Methods

In this study, we aim to investigate weather "air Cleaning techologies" (ACT) had an impact on illness-related absences in primary schools. Thirty one primary schools in the Bradford area were recruited and fitted with air quality monitors and were assigned to one of three study arms: HEPA-ACT, UV-ACT and a control arm with no ACTs. Regulatory complications prevented the activation of the UV-ACT devices, so this study we will examine the effect of the HEPA-ACT only.

The study period covers September 2021, when in-person compulsory education resumed following the covid pandemic, until April 2022, at which point public health guidance changed significantly. Our primary unit of observation is a school, and our outcome of interest is a rate of the number of school sessions missed due to illness per one hundred sessions attended in-school. The principal independent variable of interest was the presence of classroom HEPA filters. We also included measures of indoor air quality, mean CO2, and the rate of covid infections in the overall populations, positive tests, as confounding independent variables. Further information on the data can be found in the supplimentary materials.

To evaluate the effect of HEPA filtration, we used a simple additive model of the overall illness absence rate as a gamma random variable, with parameters estimated by a regression on HEPA filtration, mean CO^2 concentration and mean weekly positive covid tests in the local population aggregated over the study period:

$$illness.rate_i \sim Gamma(\mu_i, \phi)$$
$$log(\mu_i) = \beta_0 + \mathbf{x}_\mathrm{i}(\beta_\mathrm{HEPA} + \beta_{\mathrm{CO}^2} + \beta_{+\mathrm{ve.tests}})$$

Results

Figure 1 shows a comparison of the distribution of monthly illness rates between schools in the HEPA and control arms. The data show illnesses following a familiar trajectory, rising to a peak over the autumn term towards the winter break and declining into the spring term. Rates of illness in schools with HEPA filters appear to be lower on average, and this difference appears to be greater in the peak winter months. An extreme outlier can be seen in one of the HEPA schools in September - this corresponds with a severe outbreak of COVID infections in the local area of the affected school which resulted in a dramatic increase in isolation and illness absences in the affected school over a 3 week period.

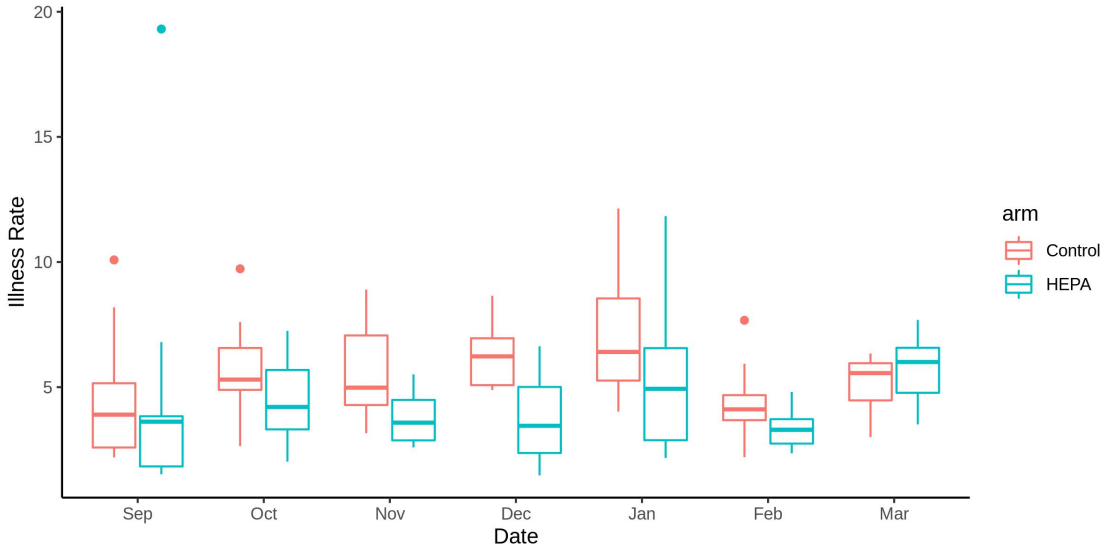

Figure 1: Illness rates for each arm of the study by month. Outliers are shown as dots above the boxplots. An extreme outlier can be seen in the HEPA arm in September as the schools returned for the first time following COVID closures

Our regression analysis of the marginal effect of HEPA filters over the study period did suggest a reduction in illnesses (marginal effect of HEPA -0.977 illness sessions per 100 attended sessions 95%CI [-2.22, 0.268]), but this effect was not statistically significant (p value 0.124). The estimated effects of mean CO^2 and positive covid tests in the population were both extremely small and not statistically significant (see supplementary materials). Further analysis suggested that the estimated effect of HEPA filters was highly influenced by an individual school. Figure 2 shows the Cook's distance measures for the total effect of HEPA filters for each of the schools: the school in which there was an extreme outbreak in September has an unusually high Cook's distance value, and thus has an exaggerated influence on the estimated effect of HEPA filters.

