIC grouping. We then calculate the average proximity across all SIC groupings within that industry in that MSOA. This is done by multiplying the proximity by the number of employees in each SIC grouping, summing across all SIC groupings within that industry, and then dividing by the number of employees in that industry in that MSOA.

S.1.1.13. Neighbourhood characteristics

In addition to the employee characteristics, we consider some characteristics of the neighbourhood population (individuals who live near to the workplace) as potential confounder variables. These are: resident IMD quintile, the most common IMD quintile among residents in the workplace MSOA; resident COVID-19 case rate, the proportion of residents in the workplace MSOA testing positive for COVID-19; resident vaccination dose 1, the proportion of residents in the workplace MSOA with 1 or more vaccination dose; resident vaccination dose 2, the proportion of residents in the workplace MSOA with 2 or more vaccination doses.

S.1.2. Identifying confounder sets for statistical analysis

Figure S1 illustrates the Directed Acyclic Graph (DAG) used to represent predefined hypothetical causal relationships between HSE key priorities relating to industry, employee, and local neighbourhood characteristics and outbreaks of COVID infections. This DAG assumes we are interested in the acute effects of each priority, defined here as their relationship with the following week. The HSE priorities defined prior to the analysis took place were:

Workplace Characteristics
- Location
- Size (Number in Employment)
- Sector (Theme 1 Sector)
    - Employment Type (Temporary/Zero Hours)
    - Younger workers (Employment by Age)
    - Physical Proximity

Employee Travel Characteristics (Employee)
- Work commute mode

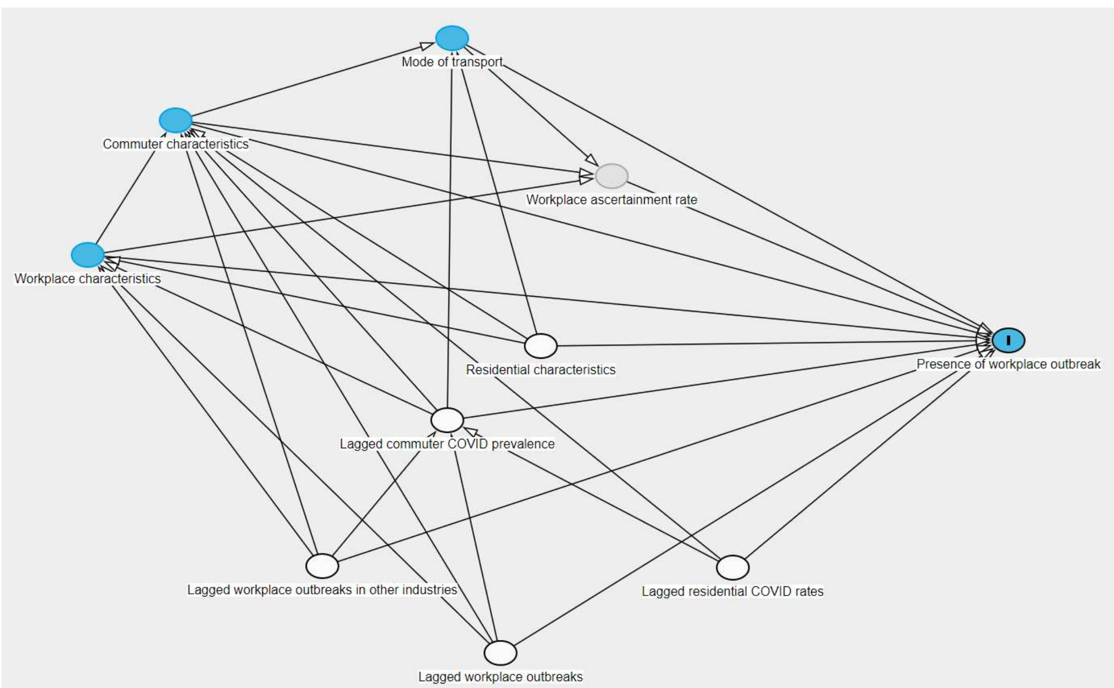
Employee Demographic Characteristics (socio-economic)
- Indices of Multiple Deprivation
- Ethnicity Group
- Vaccination

These priorities were translated on the Figure S1 DAG into three risk factor groups
1. Workplace Characteristics
2. Employee Characteristics
3. Work Commute Mode

We then proposed the following three confounder sets for each of the HSE priorities, no confounders, minimal confounders, and then HSE priority specific full confounders. In each case we aim to capture the total association between the HSE priority risk factor and outcome, and so aim to not adjust for characteristics later in the causal path.

