Open access | Original research

BMJ Open

Association between community-based self-reported COVID-19 symptoms and social deprivation explored using symptom tracker apps: a repeated cross-sectional study in Northern Ireland

Jennifer M McKinley ,[1] David Cutting ,[2] Neil Anderson,[2] Conor Graham ,[1] Brian Johnston ,[1] Ute Mueller ,[3] Peter M Atkinson ,[4] Hugo Van Woerden ,[5,6] Declan T Bradley ,[5,7] Frank Kee[5,7]

To cite: McKinley JM, Cutting D, Anderson N, *et al*. Association between community-based self-reported COVID-19 symptoms and social deprivation explored using symptom tracker apps: a repeated cross-sectional study in Northern Ireland. *BMJ Open* 2021;**11**:e048333. doi:10.1136/bmjopen-2020-048333

► Prepublication history and additional online supplemental material for this paper are available online. To view these files, please visit the journal online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-048333).

Received 22 December 2020
Accepted 28 May 2021

Check for updates

© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Prof Jennifer M McKinley; j.mckinley@qub.ac.uk

ABSTRACT

Objectives The aim of the study was to investigate the spatial and temporal relationships between the prevalence of COVID-19 symptoms in the community-level and area-level social deprivation.

Design Spatial mapping, generalised linear models, using time as a factor and spatial-lag models were used to explore the relationship between self-reported COVID-19 symptom prevalence as recorded through two smartphone symptom tracker apps and a range of socioeconomic factors using a repeated cross-sectional study design.

Setting In the community in Northern Ireland, UK. The analysis period included the earliest stages of non-pharmaceutical interventions and societal restrictions or 'lockdown' in 2020.

Participants Users of two smartphone symptom tracker apps recording self-reported health information who recorded their location as Northern Ireland, UK.

Primary outcome measures Population standardised self-reported COVID-19 symptoms and correlation between population standardised self-reported COVID-19 symptoms and area-level characteristics from measures of multiple deprivation including employment levels and population housing density, derived as the mean number of residents per household for each census super output area.

Results Higher self-reported prevalence of COVID-19 symptoms was associated with the most deprived areas ($p<0.001$) and with those areas with the lowest employment levels ($p<0.001$). Higher rates of self-reported COVID-19 symptoms within the age groups, 18–24 and 25–34 years were found within the most deprived areas during the earliest stages of non-pharmaceutical interventions and societal restrictions ('lockdown').

Conclusions Through spatial regression of self-reporting COVID-19 smartphone data in the community, this research shows how a lens of social deprivation can deepen our understanding of COVID-19 transmission and prevention. Our findings indicate that social inequality, as measured by area-level deprivation, is associated with disparities in potential COVID-19 infection, with higher prevalence of self-reported COVID-19 symptoms in urban areas associated with area-level social deprivation, housing density and age.

Strengths and limitations of this study

- ► The geographical spread of the self-reporting participants using the smartphone apps was investigated through spatial mapping and regression using time as a factor.
- ► The use of two apps from different smartphone app providers enabled a broad sampling of the general population using a repeated cross-sectional study design.
- ► The predicted variable in the study is the reporting of COVID-19 symptoms rather than true disease prevalence and therefore caution must be exercised in interpreting the results.
- ► Nevertheless, the results may inform the search for effective interventions to reduce health inequalities and improve prevention of COVID-19 in the population.

INTRODUCTION

Measuring and managing transmission of the novel SARS-CoV-2 virus has presented public health authorities and policy-makers with considerable challenges during the evolution of the COVID-19 pandemic.[1] The variety of approaches adopted by different countries for monitoring the spread of the virus, included spatiotemporal epidemiology, contagion risk models and monitoring platforms,[2–5] to inform their policy responses. Measurement of the number of cases is key to monitoring transmission, risk assessment and evaluating the effectiveness of non-pharmaceutical societal interventions. National agencies record data on numbers of COVID-19 positive tests, hospital admissions and deaths, but these are biased towards the higher parts of the epidemiological pyramid,[6] representing mainly people with more severe disease and timely access to testing. The challenge during the COVID-19 pandemic has been recording those in the community with mild symptoms who may not seek care or be able to access testing. Moreover, the number of infected people in the community depends on individual and social behaviours and these data have been more difficult to record. The introduction of COVID-19 symptom trackers as free smartphone apps (launched in UK 24 March 2020 and US 29 March 2020) provided a way to track in real time how the virus might be transmitting by recording self-reported health information from both asymptomatic and symptomatic individuals on a daily basis.[7–10] At this stage in the pandemic, during the earliest stages of non-pharmaceutical interventions, viral or other positive testing methods were not widely available.[11] However, the COVID-19 symptom trackers provided a way to record self-reported health information from both non-symptomatic and symptomatic individuals in the community.

