Risk factors for SARS-Cov-2 infection at a United Kingdom electricity-generating company: a test-negative design case-control study

Charlotte E Rutter[1*], Martie Van Tongeren[2 3], Tony Fletcher[1], Sarah A Rhodes[4], Yiqun Chen[5], Ian Hall[2 3 6], Nick Warren[5], and Neil Pearce[1]

[1]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK

[2]Division of Population Health, Health Services Research & Primary Care, School of Health Sciences, University of Manchester, Manchester, United Kingdom.

[3]Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom.

[4]Centre for Occupational and Environmental Health, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

[5]Science Division, Health and Safety Executive, Buxton, UK

[6]Public Health, Advice, Guidance and Expertise, UK Health Security Agency, London, United Kingdom

***Corresponding author:**

Department of Medical Statistics, London School of Hygiene and Tropical Medicine,

Keppel St, London

Charlotte.Rutter1@lshtm.ac.uk

Abstract

Objectives

Identify workplace risk factors for SARS-Cov-2 infection, using data collected by a United Kingdom electricity-generating company.

Methods

Using a test-negative design case-control study we estimated the odds ratios (OR) of infection by job category, site, test reason, sex, vaccination status, vulnerability, site outage, and site COVID-19 weekly risk rating, adjusting for age, test date and test type.

Results

From an original 80,077 COVID-19 tests, there were 70,646 included in the final analysis. Most exclusions were due to being visitor tests (5,030) or tests after an individual first tested positive (2,968).

Women were less likely to test positive than men (OR=0.71; 95% confidence interval=0.58-0.86). Test reason was strongly associated with positivity and although not a cause of infection itself, due to differing test regimes by area it was a strong confounder for other variables. Compared to routine tests, tests due to symptoms were highest risk (94.99; 78.29-115.24), followed by close contacts (16.73; 13.80-20.29) and looser work contacts 2.66 (1.99-3.56). After adjustment, we found little difference in risk by job category, but some differences by site with three sites showing substantially lower risks, and one site showing higher risks in the final model.

Conclusions

Infection risk was not associated with job category. Vulnerable individuals were at slightly lower risk, tests during outages were higher risk, vaccination showed no evidence of an effect on testing positive, and site COVID-19 risk rating did not show an ordered trend in positivity rates.

Key messages

- The effects of different job types on risk of COVID-19 infection are unknown and the effect estimates are often confounded by date of test due to differing prevalence rates and testing protocols over time.
- Both date and reason of test are important confounders to be included when estimating odds ratios for COVID-19 infection
- There was little difference in COVID-19 infection risk by job category after adjusting for test reason; however women were less likely to test positive than men

Introduction

In July 2020, government scientific advisers and key funders identified where the United Kingdom (UK) must increase research to respond to near term strategic, policy and operational needs, and ultimately improve resilience against COVID-19 through 2021 and beyond. Six COVID-19 National Core Studies (NCS) have been established to meet these needs, including the NCS project on transmission of the SARS-CoV-2, led by the UK's Health and Safety Executive (HSE). This project is known as "PROTECT": The Partnership for Research into Occupational, Transport and Environmental Covid NCS, which brings together more than 70 researchers from 16 different institutions.

One of the six key themes of the "PROTECT" project is to collect data from outbreak investigations in a range of workplaces to understand SARS-CoV-2 transmission risk factors, potential causes for COVID-19 outbreaks and the effectiveness of a range of measures to control and prevent these outbreaks. In addition to specific outbreak investigations conducted as part of PROTECT, some companies have been identified which succeeded in assembling some detailed data on testing in their workforces including relevant data on outbreaks they have experienced. One of these is a large electricity-generating company. We here report the findings of a test-negative design case-control study conducted using the data collected by this company. The main aim of these analyses was to investigate contextual-level, workplace subgroups and individual-level risk factors for SARS-Cov-2 infections.