1. Unadjusted model, i.e. risk factor only with no adjustment for confounding.
2. Minimal adjusted model, i.e. common confounders to all risk factors including the lagged COVID rate variables, MSOA mobility cluster (local to the workplace), and neighbourhood characteristics (local to the workplace).
3. Priority risk factor fully adjusted model, i.e. also adjusted for all variables in the two risk factor groups occurring prior in the hypothetical causal chain to the HSE priority. For example, if we examine a variable in the workplace characteristic risk factor group, we would adjust for mode of transport characteristics, and employee characteristics. Note, variables within a risk factor group will not be adjusted for due to the complexity of the causal relationships present.

Figure S1: *Proposed DAG of hypothetical causal relationships between key MSOA industry characteristics, employee characteristics, work commute mode, and neighbourhood characteristics, and outbreak clusters of COVID infections. White variables are the common confounder set, grey variables are unobserved, and blue variables are our key variables of interest.*



S.1.3. Bias

The main bias in our data is that different industries will have different testing rates. For example, some industries may have had mandatory testing mandates, requiring all employees to get regularly tested. Therefore, since the outbreaks are defined in terms of positive tests rather than infections, the probability of outbreaks being identified are not comparable across industries. To avoid this bias, the primary analysis uses industry-specific models, and we do not compare across industry. In the interpretation of our results, we have converted the effect sizes to instead report the relative increase in outbreak risk from changing the focal variable. Since this is a relative measure, these can be compared across industries.

Changes to testing policy would affect the analysis conducted. Over the study period, testing policy was reasonably consistent temporally. However, we do not account for spatial variation in testing policy. If some areas have fewer tests at random, this will be adjusted for in the Markov Random Field used to smooth the MSOA level testing data. However, if there is a spatial bias in the testing policy (for example a spatial bias in test availability), this will not be captured by our model. Therefore, both the employee COVID-19 case rates and the number of outbreak clusters could have some spatial bias, which could be amplified by work from policies (33).

The final major bias is in the permanence and proximity variables. These rely on having sufficient variation within each industry sector in order to be used. However, for some industries, there is insufficient variation, with just a single value for these variables for all MSOAs, or highly correlated permanence and proximity. For such scenarios, these variables do not change enough to accurately measure their influence on workplace outbreaks. To account for this, we only consider these variables for the following industry sectors: Services; Utilities; Transport, distribution and warehousing; Manufacturing; and Construction.

S.2. Supplementary tables

Table S1: *Variables considered in this study. Risk factor variables are the risk factors of interest in this study. Confounder variables are included in the minimum confounder set.*

Variable	Type	Group
Physical proximity in the workplace	Risk factor	**Workplace**
Proportion of workers on permanent contracts	Risk factor	**Workplace**
Mobility class	Risk factor	**Workplace**
Proportion of employees with 2 or more vaccination doses	Risk factor	**Employee**
Proportion of employees aged 18-29	Risk factor	**Employee**
Proportion of employees aged 30-44	Risk factor	**Employee**
Proportion of employees aged 45-59	Risk factor	**Employee**
Proportion of employees aged 60-64	Risk factor	**Employee**
Proportion of employees with an asian ethnicity	Risk factor	**Employee**
Proportion of employees with a black/african/caribbean ethnicity	Risk factor	**Employee**
Proportion of employees with a mixed/multiple/other ethnicity	Risk factor	**Employee**
Proportion of employees with a white ethnicity	Risk factor	**Employee**
Proportion of employees who identify as Female	Risk factor	**Employee**
Proportion of employees who identify as Male	Risk factor	**Employee**
Employee indices of multiple deprivation (IMD) quintile	Risk factor	**Employee**
Proportion of employees using bus/metro/tram	Risk factor	**Work commute mode**
Proportion of employees using taxi/vehicle passenger	Risk factor	**Work commute mode**