The importance of the link between health and place is widely recognised.[12] Health inequalities are defined as differences in health across the population, and between different groups within society.[13] An interplay of factors at multiple levels can influence health inequalities, including the physical and socioeconomic environment.[14–17] Limitations in data sampling, data collection and analysis techniques have constrained our understanding of the causes of these disparities.[18] This has hindered the opportunity to provide evidence for effective interventions to reduce these disparities and improve overall health outcomes. Health inequalities have been documented between population groups across socioeconomic status and deprivation, vulnerable groups of society or 'inclusion health' groups and geography.[13] The main driver for these differences is contact networks which arise as a function of social behaviour (culture) and urban and rural geographies. It is now recognised that the COVID-19 crisis has disproportionately affected certain at-risk communities, based on their previous health, socioeconomic position and ethnic characteristics.[19–25] While most of the clinical research has reported on people experiencing severe illness, in this research we investigate the spatial and temporal relationships between the prevalence of COVID-19 symptoms in the community and area-level social deprivation using a repeated cross-sectional study design.

METHODS

The current study concentrates on the reporting period 24 March 2020–22 June 2020 at the earliest stages of non-pharmaceutical interventions and societal restrictions ('lockdown'), when viral or other positive testing methods were not widely available.[11] A repeated cross-sectional study design using self-reported COVID-19 symptoms smartphone apps provided a way to track the spatial and temporal spread of the virus through Northern Ireland (NI) by self-reported health information from both asymptomatic and symptomatic individuals.

In the UK, administrations in England, Scotland, Wales and NI have responsibility for public health functions, including most aspects of responding to the COVID-19 pandemic. Our study setting is NI, one of the devolved UK nations, with an estimated mid-year population of 1 893 700 (30 June 2019).[26] Two major symptom tracking apps were available and used in NI. The UK COVID-19 symptom tracker was developed by King's College London (KCL) and the health science company ZOE (https://COVID-19.joinzoe.com/) and is available to download throughout the UK.[10] The NI Health and Social Care (HSC) service launched its own symptom tracker app, COVIDCare NI (formerly known as 'COVID-19 NI'), on 6[20]April 2020. The COVIDCare NI symptom checker app, developed primarily as part of a triage system, provided advice for users on whether they should self-isolate and/or seek medical assistance. The UK KCL ZOE symptom tracker app provided data for NI for the current study for the period (24 March 2020–22 June 2020) whereas the HSC NI Symptom checker feature (COVIDCare NI) provided data for the reporting period 6 April 2020–22 June 2020. Smartphone ownership does not vary significantly by urban or rural location in NI and shows a strong alignment with UK prevalence.[27] In 2019, 76% of adults in the UK reported smartphone ownership.[28]

Data from both smartphone symptom tracking apps were generated on a series of 7 and 14 day periods, known as sliding windows. Each period contained: (1) total individual active users who have used the COVID-19 symptom checking/recording features and (2) total individual users recording an assessment, with symptoms meeting the classic (new continuous cough or high temperature) or refined (new continuous cough or high temperature or anosmia) Public Health England (PHE) COVID-19 case definitions.[29] There are some differences between the two symptom tracker apps especially with the 'new' PHE definition which included anosmia. These are:1) The symptom of anosmia was included in the KCL ZOE symptom tracker app from the start but was only included later in the presumptive positive definition. In this research study, we, therefore, refer to classic (new continuous cough or high temperature) or refined symptoms (new continuous cough or high temperature or anosmia) as defined by PHE; (2) The COVIDcare NI app initially included anosmia as a symptom which could only be reported if one of the traditional symptoms was also present. The definition was changed at the same time as the KCL ZOE symptom tracker app refined the symptoms in line with PHE guidelines.[29] In summary both symptom tracker apps provide a cross sectional study but may not be sampling the same repeated cross section of the population of NI. Therefore, the symptom tracker app data sets were not combined in the analysis but instead a comparative analysis was provided.