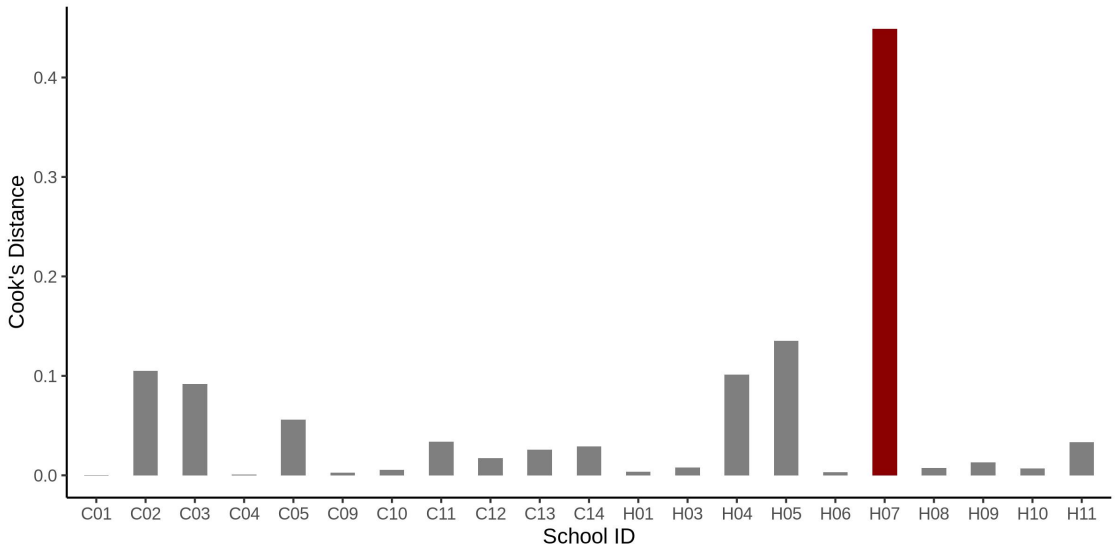

Figure 2: Cooks distances for each observation as an estimate of their influence on the estimated effect of HEPA filtration. School IDs prefixed with a C are control schools, and with an H are HEPA schools. Figures are calculated using the total effect of the HEPA intervention, using a regression on presence of HEPA filters only. School H07 highlighted in red is the same school that experienced the extreme outbreak seen in September 2022 in Figure 1

We defined a threshold above which illness rates were considered extreme outliers, and removed these observations from the weekly data; subsequent regression estimates of the effect of HEPA filters without without these outliers showed a greater reduction in illnesses as our previous analyses and the effect was now statistically significant (p value 0.0212). In the absence of extreme outbreaks we estimated a marginal -1.23 (95% CI [-2.20, -0.261]) reduction in illness sessions per 100 attended sessions with the use of HEPA filters, a 21.8% decrease in illness absences compared to the mean illness rates in control schools (Figure 3). Estimates of the effects of the other co-variates remained small and not statistically significant (see supplementary materials).

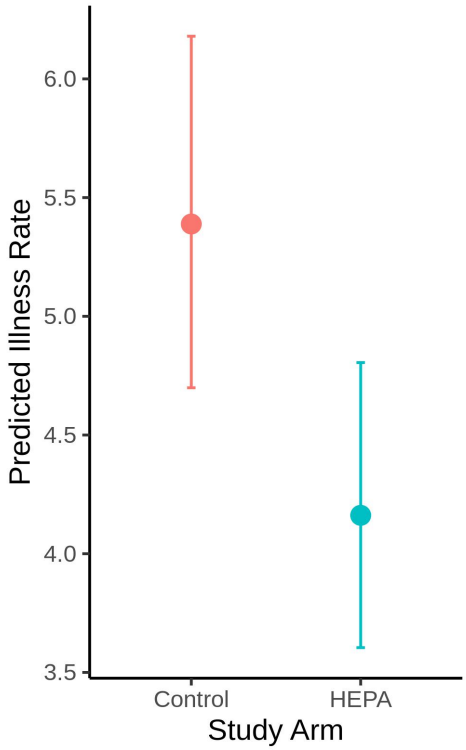

Figure 3: Estimates of the average marginal effect of the presence of HEPA filters from our regression analyses where outliers were removed from the data. Error bars represent a 95% confidence interval