Proportion of employees using other transport	Risk factor	**Work commute mode**
Proportion of employees using train	Risk factor	**Work commute mode**
Proportion of employees using single occupancy	Risk factor	**Work commute mode**
Workplace COVID-19 outbreak rate	Confounder	**COVID-19 case and outbreak rates**
Workplace COVID-19 outbreak rate - other industries	Confounder	**COVID-19 case and outbreak rates**
Employee COVID-19 case rate	Confounder	**COVID-19 case and outbreak rates**
Neighbourhood indices of multiple deprivation (IMD) quintile	Confounder	**Neighbourhood**
Proportion of neighbourhood residents with 1 or more vaccination doses	Confounder	**Neighbourhood**
Proportion of neighbourhood residents with 2 or more vaccination doses	Confounder	**Neighbourhood**
Neighbourhood COVID-19 case rate	Confounder	**Neighbourhood**

Table S2: *Unadjusted model, percentage change in risk of workplace COVID-19 outbreak rate by industry.*

Variable	Percentage change in risk by industry (unadjusted model)										
	Services	Utilities	Education	Transport, distribution and warehousing	Mining and Quarrying	Manufacturing	Public service activities	Construction	Human health and social work	Waste management and remediation	Agriculture, forestry and fishing
Workplace - proximity											
Physical proximity in the workplace	3.5 (3.3,3.8)	26 (22,30)	NA	18 (17,19)	NA	6.6 (6.1,7.1)	NA	-9.7 (-11,-8.1)	NA	NA	NA
Workplace - permanence											
Proportion of workers on permanent contracts	0.22 (-0.37,0.81)	-27 (-31,-23)	NA	-43 (-44,-41)	NA	-20 (-21,-20)	NA	-43 (-56,-27)	NA	NA	NA
Workplace - mobility class											
Mobility class - Exurban	-30 (-31,-28)	-9.8 (-21,2.9)	-13 (-14,-11)	-16 (-19,-12)	93 (38,1.7e+02)	-13 (-15,-10)	-38 (-41,-35)	-31 (-36,-25)	-12 (-14,-9.6)	5.1 (-9,21)	-26 (-37,-12)
Mobility class - Metropolitan	-18 (-19,-16)	-39 (-45,-32)	-10 (-12,-8.4)	-15 (-17,-12)	-74 (-85,-52)	-15 (-18,-12)	13 (8.8,17)	0.079 (-6,6.6)	-11 (-13,-9.4)	52 (33,74)	7e+02 (1.5e+02,4.7e+0
Mobility class - Rural	-32 (-34,-30)	-31 (-44,-15)	-53 (-55,-52)	-17 (-22,-13)	15 (-21,68)	-17 (-21,-14)	16 (8,25)	-37 (-44,-29)	-6.2 (-10,-1.8)	52 (26,84)	-71 (-76,-65)
Mobility class - Suburban	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Employee - vaccination											
Proportion of employees with 2 or more vaccination doses - Delta	-1.6 (-1.7,-1.5)	-2.6 (-3.5,-1.7)	3.1 (2.9,3.2)	-1.1 (-1.3,-0.83)	3.9 (0.84,7)	-1.1 (-1.3,-0.84)	-0.3 (-0.66,0.059)	-2.9 (-3.4,-2.4)	0.56 (0.36,0.75)	-2.6 (-3.8,-1.4)	-0.53 (-2.1,1.1)
Proportion of employees with 2 or more vaccination doses - Omicron	-2.7 (-2.9,-2.5)	-4.7 (-6.2,-3.1)	0.5 (0.27,0.73)	-3.1 (-3.5,-2.7)	2.3 (-3.7,8.6)	-1.7 (-2.1,-1.3)	-4.2 (-4.8,-3.6)	-4.8 (-5.8,-3.9)	0.59 (0.34,0.83)	-0.11 (-1.9,1.7)	-17 (-19,-14)
Employee - age											
Proportion of employees aged 18-29	3 (2.7,3.2)	3.6 (2.4,4.7)	-0.66 (-0.96,-0.37)	1.2 (0.73,1.7)	-0.78 (-4.6,3.2)	-1.8 (-2.3,-1.3)	-0.81 (-1.4,-0.26)	-0.58 (-1.7,0.57)	-4.5 (-4.9,-4.2)	5.4 (3.3,7.6)	3 (0.31,5.7)