Both tracker apps require recording a location. As the period of the analysis coincided with the first societal

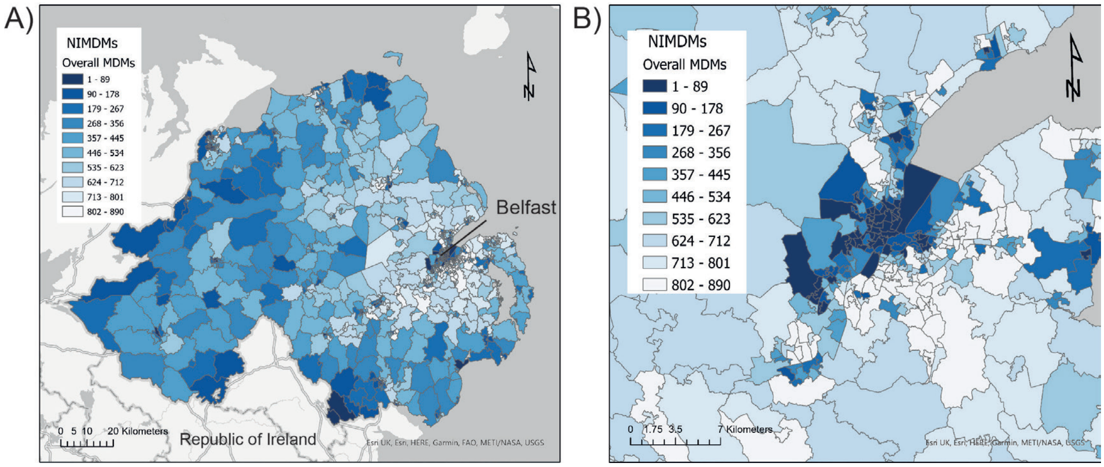

Figure 1 Northern Ireland Multiple Deprivation Measures 2017 (NIMDM) provided by the Northern Ireland Statistics and Research Agency[30] including information on overall social deprivation MDMs ranking (A) for NI and (B) for Belfast urban area. Low ranking indicates highest deprivation. MDMs, Multiple Deprivation Measures; NI, Northern Ireland.

restrictions ('lockdown') in NI, it is reasonable to assume that for most people this would have been their home location. The KCL ZOE symptom tracker app report is linked to one distinct individual record. The COVIDcare NI app records events without specific individuals. The authors introduced a pseudo-individual marker based on a combination of individual factors (handset used, age, gender) which proved to be very effective in providing distinctive records. Therefore, in the current study we assume that for both apps a COVID-symptomatic individual has been included once. For this research study, both symptom tracker app datasets were analysed at super output area (SOA) level. The KCL ZOE tracker app generates data geocoded to SOAs, while in the case of COVIDCare NI, data were converted from postal code to SOAs by the authors. Data containing invalid postcodes or postcodes outside of NI were removed during this postprocessing. There are 890 SOA administrative areas across NI.[30] When the numbers of users or those reporting symptoms (from either app) were too small in any SOA (n≤5) these small cell counts were suppressed to avoid disclosure risk. By 'reporting symptoms' we mean that, on any given date, symptoms would have satisfied the PHE case definition.[29]

Area-level deprivation was characterised using the orthern Ireland Multiple Deprivation Measures 2017 (NIMDM) provided by the Northern Ireland Statistics and Research Agency (NISRA; figure 1).[30] The NIMDMs are derived from the 2011 census and were made available by the NISRA in 2017. The 2011 census is currently the most comprehensive population census for NI. Results for the next census are not yet available[31] as it took place in March 2021. The NIMDMs provide information on seven individual domains of deprivation and an overall score