Methods

The large electricity-generating company which is the subject of this report, tested staff frequently on site throughout the course of the pandemic. The testing strategy and method varied over time and by facility. During some time-periods, all staff were tested routinely, whereas in other time-periods, most workers were only tested because they had symptoms, or were identified as a contact of a positive case. These practices also varied across sites, so that at any given time, some sites may have been testing routinely, while others were only doing symptomatic and contact testing. Reason for test was collected and categorised into 4 groups: testing due to symptoms (using a lower threshold than government recommendations), testing for close contacts (using government defined criteria), testing for looser work contacts (as per company protocols) and routine testing.

Tests with a missing or inconclusive result were excluded, along with those from visitors (single tests), and any person with missing job type. We also excluded tests that were missing one of the following a priori confounders: age, sex, site, test date or test type.

We included a maximum of one test per day for each person. Where there were multiple tests in a day with different outcomes or reasons (only a small number were identified), we prioritised positive results over negative (as false positives are less common than false negatives), and test reason in order of strength of reason (i.e., symptoms, close contact, loose contact, screening, and then missing). For each person, we used tests only up to and including their first positive result. Thus, the analyses presented here relate to the risks of a first infection, and subsequent infections were not considered.

We used a test-negative design, in which positive tests (cases) were compared with negative tests (controls) during each quarter (3-month period). This approach is intended to control for factors that affect the propensity to be tested at different time points (e.g. changing testing protocols and recognition of symptoms). It also has the advantage of being feasible, since we only had access to the test data, and not to data on individuals who were not tested. It has been widely used for assessing vaccine effectiveness, both for COVID-19(1) and for other infections(2). More recently, it has been used for assessing risk factors for COVID-19 infection(3, 4).

Site risk rating was assessed by the Outbreak Management Team consisting of the company doctors, occupational health advisors and site representatives for each power station, approximately once a week, based on the background prevalence of disease and the number of cases on site. The risk rating from 0 to 5 determined the COVID-19 mitigation requirements (e.g. cleaning, PPE, testing, social distancing). If there was no available risk rating for a particular week then the rating for the closest previous date was chosen. For sites with no risk rating assigned then the average risk rating across all sites was used.

An outage is a statutory shutdown, when a power station is offline, and maintenance can be undertaken. It is often a time with an increased number of external visitors/contractors to the site. We have a binary flag determining whether a test was taken during an outage at the relevant site.

Vulnerability status was determined by increased risk of severe disease or death from COVID-19 due to a pre-existing health condition as determined by the literature and was based on an employee's request for assessment. We have a binary flag for identified vulnerable staff members, who followed different protocols for their own protection (such as home working practices).

Vaccination information was captured for most workers on a voluntary basis. Where we had date of vaccine then we could determine vaccine status at the time of the test, defining vaccine immunity as beginning 10 days after the vaccination date. Partial vaccination was defined as having received one vaccine (10 days or more prior) and full vaccination as two or more. Individuals without vaccination information were assumed unvaccinated as negative information was not captured.

There were different types of tests available at different sites at different times and for different reasons. We categorised the tests as PCR LAMP (polymerase chain reaction loop-mediated isothermal amplification), other PCR (polymerase chain reaction), and LFT (lateral flow test). For PCR LAMP, all positives and 10% of negatives were then confirmed by PCR.

We fitted logistic regression models comparing tests with positive outcomes to those with negative outcomes adjusting for the time-period. There may have been more than one test per person during a time-period but unless the tests were very close together there would not be much dependency. In addition to the a priori confounders of age, sex, date of test, and test type, we considered the other available information detailed above as potential confounders or other explanatory variables of interest. Some of these factors were related, e.g., routine testing was more common during site outages. All potential confounders were included in the final model, provided there were no problems of collinearity or non-convergence of the model. Ordinal variables were tested for linear trend using a likelihood ratio test and included as either categorical or continuous based on the result. The analysis used Stata version 17.

Results

From an original file of 80,077 tests there were 70,878 included in the analysis (Table 1). Most exclusions were due to being visitor tests (5,030) or being tests for an individual after they first tested positive (2,968, of which 433 [14.6%] were also positive).