Proportion of employees aged 30-44	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Proportion of employees aged 45-59	6.8 (6.4,7.1)	9 (7.7,10)	2.7 (2.4,3)	6.5 (5.9,7)	13 (9.2,18)	5.5 (5,6)	-1.1 (-1.6,-0.51)	2.7 (1.5,3.8)	3.1 (2.7,3.5)	7.9 (5.8,10)	-3.4 (-6.5,-0.13)
Proportion of employees aged 60-64	-11 (-11,-10)	-9.3 (-12,-6.9)	-6.9 (-7.3,-6.4)	-14 (-15,-13)	-16 (-21,-12)	-20 (-21,-20)	-8.1 (-9.2,-7)	-13 (-15,-11)	-10 (-11,-9.5)	-3.1 (-6.2,0.21)	-20 (-25,-16)
Employee - ethnicity											
Proportion of employees with an asian ethnicity	0.044 (-0.073,0.16)	4.4 (3.2,5.5)	-0.22 (-0.35,-0.094)	-1.1 (-1.4,-0.92)	4.6 (0.64,8.8)	-0.66 (-0.88,-0.44)	-0.17 (-0.48,0.14)	0.41 (-0.13,0.96)	0.59 (0.44,0.74)	0.029 (-1.1,1.2)	3.7 (1.6,5.8)
Proportion of employees with a black/african/caribbean ethnicity	1.7 (1.4,2)	6.9 (4.9,9)	-1.4 (-1.7,-1.1)	3.4 (2.8,4)	14 (-13,48)	1.4 (0.47,2.3)	1 (0.49,1.5)	1.9 (0.6,3.3)	3.4 (3.1,3.7)	7.9 (5,11)	-0.94 (-9.2,8.1)
Proportion of employees with a mixed/multiple/other ethnicity	-9.4 (-10,-8.7)	-35 (-38,-31)	0.62 (-0.11,1.4)	-7.3 (-8.7,-5.8)	-52 (-65,-33)	-11 (-12,-8.8)	7.2 (5.4,9)	-5 (-7.9,-1.9)	-13 (-14,-12)	-11 (-17,-4.6)	84 (58,1.1e+02)
Proportion of employees with a white ethnicity	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Employee - sex											
Proportion of employees who identify as Female	-0.44 (-0.53,-0.35)	-0.66 (-1,-0.27)	3 (2.9,3.1)	-1.4 (-1.6,-1.3)	-5.1 (-6,-4.1)	-3.5 (-3.6,-3.3)	-1.5 (-1.7,-1.4)	-2.8 (-3.1,-2.4)	3.1 (3,3.3)	0.92 (0.36,1.5)	0.14 (-1.3,1.6)
Proportion of employees who identify as Male	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Employee - IMD											
Employee IMD quintile	-14 (-15,-13)	-26 (-31,-21)	3.5 (2.4,4.5)	-8.7 (-10,-7)	-42 (-53,-27)	-15 (-17,-14)	-4.5 (-6.8,-2.1)	-23 (-26,-20)	-7.8 (-8.9,-6.6)	-3.7 (-11,3.7)	-10 (-19,0.15)
Method of travel to work											
Proportion of employees using bus/metro/tram	0.43 (0.35,0.52)	-0.6 (-1.1,-0.059)	-0.71 (-0.82,-0.6)	0.77 (0.57,0.97)	-5.2 (-9.1,-1.1)	0.81 (0.55,1.1)	0.17 (-0.048,0.39)	-0.13 (-0.51,0.26)	0.84 (0.71,0.97)	3.2 (2,4.5)	1.4 (-1.1,3.9)
Proportion of employees using taxi/vehicle passenger	2 (0.029,4.1)	-13 (-24,0.89)	-18 (-20,-17)	-11 (-14,-7.2)	-51 (-63,-35)	-35 (-38,-32)	15 (11,20)	-30 (-37,-22)	-16 (-18,-13)	12 (-3.7,31)	-29 (-40,-15)
Proportion of employees using other transport	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Proportion of employees using train	8.8 (8.4,9.2)	-0.49 (-2.2,1.3)	-3.5 (-3.9,-3)	5.7 (5.1,6.3)	-11 (-16,-6)	7.1 (6.6,7.7)	-6.1 (-6.9,-5.2)	7.2 (5.6,8.9)	0.92 (0.4,1.4)	-2.6 (-4.8,-0.38)	14 (12,17)
Proportion of employees using single occupancy	-0.91 (-1.1,-0.73)	-5.9 (-6.6,-5.1)	-1.6 (-1.8,-1.3)	-1.3 (-1.6,-0.98)	-2.9 (-8.1,2.5)	-6.9 (-7.5,-6.3)	-0.95 (-1.2,-0.68)	1.8 (1.3,2.4)	-4.1 (-4.4,-3.8)	-3.3 (-5.7,-0.86)	9.2 (5.7,13)