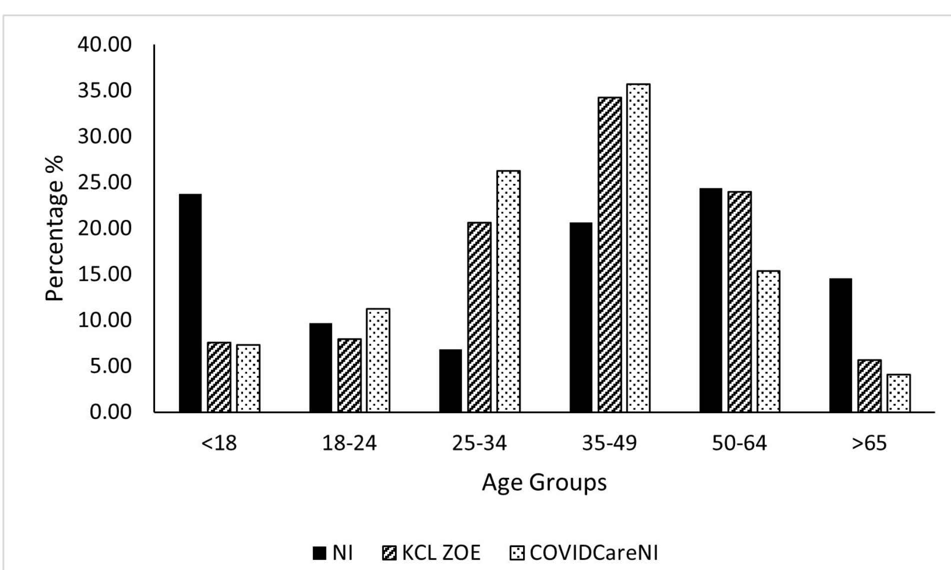

Figure 2 Smartphone symptom tracker app user demographic profile compared with the population profile of NI. Self-reporting COVID-19 symptom data provided by the KCl ZOE symptom tracker app data for NI (reporting period 24 March 2020–22 June 2020) and COVIDCare NI symptom checker feature, (reporting period 6 April–22 June 2020). NI, Northern Ireland.

for relative social deprivation, comparable to the Index of Multiple Deprivation in England.[32 33] The ranking scale is from 1 (most deprived) to 890 (least deprived). Population household density was used as a further explanatory variable to investigate the relationship with self-reported COVID-19[31] and was derived as the number of residents divided by number of households for each SOA.

Regression analysis

The reporting period 6 April 2020–30 May 2020 was used for regression analysis, to provide a repeated cross-sectional study. KCL ZOE symptom tracker app data for NI with revised PHE case definitions and COVIDCare NI, based on a repeated 14 day sliding window (resulting in a 1-week overlap of data), attributed to the last day of the period, were used for regression analysis. The 7-day sliding window data were not used due to low number issues for some SOAs. For both COVID-19 self-reporting symptom mobile platforms, the data were analysed in the form of:

- Rates calculated as the proportion of active users reporting symptoms for each SOA that occurred in the defined periods of time, standardised according to the population of each SOA. This allowed comparison of a repeated cross-sectional study of self-reported prevalence of COVID-19 in terms of active app users reporting PHE case definition symptoms.
- Age-standardised rates based on the 2011 Census population of NI.[26] The age brackets used based on 2011 Census population data (as the most comprehensive age band data available) comprised <18, 18–24, 25–34, 35–49, 50–65 and >65 years.

Generalised regression models (with log link) were fitted between the dependent variable 'population standardised self-reported COVID-19 symptoms' and time as a factor. The independent variables included area-level deprivation indices using overall Multiple Deprivation Measures (MDM), individual deprivation domains (adjusting p values for multiple comparisons using the Bonferroni correction) and population household density. All regression analysis was conducted using glm R package and R V.4.0.0. To account for spatial autocorrelation the Moran's I statistic and a spatial lag model, using spatialreg R package, were used to test the residuals computed from the regression models.[34 35] Where the Moran's I for the residuals was found to be significantly different from random, the generalised linear model regression results were compared with a spatial lag model and the model fit compared using an Akaike information criterion.

Patient and public involvement

No patients were involved.

RESULTS

The smartphone symptom tracker apps user demographic profile is most comparable with the population profile of NI for the age groups 18–24 years and 50–64 years and shows a higher percentage of users fall within the age groups 25–34 and 35–49 years relative to the other age groups (figure 2). The self-reporting COVID-19 symptom data represent a time series of the prevalence of self-reported symptoms. The earlier release date of the UK KCL ZOE symptom tracker app, compared with the COVIDCare NI app, allowed analysis of COVID-19 self-reporting symptom data at the earliest stages of non-pharmaceutical interventions and societal restrictions ('lockdown') in NI (14 day window data from 30 March 2020). An increase in active users of the KCL ZOE tracker App reporting COVID-19 symptoms was observed between 30 March 2020 and 6 April 2020, followed by a sharp decrease after 6 April 2020 (figure 3A). The COVIDCare NI app shows a decrease in active users reporting COVID-19 symptoms from the start of reporting period 20 April 2020 until 22 June 2020. However, there was an