Table 2 shows the demographic characteristics of the study participants. Almost 90% of the workers tested were men. There was a wide spread of ages from under 20 to over 70 and the median age group was 41-45. The largest proportion of workers were external contractors (53%), followed by engineering (16%) and operations (13%). Jobs were spread in a variety of locations, most at power stations, and there were many more tests per person at power stations 3, 7 and 8 (around 10 per person on average) than at other sites (between 2-5 per person on average).

The number of tests varied hugely by date, reflecting different stages of the pandemic, as well as changes in regulations and protocols of testing and the general prevalence of COVID-19 in the UK. Most tests were in the first half of 2021, but most positive tests were in the first half of 2022 (Figure 1).

The proportions of positive tests that were identified, according to the reasons for testing, are shown in Table 3. Overall, testing those with symptoms identified more than half (54%) of the positive results, whereas testing close contacts picked up 24% and looser contacts 3%. Nevertheless, a significant minority of cases (16%) were identified by routine screening. The remaining 3% had missing test reason.

Table 4 shows the findings for the main risk factors under study. Our "base model" adjusted for test date, age-group and test type (results not shown), as well as the other risk factors shown in the table. There was strong evidence against a linear trend for time-period ($p<0.0001$) and weak evidence for age group ($p=0.05$) so these were both included as categorical.

We found that the reason for testing was a strong confounder; in particular, the odds ratio for "external" workers (i.e. contractors) changed from 0.74 (95% CI=0.61-0.89) to 1.14 (0.89-1.45) after adjusting for the test reason. Also, women showed lower risks than men, after adjustment for test reason (0.70; 0.58-0.86) but not before 1.00 (0.85-1.17). As well as being a strong confounder of job type, reason for testing was a very strong factor itself (Table 4).

Vulnerability, outages, vaccination status and site risk rating were not identified as confounders (on introduction to the model, results not shown) but were included in the final model for completeness. There was strong evidence against a linear trend for site risk rating ($p<0.0001$) so this was included as categorical. There was no evidence against a linear trend for vaccine status ($p=0.92$) so this was included as continuous.

The final model included job type, age, sex, test date, test type, test reason, job site, vaccination status, vulnerability status, outage, and site risk rating. This model showed that there were few differences between job types and likelihood of testing positive after adjusting for all the other included factors. Women were less likely to test positive than men (0.71; 0.58-0.86).

The relationship between site and test positivity showed that power station 8 was higher risk (2.05; 1.52-2.77) than power station 7, the test site with the most tests. All other sites were estimated to be similar or lower risk than power station 7 but with varying magnitudes and levels of evidence (Table 4).

There was a large effect of test reason, with those testing due to symptoms having 94.99 (78.29-115.24) times the odds of those from routine screening, those testing due to a positive close contact had 16.73 (13.80-20.29), and looser contact 2.66 (1.99-3.56) times the odds of those tested in routine screening (Table 4).

There was no evidence of a difference in risk for vaccinated workers (0.97 per vaccination; 0.88-1.06) but vulnerable workers were at lower risk of testing positive than other workers (0.78; 0.63-0.96) and workers testing during an outage were at increased risk (1.35; 1.12-1.63). The site risk rating did not have a linear relationship with an individual worker's risk of testing positive; category 2 was lower risk than category 1 (baseline), but categories 3 and 4 were higher risk (Table 4).

We additionally ran the final model separately for each reason for testing (Supplementary Table S1). This showed markedly different findings according to the reason for testing, e.g., with some sites and job categories showing reduced risks if the testing was for symptoms, but increased risks if it was completed as part of routine screening. However, these findings are difficult to interpret because there is a problem with data sparsity in some time-periods due to the changing testing protocols (Figure 1). Nevertheless, table S1 shows large differences by site for positivity after routine testing.