Supplimentary Materials

Data Preparation

Attendance Reports

Every school has a statutory responsibility to collecte and report attendance records in accordance with DfE guidance (REFERENCE). Attendance data for each school in the study were provided by the Information Management Team at Bradford Council covering the full study period. The data provided comprise anonymised attendance "strings" (lines of text) represented as individual characters that signify one of 30+ attendance types, per student for each of 14 weekly am/pm sessions Monday-Sunday. These codes were cleaned and aggregated into weekly counts of the various attendance types. Weeks in which schools were closed were removed from the data, as they contain no information about attendance, illness or otherwise. A minority of codes were blank or did not correspond to a recognised attendance type. Pending potential future data quality improvement efforts, school weeks with a 1% or greater proportion of "junk" sessions were removed from the dataset - 2.4% of the aggregated data after removing school closures.

The aggregated attendance records were then divided into one of three categories: codes signifying an illness-related absence, codes signifying in-school attendance, and remaining codes that signify absence from school for reasons other than illness (TABLE REF). We then defined our response/outcome variable as a rate of the illness-related absence sessions per one hundred in-school attended sessions.

	Code	Description
In-school Attendance Codes:	/	AM: Present in school
	\	PM: Present in school
	L	Late arrival before the register has closed
	U	Arrived in school after registration closed
Illness Attendance Codes:	**I**	Illness (not medical or dental appointments)
	I01*	Illness - "non-coronavirus (COVID-19) related illness or sickness"
	I02*	Illness Confirmed case of coronavirus (COVID-19)
	X02*	Pupil self-isolating with coronavirus (COVID-19) symptoms

Figure 4: DfE attendance codes that signify a pupil's attendance in the school building or absence from school because of illness. The remaining attendance codes are not considered in this study.

***** Added to the pre-2020 attendance codes following the DfE's updates to the attendance guidelines following the Covid-19 pandemic

Air Quality Data

An important mediator of the transmission of airborne illnesses is the ventilation of the classroom environment. Ventilation rates have a potential to confound the effect of the HEPA intervention in this study, as better ventillated classrooms have a lower potential for illness to be communicated from person to person. Each of the participating Class-Act schools are fitted with air quality sensors, collecting timestamped recordings of CO^2 concentration (along with other IAQ measures) each minute. These data were collected for the study period, and cleaned to remove measurements above unreasonably high thresholds values and dates with greater than 25% missing/dropped recordings. The cleaned data were then aggregated to calculate a weekly mean CO^2 concentration over the period covered in this study for each of the participating HEPA/control schools.

Covid Rates

Rates of Covid-19 in the general population also have the potential to confound the effect of the HEPA intervention. The United Kindom government provide statistics on Covid-19 related metrics as a publicly available resource (REFERENCE). Data are available at varying granularities, from National at the highest level to Middle Layer Super Output Areas (MSOA) at the lowest level. For the purposes of this study, "new cases by specimen date" was selected as the metric to be used as a potential covariate. This metric is available at different levels of specificity. At the lower tier local authority level, daily 7-day rolling rates are available broken down into demographic age groups. At the MSOA-level weekly data are available for the total population for the 7-day period ending each Saturday. It was unclear which level of specificity would be more valuable in describing illness outcomes in schools, age or geographic location and, as such, both the local authority and MSOA levels of the "new cases by specimen date" were selected as covariates for analysis (REFERENCE DEFINITIONS). To ensure parity between the data recorded weekly, rolling rates were taken on the Saturday of each week for the data available daily. The MSOA level statistics proved to better explain illness rates in the schools, so this measure was used in the final analysis.

Variable Name	newCasesBySpecimenDateAgeDemographics
Lowest Geographic Level	Lower Tier Local Authority
Frequency	Daily
Description	New cases by specimen date age demographics - Age breakdown of new cases. Data are shown by the date the sample was taken from the person being tested. Data are shown for rate per 100,000 people of the number of new cases in the rolling 7-day period ending on the dates shown .