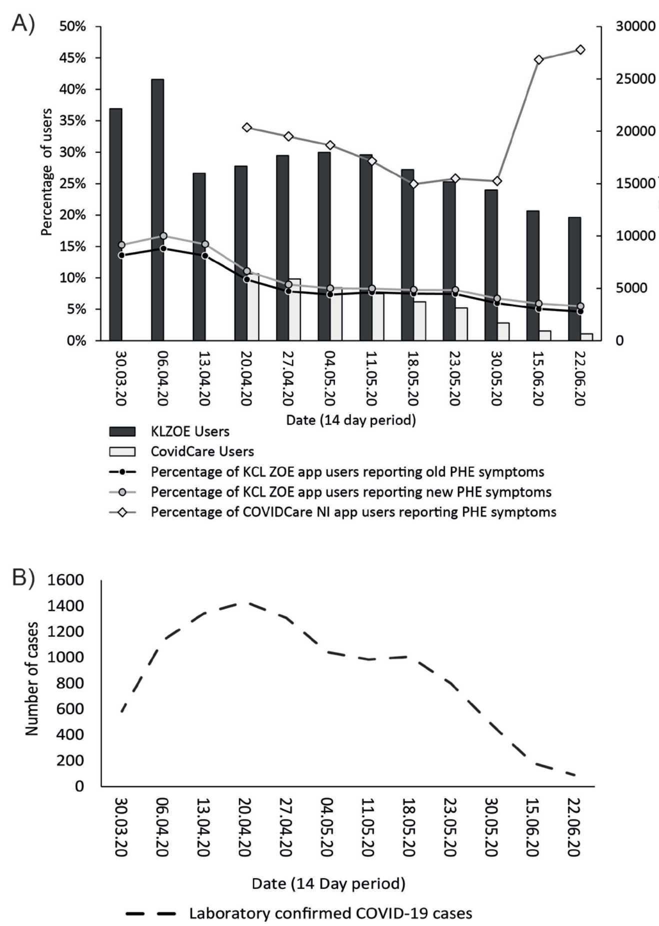

Figure 3 (A) comparison of percentage of users self-reporting COVID-19 symptom data (using data from table 1) as provided by the KCl ZOE symptom tracker app data for NI (reporting period 24 March 2020–22 June 2020), COVIDCare NI symptom checker feature, (reporting period 6 April 2020–22 June 2020) and (B) laboratory confirmed COVID-19 cases. The dates correspond to the end date of 14-day symptom reporting sliding window. (B) laboratory confirmed COVID-19 cases based on published data HSC NI public health agency reports.[11] HSC, health and social care; NI, Northern Ireland; PHE, Public Health England.

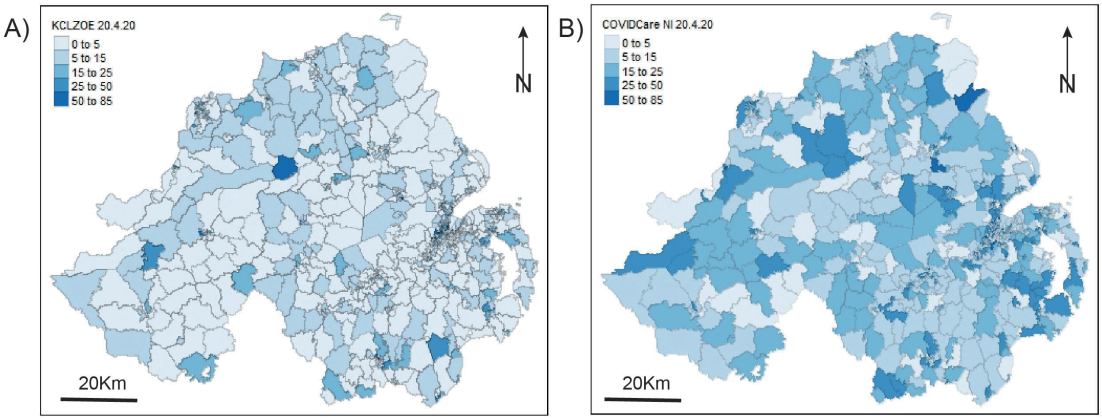

Figure 4 Maps of COVID-19 symptom data for the reporting period ending 20 April 2020, provided by two sources: (A) KCl ZOE symptom tracker app data for new PHE symptoms for Northern Ireland (reported symptoms in 592 SOAs) and (B) COVIDCare NI symptom checker feature (reported symptoms in 758 SOAs). The date 20 April corresponds to the end date of a 14-day symptom reporting sliding window. Self-reported prevalence rates are standardised for 100 000 population. NI, Northern Ireland; PHE, Public Health England; SOA, super output area.