Discussion

In this test-negative design study, based on data from a United Kingdom electricity-generating company, we estimated the odds ratios for infection by job category, site, reason for testing, vulnerability, sex, reason for testing and the COVID-19 weekly risk rating for each site, adjusting for age and test date. There was little difference in risk by job category, and few differences by site. Women were less likely to be infected than men (OR=0.71). As might be expected, the reason for testing was very strongly associated with test positivity with ORs of 94.99 for those tested because of symptoms, 16.73 for those tested as close contacts of positive cases, and 2.66 for looser contacts, compared to those tested routinely. Reason for testing was also a strong confounder for the other independent variables. Thus, in studies of this type, it is important to collect information on, and adjust for, the reason for testing, as previously suggested by Vandenbroucke et al(3).

It is also notable, that despite these strong associations with symptomatic and contact testing, across the pandemic, 16% of cases were identified by routine testing. Thus, routine testing may have played an important role in identifying a significant minority of cases, and thereby also reducing the spread of infection to contacts.

One limitation of this study is that we have included multiple tests on the same person in each time-period (Supplementary Table S2). If the tests are well-spaced then they should still be independent. However, tests close together on the same person will be more likely to have the same result as each other. One option is to shorten each time-period and remove multiple tests, but if the time-periods are very short then there are problems of sparse data due to the large number of parameters in the model. In addition, although most individuals had up to 4 tests in a time-period, there were some with substantially more. These could bias results towards the null as there is a maximum of one positive result per person and so the multiple negative tests would outweigh the positive one when estimating the effects.

Overall, these findings showed little difference in positivity rates by job category once the analyses were adjusted for test reason. There were some differences by site, with four sites showing substantially lower risks, and one site showing higher risks in the final model. Vulnerable individuals showed slightly lower risks, possibly due to those individuals taking more care. Positivity rates were slightly higher during outages when more people were on site. The site risk rating did not show consistent associations with positivity rates, which may mean that the Covid rules had varying effects at the different levels. Vaccination did not show a protective effect on testing positive which is perhaps surprising.

Research Ethics Approval

The Observational Research Committee of the London School of Hygiene and Tropical Medicine gave ethical approval for this work (ref 28125).

Funding sources

This work was supported by the PROTECT (Partnership for Research in Occupational, Transport and Environmental COVID Transmission) COVID-19 National Core Study on Transmission and Environment and managed by the Health and Safety Executive on behalf of HM Government (grant number 1.11.4.3941).

Conflict of Interest

None declared.

Disclosure Statement

The contents of this paper, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect Health and Safety Executive policy.

References

1. Cerqueira-Silva T, Katikireddi SV, de Araujo Oliveira V, Flores-Ortiz R, Júnior JB, Paixão ES, et al. Vaccine effectiveness of heterologous CoronaVac plus BNT162b2 in Brazil. Nat Med. 2022;28(4):838-43.

2. Vandenbroucke JP, Pearce N. Test-negative designs: differences and commonalities with other case-control studies with 'other patient' controls. Epidemiology. 2019;30:838-44.

3. Vandenbroucke JP, Brickley EB, Pearce N, Vandenbroucke-Grauls C. The Evolving Usefulness of the Test-negative Design in Studying Risk Factors for COVID-19. Epidemiology. 2022;33(2):e7-e8.

4. Vandenbroucke JP, Brickley EB, Vandenbroucke-Grauls CMJE, Pearce N. The test-negative design with additional population controls: a practical approach to rapidly obtain information on the causes of the SARS-CoV-2 epidemic. Epidemiology. 2020:in press.