Variable Name	newCasesBySpecimenDateRollingRate
Lowest Geographic Level	Medium Super Output Area (MSOA)
Frequency	Weekly
Description	New cases rolling rate by specimen date - Rate per 100,000 people of the number of new cases in the rolling 7-day period ending on the dates shown. Data are shown by the date the sample was taken from the person being tested.

Figure 5: Details of the UK Health Security Agency (UKHSA) Covid-19 statistics used in this study as described by the documentation at coronavirus.data.gov.uk

Aggregation

	Control (N=285)	**HEPA** (N=260)	**Overall** (N=545)
Illness Ratio:			
Mean (SD)	0.0569 (0.0304)	0.0459 (0.0392)	0.0517 (0.0353)
Median [Min, Max]	0.0507 [0.00332, 0.177]	0.0375 [0.00498, 0.398]	0.0436 [0.00332, 0.398]
Missing	4 (1.4%)	6 (2.3%)	10 (1.8%)
Mean CO^2 Concentration:			
Mean (SD)	1220 (277)	1270 (288)	1240 (284)
Median [Min, Max]	1230 [609, 2110]	1270 [651, 1880]	1250 [609, 2110]
Missing	31 (10.9%)	28 (10.8%)	59 (10.8%)
Covid Rate - MSOA:			
Mean (SD)	522 (531)	516 (485)	519 (509)
Median [Min, Max]	319 [72.3, 3100]	344 [49.7, 2460]	332 [49.7, 3100]
Missing	1 (0.4%)	1 (0.4%)	2 (0.4%)
Covid Rate - Age 4-14:			
Mean (SD)	712 (617)	712 (616)	712 (616)
Median [Min, Max]	417 [80.4, 2360]	432 [80.4, 2360]	417 [80.4, 2360]

Table 1: Table 1 statistics of the weekly observations for each variable prior to aggregation

OUTLIER REMOVAL

Model Results

Full Dataset

```
Call:
glm(formula = illness_rate ~ 1 + arm + mean_co2 + mean_positive_tests,
    family = Gamma(link = "log"), data = agg_data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.54665  -0.18318  -0.01851   0.11960   0.59955

Coefficients:
```

```
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.4063099  0.4251992   3.307  0.00416 **
armHEPA             -0.1960541  0.1262701  -1.553  0.13892
mean_co2             0.0001574  0.0002908   0.541  0.59520
mean_positive_tests  0.0002012  0.0003749   0.537  0.59835
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 0.08138246)

    Null deviance: 1.6095  on 20  degrees of freedom
Residual deviance: 1.3639  on 17  degrees of freedom
AIC: 78.603

Number of Fisher Scoring iterations: 5
```

	Contrast	Estimate	Std Error	t Value	p Value
Study Arm	HEPA - Control	-0.977 (-2.22, 0.268)	0.635	-1.54	0.124
Mean CO^2	+100	0.0791 (-0.249, 0.408)	0.168	0.472	0.637
Positive Covid Tests	+100	0.101 (-0.289, 0.492)	0.199	0.507	0.612

Table 2: Marginal effect estimates for each of the input parameters from the model using the full dataset. P values suggest none of the effects are statistically significant

Data With Outliers Removed

```
Call:
glm(formula = illness_rate ~ 1 + arm + mean_co2 + mean_positive_tests,
    family = Gamma(link = "log"), data = agg_data_wo_outliers)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.44980  -0.07781  -0.02459   0.14408   0.35029

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.431e+00  3.419e-01   4.185 0.000622 ***
armHEPA             -2.584e-01  1.021e-01  -2.530 0.021584 *
mean_co2             2.201e-04  2.355e-04   0.935 0.363124
mean_positive_tests -3.896e-05  2.983e-04  -0.131 0.897603
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 0.05286915)

    Null deviance: 1.32849  on 20  degrees of freedom
Residual deviance: 0.96697  on 17  degrees of freedom
AIC: 69.723

Number of Fisher Scoring iterations: 5
```

	Contrast	Estimate	Std Error	t Value	p Value
Study Arm	HEPA - Control	-1.23 (-2.2, -0.261)	0.494	-2.49	0.0129
Mean CO^2	+100	0.106 (-0.15, 0.362)	0.131	0.809	0.419
Positive Covid Tests	+100	-0.0187 (-0.314, 0.277)	0.151	-0.124	0.901

Table 3: Marginal effect estimates for each of the input parameters from the model estimated from the data with outliers removed. The effect of HEPA filtration is statistically significant. Mean CO^2 and positive covid tests are not statistically significant