increase in the percentage of COVIDCare NI app users reporting PHE symptoms for the time period 30 May 2020 to 22 June 2020. Although the overall number of users for both apps has decreased by this time, the increase percentage of COVIDCare NI app users reporting symptoms may reflect the NI Public Health Agency messaging to use the COVIDCare NI app as a triage system to provide advice for symptoms on whether they should self-isolate and/or seek medical assistance. Both apps show an overall decrease in reported COVID-19 symptoms over time which mirrors the reported peak and subsequent decline in laboratory confirmed COVID-19 cases (figure 3B).[11] Although the overall rate (per 100 000 population) of self-reported symptoms is comparable, the geographical coverage varies across the time periods (figure 4A,B). The spatial maps (figure 4A,B) indicate that KCL ZOE symptom tracker app users were located more frequently in high population areas including Belfast, whereas coverage for the COVIDCare NI app was more even across NI.

Regression analysis revealed a statistically significant negative correlation between active users of both mobile platforms reporting symptoms and area-level deprivation (MDM; p<0.001; table 1). These findings indicate that throughout the reporting period, from initial lockdown when restrictions were most stringent, the most deprived SOAs (lowest social deprivation rankings) were associated with higher population standardised prevalence rates of self-reported COVID-19 symptoms. A statistically significant negative correlation was found between users reporting COVID-19 symptoms and the area-level deprivation measures of employment (p<0.001) and living environment (p=0.01) using data from both mobile platforms, (online supplemental table 1).

Using the mean number of residents per household for each SOA as a proxy for population housing density and time as a factor, a statistically significant negative correlation, was found between prevalence rates of self-reported COVID-19 symptoms and mean number of residents per household for both tracker apps (online supplemental table 2). The findings indicate that higher self-reported prevalence rates are associated with SOAs that have a lower mean number of residents per household. This seems counterintuitive with the expectation that higher density housing would increase the risk of transmission. As urban areas have a greater proportion of higher density housing, an analysis was carried out for the Belfast urban area, the capital city of NI, UK. The Belfast urban area comprises 150 SOAs and a population of 287, 535 (as defined by the local government districts identifier[31]). For Belfast, a statistically significant negative correlation was found between self-reported prevalence rates and social deprivation (MDM; p<0.001) for both Symptom Tracker apps (online supplemental table 3). The findings for the urban area of Belfast are consistent with that for NI and indicate that during lockdown restrictions, the most deprived SOAs were associated with higher population standardised prevalence rates of self-reported COVID-19 symptoms. However, the relationship between population standardised prevalence rates of self-reported COVID-19 symptoms for Belfast in relation to housing density is quite different from that observed for overall NI. A positive relationship is observed indicating higher population standardised prevalence rates of self-reported COVID-19 symptoms with higher numbers of residents per household (p<0.001; online supplemental table 3). The findings from the current study indicate that higher self-reported prevalence rates are associated with SOAs that have a higher mean number of residents per household suggesting that higher density housing in urban areas increases the risk of transmission.

The relationship between self-reported prevalence rates of COVID-19 symptoms and the measures of area-level deprivation was explored using age standardised rates of self-reported COVID-19 symptom data using COVIDCare NI for two 14-day time periods (ending 20 April 2020 and 11 May 2020) as these time periods provided sufficient age standardised data within all age brackets (online

Table 1 Regression analysis (GLM log link), with time as a factor, of COVID-19 symptom mobile data platforms provided by two sources: KCL ZOE symptom tracker app data for NI and HSC NI Symptom checker feature (COVIDCare NI) and covariate area-level deprivation (MDM)

Date	Estimate	Std. error	T value	Pr(>|t|)	Signif. codes
KCL ZOE					
Intercept	3.135	0.0325	96.505	<2.00E-16	<0.001
06 Apr 2020	0.0002	0.0425	0.005	0.99590	
13 Apr 2020	0.2441	0.0401	6.084	1.25E-09	<0.001
20 Apr 2020	0.2053	0.0431	4.77	1.89E-06	<0.001
27 Apr 2020	0.1405	0.0467	3.011	0.0026	0.001
04 May 2020	0.1508	0.0468	3.225	0.0013	<0.001
11 May 2020	0.1775	0.0465	3.82	0.0001	<0.001
18 May 2020	0.3503	0.0437	8.019	1.29E-15	<0.001
23 May 2020	0.3422	0.0449	7.627	2.8E-14	<0.001
30 May 2020	0.3662	0.0471	7.78	8.57E-15	<0.001
MDM	−0.0022	5.07E-05	−43.457	<2.00E-16	<0.001
COVIDCare NI					
Intercept	3.189	0.0327	97.495	<2e-16	<0.001
27 Apr 2020	−0.0138	0.0414	−0.333	0.7391	
04 May 2020	−0.0009	0.0421	−0.02	0.9837	
11 May 2020	0.0278	0.0424	0.654	0.5128	
18 May 2020	0.0450	0.0433	1.038	0.2995	
23 May 2020	0.1341	0.0423	3.173	0.0015	0.001
30 May 2020	0.4452	0.0384	11.598	<2e-16	<0.001
MDM	−0.0005	0.0004	−11.361	<2e-16	<0.001