Table S3: *Minimal adjusted model, percentage change in risk of workplace COVID-19 outbreak rate by industry. NA values for workplace characteristics since there is no minimal adjusted model for these* *variables.*

Variable	Percentage change in risk by industry (partially adjusted model)										
	Services	**Utilities**	**Education**	**Transport, distribution and warehousing**	**Mining and Quarrying**	**Manufacturing**	**Public service activities**	**Construction**	**Human health and social work**	**Waste management and remediation**	**Agriculture, forestry and fishing**
Workplace - proximity											
Physical proximity in the workplace	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Workplace - permanence											
Proportion of workers on permanent contracts	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Workplace - mobility class											
Mobility class - Exurban	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mobility class - Metropolitan	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mobility class - Rural	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mobility class - Suburban	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Employee - vaccination											
Proportion of employees with 2 or more vaccination doses - Delta	-3.8 (-4.1,-3.5)	-2.8 (-4.9,-0.76)	1.7 (1.2,2.1)	-3.6 (-4.2,-3)	7.5 (-3.6,20)	-2.3 (-2.9,-1.8)	-3.3 (-4.1,-2.5)	-1.1 (-2.4,0.22)	-2 (-2.5,-1.6)	-2.3 (-5.4,0.86)	-6 (-11,-0.67)
Proportion of employees with 2 or more vaccination doses - Omicron	-3.3 (-3.6,-3.1)	-7.1 (-9.1,-5.1)	-0.2 (-0.55,0.14)	-4.4 (-4.9,-3.9)	8 (-0.36,17)	-2.3 (-2.8,-1.7)	-6.1 (-6.8,-5.4)	-0.44 (-1.8,0.97)	-2.5 (-2.8,-2.2)	0.62 (-1.9,3.2)	-11 (-14,-7.1)
Employee - age											
Proportion of employees aged 18-29	1 (0.83,1.2)	5.5 (4.3,6.8)	-0.2 (-0.48,0.072)	3.2 (2.8,3.7)	1.7 (-2.7,6.2)	0.19 (-0.29,0.66)	0.76 (0.21,1.3)	1 (-0.16,2.2)	-2.5 (-2.8,-2.1)	4.2 (2,6.5)	1.2 (-1.5,3.9)
Proportion of employees aged 30-44	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Proportion of employees aged 45-59	5.3 (5,5.6)	9.1 (7.8,10)	2.1 (1.8,2.4)	6.9 (6.3,7.4)	13 (8.3,18)	4.2 (3.7,4.6)	-0.47 (-1.1,0.13)	4.5 (3.4,5.6)	2.2 (1.8,2.5)	5.2 (3.1,7.3)	-2.5 (-5.7,0.71)
Proportion of employees aged 60-64	-4.8 (-5.3,-4.2)	1.4 (-1.3,4.2)	-4.3 (-4.8,-3.8)	-6.3 (-7.2,-5.4)	-11 (-16,-6)	-12 (-13,-11)	-3.6 (-4.8,-2.4)	-5.2 (-7.2,-3.1)	-4.1 (-4.7,-3.5)	-0.45 (-3.8,3)	-15 (-19,-10)