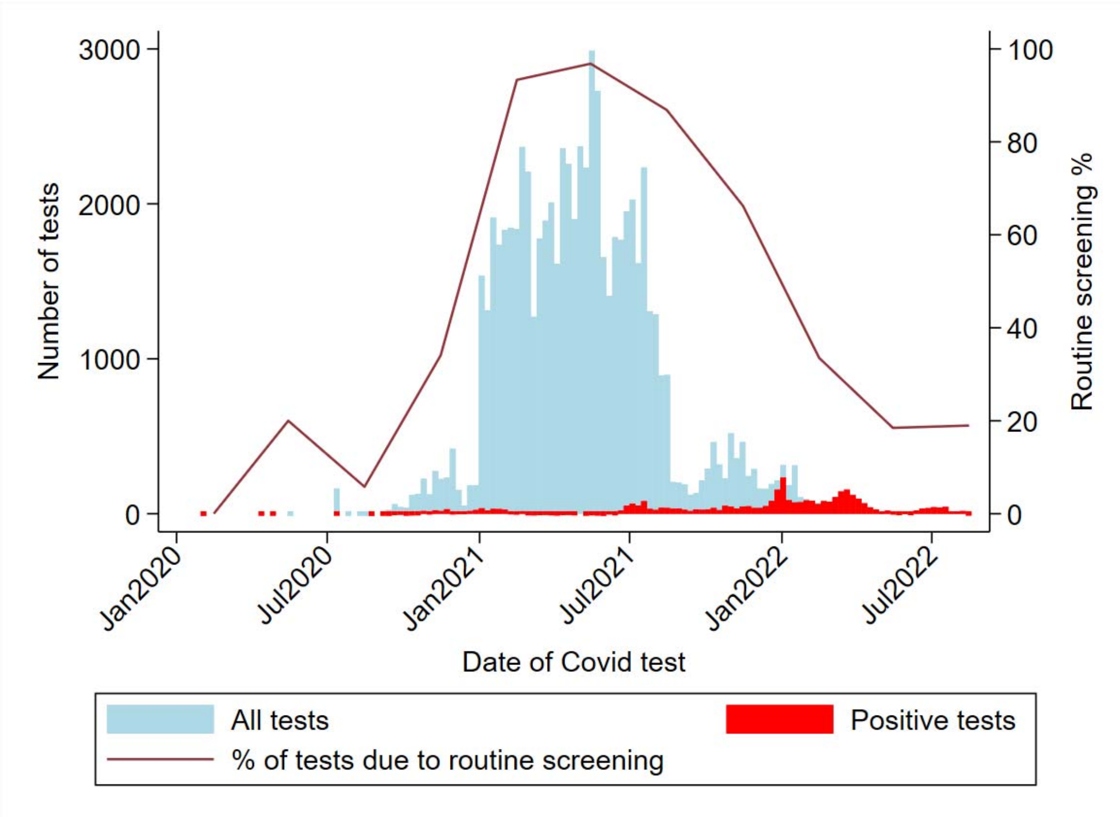

Figure 1 – Number of tests, positive tests, and percentage due to routine screening, by date

Table 1: Test numbers and exclusions

Exclusion reason (in order)	Excluded tests	Remaining tests
Total		80,077
Missing/invalid test outcome	200	79,877
Missing test date	0	79,877
Missing job category	267	79,610
Visitor job category	5,030	74,580
Missing site	0	74,580
Missing sex	0	74,580
Missing age group	1	74,579
Missing test type	601	74,579
Multiple tests in single day[a]	132	74,447
Tests after first testing positive	2,968	70,878

[a] duplicates deleted; 17 people had both negative and positive tests of which one positive test was kept. Of those with matching outcomes, 31 people had different test reasons of which the highest priority one was kept in this order: symptoms, close contact, loose contact, screening, missing reason.