The dates shown correspond to the end date of the 14-day symptom reporting sliding window (resulting in a 1-week overlap of data).
GLM, generalised linear model; HSC, health and social care; MDM, Multiple Deprivation Measure; NI, Northern Ireland.

supplemental table 4). A statistically significant negative correlation was found between prevalence rates of self-reported COVID-19 symptoms and area-level deprivation for the age groups 18–24 years and 25–34 years ($p<0.001$ for both time periods for age group 25–34 years). In contrast, a statistically significant positive correlation with area-level deprivation was found for the age groups 50–64 years and >65 years (online supplemental table 4). The results of this current study reveal a statistically significant positive relationship between self-reported prevalence rates of COVID-19 symptoms and mean number of residents per household (housing density) for the age groups<18 years (for both time periods), 35–49 years and 50–64 years (shown for time period ending 11 May; online supplemental table 4). In contrast, a statistically significant negative correlation with population housing density was found for the age group 25–34 years ($p<0.001$ for both time periods for age group 25–34 years). A statistically significant negative correlation was found between self-reported prevalence rates of COVID-19 symptoms and overall social deprivation for the age group 25–34 ($p<0.001$ for both time periods for age group 25–34 years).

DISCUSSION

Our research has shown how a lens of area-level deprivation can deepen our understanding of COVID-19 transmission and prevention. Our findings indicate that social inequality, as measured by area-level deprivation, is associated with disparities in potential COVID-19 infection, with higher prevalence of self-reported COVID-19 symptoms in urban areas associated with area-level deprivation, housing density and age. The findings from the current study provide evidence for the disproportionate adverse effects of the interventions of societal restrictions (eg, ‘lockdown’) on areas of greater deprivation and in particular the impact of higher prevalence of self-reported COVID-19 symptoms in younger populations who have a higher likelihood of living in higher density housing types in urban areas.

There has been much debate and research on the increased risk for the socially vulnerable during natural and human disasters, including the COVID-19 pandemic.[36–38] The pandemic has magnified the heterogeneity in society’s health burden with a disproportionately higher impact on socially vulnerable communities.[20–25] These socioeconomic inequalities are linked directly to area-level deprivation indices including income, education, employment, housing and environment, which contribute to greater risk of poor health.[39–42] Our study has shown the value of using symptom reporting to enable a more granular exploration of social deprivation, housing density and age effect. The findings from our research reveal that the highest self-reported prevalence rates of COVID-19 symptoms were found to be associated with the most deprived areas (lowest social deprivation rankings) and the most deprived areas with lowest ranking for employment. Studies from other countries based on the same time period (March 2020–July 2020) concur with these findings in that the impact of the SARS-CoV-2 infection was found to be higher (up to three times higher on deprived communities.[23] Other studies indicate a link between deprivation and higher mortality rates after infection.[24]

The findings from our research reveal differing relationships with the domains of area-level deprivation across age groups. Higher self-reported prevalence rates of COVID-19 symptoms in the age groups 18–24 and 25oc–34oc occurred in the most deprived areas. This finding suggests that population density may be an important factor for these age groups, which may not be the most at-risk groups for the consequences of infection, but may spread the virus through the community. This research indicates that other factors such as area-level deprivation are more important for the prevalence rates of COVID-19 symptoms for the age groups <18 years, 35–49 years and 50–64 years age group. Incidence across time periods was also affected by the average number of people in a household. The current research suggests that a more in-depth analysis by location is required to examine the influence of rural and urban geography on the effects of area-level deprivation and population housing density on prevalence rates of COVID-19 symptoms. The disproportionate impact of societal restrictions (including 'lockdown') on areas of greater deprivation is a cause for concern and suggests targeted interventions (increased availability and accessibility of testing) are required to mitigate the impacts in areas of higher deprivation.