Employee - ethnicity											
Proportion of employees with an asian ethnicity	0.19 (0.075,0.3)	2 (0.84,3.1)	-0.056 (-0.18,0.069)	-1.1 (-1.3,-0.89)	3.5 (-1.1,8.2)	-0.51 (-0.72,-0.31)	0.29 (-0.033,0.61)	1.3 (0.71,1.9)	0.59 (0.44,0.74)	1.3 (0.089,2.5)	3.1 (0.85,5.4)
Proportion of employees with a black/african/caribbean ethnicity	1.8 (1.6,2.1)	6.6 (4.6,8.7)	0.38 (0.087,0.68)	3.8 (3.2,4.5)	7.8 (-19,44)	2.1 (1.2,3)	0.22 (-0.32,0.75)	0.98 (-0.32,2.3)	3 (2.7,3.3)	5.5 (2.5,8.6)	-4.3 (-12,4.4)
Proportion of employees with a mixed/multiple/other ethnicity	-9.7 (-10,-9.1)	-33 (-36,-29)	-2.3 (-3,-1.6)	-7.9 (-9.3,-6.4)	-45 (-63,-18)	-8.5 (-10,-6.8)	6.5 (4.6,8.4)	-13 (-15,-9.7)	-11 (-12,-9.8)	-9.4 (-16,-2.9)	78 (52,1.1e+02)
Proportion of employees with a white ethnicity	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Employee - sex											
Proportion of employees who identify as Female	-0.63 (-0.71,-0.54)	-0.69 (-1.1,-0.3)	1.8 (1.7,1.9)	0.5 (0.34,0.65)	-3.1 (-4.2,-2.1)	-1 (-1.2,-0.87)	-1.1 (-1.3,-0.99)	-1.4 (-1.8,-1.1)	1.2 (1.1,1.3)	1.2 (0.6,1.7)	1.8 (0.39,3.2)
Proportion of employees who identify as Male	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Employee - IMD											
Employee IMD quintile	-19 (-20,-18)	-31 (-37,-26)	-6.2 (-7.4,-5)	-15 (-17,-13)	-15 (-33,9.8)	-14 (-16,-13)	-14 (-16,-11)	-21 (-25,-17)	-19 (-20,-17)	-11 (-19,-1.2)	-7.3 (-19,6)
Method of travel to work											
Proportion of employees using bus/metro/tram	0.26 (0.17,0.34)	-0.81 (-1.4,-0.2)	-0.11 (-0.23,0.0039)	1.2 (1,1.4)	-4 (-8.8,1)	0.71 (0.44,0.99)	0.39 (0.14,0.65)	-0.54 (-0.94,-0.13)	1.2 (1.1,1.3)	3.2 (1.9,4.5)	4.5 (1.9,7.2)
Proportion of employees using taxi/vehicle passenger	28 (26,30)	17 (2.2,34)	-4.5 (-6.2,-2.8)	10 (7.4,13)	-41 (-56,-22)	-5.5 (-9.2,-1.7)	14 (9.6,19)	-1.5 (-11,9.1)	9.5 (7.2,12)	15 (-1.7,35)	-27 (-39,-13)
Proportion of employees using other transport	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline	baseline
Proportion of employees using train	5.8 (5.4,6.2)	-2 (-3.7,-0.24)	-2.2 (-2.6,-1.8)	3.6 (3.1,4.2)	-9.3 (-15,-3.3)	3.6 (3.2,4.1)	-4.6 (-5.5,-3.6)	4.4 (2.8,6)	2.9 (2.4,3.4)	-2 (-4.3,0.27)	7.9 (5.8,10)
Proportion of employees using single occupancy	-1.4 (-1.6,-1.3)	-6.1 (-6.8,-5.5)	-3 (-3.2,-2.8)	-2.6 (-2.9,-2.3)	-3.7 (-10,3.1)	-4.5 (-5,-3.9)	-1.6 (-1.9,-1.4)	0.042 (-0.44,0.52)	-3.7 (-3.9,-3.5)	-3.1 (-5.5,-0.71)	3.1 (-0.2,6.6)