Table 2: Demographic characteristics of analysis sample

Variable	Category	Test level		Individual level (at time of first test)	
		Number	%	Number	%
Total		70,878		10,768	
Sex	Male	63,197	89.2%	9,571	88.9%
	Female	7,681	10.8%	1,197	11.1%
Age group	16-20	1,012	1.4%	189	1.8%
	21-25	4,950	7.0%	760	7.1%
	26-30	7,692	10.9%	1,184	11.0%
	31-35	8,775	12.4%	1,300	12.1%
	36-40	7,875	11.1%	1,237	11.5%
	41-45	7,037	9.9%	1,098	10.2%
	46-50	7,969	11.2%	1,231	11.4%
	51-55	9,727	13.7%	1,480	13.7%
	56-60	9,352	13.2%	1,348	12.5%
	61-65	5,106	7.2%	715	6.6%
	66-70	1,225	1.7%	193	1.8%
	71+	158	0.2%	33	0.3%
Job category	Energy operations	9,067	12.8%	1,420	13.2%
	Engineering	12,752	18.0%	1,716	15.9%
	External contractors	37,064	52.3%	5,710	53.0%
	HSE & security	2,621	3.7%	373	3.5%
	Nuclear & scientific	3,175	4.5%	460	4.3%
	Office-based	4,395	6.2%	783	7.3%
	Project management	1,804	2.6%	306	2.8%
Job site	Head Office	1,273	1.8%	576	5.4%
	Power station 1	2,573	3.6%	1,099	10.2%
	Power station 2	5,758	8.1%	1,164	10.8%
	Power station 3	5,692	8.0%	1,011	9.4%
	Power station 4	15,333	21.6%	1,494	13.9%
	Power station 5	2,926	4.1%	981	9.1%
	Power station 6	2,555	3.6%	865	8.0%
	Power station 7	18,403	26.0%	1,869	17.4%
	Power station 8	15,516	21.9%	1,357	12.6%
	Other	849	1.2%	352	3.3%
Test date	Q1-3 2020[a]	304	0.4%		
	Q4 2020	2,379	3.4%		
	Q1 2021	23,516	33.2%		
	Q2 2021	27,016	38.1%		
	Q3 2021	11,385	16.1%		
	Q4 2021	3,933	5.6%		
	Q1 2022	1,786	2.5%		
	Q2-3 2022[b]	559	0.8%		

Test type	PCR LAMP	51,935	73.3%		
	Other PCR	16,572	23.4%		
	Lateral Flow Test	2,371	3.4%		
Test reason	Symptoms	2,773	3.9%		
	Close contact	3,174	4.5%		
	Looser contact	2,666	3.8%		
	Routine screening	62,033	87.5%		
	Missing[c]	232	0.3%		
Vaccination status[d]	Not vaccinated	57,718	81.4%		
	Partially vaccinated (1)	5,078	7.2%		
	Fully vaccinated (2+)	8,082	11.4%		
Vulnerability status[e]	Not vulnerable	65,262	92.1%	9,819	91.2%
	Vulnerable (Cat1-3)	5,616	7.9%	949	8.8%
Outage[f]	Not during outage	35,580	50.2%		
	During outage	35,298	49.8%		
Site risk rating[g]	0/1[h] (lowest risk)	1,886	2.7%		
	2	15,008	21.2%		
	3	47,208	66.6%		
	4	6,776	9.6%		
	5 (highest risk)	0	0%		

PCR: polymerase chain reaction; LAMP: loop-mediated isothermal amplification;
[a] 3 quarters combined due to low volumes (Q1=1, Q2=5, Q3=298)
[b] 2 quarters combined due to low volumes, especially negative tests (Q2=390, Q3=169)
[c] kept in sample as not all models use test reason
[d] based on vaccination date being populated and dated at least 10 days before test
[e] assumed to be not vulnerable if no vulnerability assessment took place
[f] based on statutory outage dates at the relevant site
[g] based on approximately weekly risk rating given to each site to determine Covid-19 safety protocols. If there was no risk rating for the week, the closest available was used. If site was Other, then average risk rating for the week was used
[h] categories 0 and 1 were combined as there were only 73 in category 0.

Table 3 Proportion of positive tests by test reason in each time-period

Reason	Quarter								
	Q1-3 2020	Q4 2020	Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2022	Q2-3 2022	All
Number of positive tests	25	183	204	120	492	742	1,324	552	3,642
% Symptoms	28%	33%	48%	37%	59%	54%	55%	65%	54%
% Close contact	44%	27%	24%	35%	20%	30%	25%	10%	24%
% Looser contact	0%	11%	10%	5%	2%	3%	2%	0%	3%
% Routine screening	4%	21%	18%	22%	18%	12%	15%	21%	16%
% Missing	24%	8%	0%	2%	1%	2%	3%	3%	3%

Table 4: Odds ratios and 95% confidence intervals of risk of testing positive on covid test

Exposure	Category	Crude job type association (n=70,878)	Base model[a] (n=70,878)	Base model[a] + test reason (n=70,646)	Fully adjusted model[b] (n=70,646)
Job category	Energy operations	0.98 (0.86, 1.12)	1.22 (0.99, 1.51)	0.91 (0.70, 1.19)	0.91 (0.70, 1.20)
	Engineering	0.78 (0.68, 0.88)	0.99 (0.81, 1.21)	0.90 (0.69, 1.16)	0.90 (0.70, 1.17)
	External	0.36 (0.32, 0.40)	0.74 (0.61, 0.89)	1.14 (0.89, 1.45)	1.05 (0.82, 1.36)
	HSE & security	0.92 (0.76, 1.10)	1.16 (0.87, 1.54)	1.10 (0.77, 1.57)	1.12 (0.78, 1.62)
	Nuclear & scientific	0.85 (0.71, 1.01)	1.09 (0.84, 1.43)	0.92 (0.66, 1.29)	0.96 (0.68, 1.35)
	Office-based	1.00 (baseline)	1.00 (baseline)	1.00 (baseline)	1.00 (baseline)
	Project management	0.92 (0.75, 1.13)	1.22 (0.89, 1.68)	1.01 (0.66, 1.53)	1.00 (0.66, 1.51)
Sex	Female		1.00 (0.85, 1.17)	0.70 (0.58, 0.86)	0.71 (0.58, 0.86)
Job site	Head Office		0.74 (0.55, 0.99)	0.67 (0.46, 0.97)	0.70 (0.48, 1.03)
	Power station 1		1.38 (1.10, 1.72)	0.61 (0.47, 0.80)	0.58 (0.43, 0.77)
	Power station 2		0.40 (0.33, 0.48)	0.44 (0.34, 0.56)	0.38 (0.29, 0.49)
	Power station 3		2.54 (2.04, 3.16)	1.12 (0.85, 1.47)	0.90 (0.67, 1.20)
	Power station 4		0.74 (0.61, 0.90)	0.57 (0.45, 0.73)	0.42 (0.32, 0.55)
	Power station 5		1.34 (1.04, 1.71)	0.93 (0.69, 1.27)	0.97 (0.70, 1.33)
	Power station 6		0.39 (0.31, 0.49)	0.26 (0.20, 0.34)	0.22 (0.16, 0.29)
	Power station 7		1.00 (baseline)	1.00 (baseline)	1.00 (baseline)
	Power station 8		4.16 (3.26, 5.31)	2.41 (1.79, 3.24)	2.05 (1.52, 2.77)
	Other		0.93 (0.69, 1.25)	0.34 (0.24, 0.48)	0.31 (0.21, 0.44)
Test reason	Symptoms			85.70 (71.37, 102.91)	94.99 (78.29, 115.24)
	Close contact			15.32 (12.75, 18.40)	16.73 (13.80, 20.29)
	Looser contact			2.51 (1.89, 3.33)	2.66 (1.99, 3.56)
	Routine screening			1.00 (baseline)	1.00 (baseline)
Vaccination status	Per vaccination[c]				0.97 (0.88, 1.06)
Vulnerability status	Vulnerable				0.78 (0.63, 0.96)
Outage	During outage				1.35 (1.12, 1.63)
Site risk rating	0/1 (lowest risk)				1.00 (baseline)
	2				0.64 (0.50, 0.82)
	3				1.30 (1.00, 1.69)
	4				1.60 (1.11, 2.31)

[a] adjusted for job category, sex, and site in the table, and also test date, test type and age group (results not shown); [b] adjusted for all variables in the table plus test date, test type and age group (results not shown); [c] up to fully vaccinated (2 or more vaccinations)