Limitations

Our research dealt with symptom reporting, which is a combination of (1) COVID-19-induced symptoms and (2) symptoms that are not due to COVID-19. Thus, the signal measured includes COVID-19 prevalence but also includes false positives. The reader is also reminded that the measured signal is a function of: (1) having the requisite symptoms, (2) the propensity to report symptoms, (3) the likelihood to participate in one or other survey, (4) ownership of a smartphone and (5) being part of the at-risk population. Census data for other confounders including adult obesity and respiratory disease were not available at SOA level.

For this research, the estimate of household density was derived by using the mean number of residents per household for each SOA as a proxy for population housing density, as the address of the phone user was not identified due to confidentiality issues. This method suggests a uniform household size per SOA when it is likely to be very heterogeneous. It is acknowledged that this is a limitation of the study which may introduce inaccuracies especially where there is a small population of app users in any SOA.

A third limitation is that symptom-based surveillance and the use of self-reported data may give rise to collider bias when observational data are recorded from non-random samples, involving voluntary participation and self-reported symptoms, which may impact the reliability and generalisability of the findings.[43] It has been suggested that voluntary participants are more likely to be highly educated and health conscious and, therefore, may differ substantially from the general population. Symptom reporting behaviour may also be different across socioeconomic groups.[44]

A fourth potential limitation of the study is that self-reporting participants came from within the adult population who had access to the use of a smartphone (estimated to be 76% of the general NI population). However, the use of two forms of smartphone app enabled a broader sampling of the general population where the geographical spread of the self-reporting participants using the different smartphone apps was investigated through spatial mapping. The greatest geographical coverage was reported for the 14 day period ending 20 April 2020 for both smartphone apps (self-reporting participants from 592 SOAs and 758 SOAs for the KCL ZOE and COVIDCare NI apps, respectively; total 850 SOAs for NI). As such, the main period for this analysis was during the first UK lockdown, when restrictions were more severe.

CONCLUSIONS

COVID-19 symptom prevalence estimates obtained from self-reporting COVID-19 smartphone data were regressed on a range of socioeconomic variables in NI. Significant associations were found between reported COVID-19 prevalence and both area-level deprivation and housing density in urban areas for a range of age groups. The findings underline that social inequality, as measured by area-level deprivation, creates disparities in risk of COVID-19 infection. Specifically, the results from our research indicate a heightening of health inequalities during the period of societal restrictions with higher self-reported prevalence of COVID-19 symptoms associated with areas with the greatest area-level deprivation and the lowest deprivation rankings for employment, particularly within the age group 18–34. This increased reporting rate in the younger population may signal increased prevalence and transmission of the virus, which is likely to have a negative impact on at-risk communities. These findings, therefore, have the potential to inform COVID-19 prevention strategies through targeted messaging to change behaviour ('mask': 'face'; 'space') to mitigate the disproportionate impact of extended periods of societal restrictions (such as 'lockdown') in areas of area-level deprivation.

Author affiliations
[1]School of Natural and Built Environment, Queen's University Belfast, Belfast, UK
[2]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, UK
[3]School of Science, Edith Cowan University, Joondalup, Western Australia, Australia
[4]Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster, UK
[5]Public Health Agency, Belfast, UK
[6]Centre for Health Science, University of the Highlands and Islands, Old Perth Road, Inverness, UK
[7]Centre for Public Health, Queen's University Belfast, Belfast, UK

Twitter Jennifer M McKinley @mckinley_geojen

Acknowledgements This work uses non-identifiable data provided through use of the DoH NI app, COVIDCare NI (formerly known as 'COVID-19 NI'). The app was produced on behalf of the DoH by Digital Health and Care Northern Ireland (DHCNI), working partnership with commercial partners Civica and BigMotive. We acknowledge the access granted to the non-identifiable data, which led to this output. This work also uses data provided by participants of the COVID-19 Symptoms Study, developed by ZOE Global with scientific and clinical input from King's College London. This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research. We acknowledge the responsibility for the interpretation of the information supplied by SAIL is the authors' alone. We acknowledge the collaborative partnership that enabled acquisition and access to the deidentified data, which led to this output. The collaboration was led by BREATHE- The Health Data Research Hub for Respiratory Health, in partnership with SAIL Databank at Swansea University, the Health Data Research UK Swansea University site team and the Usher Institute at the University of Edinburgh. We acknowledge the input of ZOE Global and King's College London in their development and sharing of the data, and their input into the understanding and contextualisation of data for COVID-19 research. All research conducted was completed under the permission and approval of SAIL independent Information Governance Review Panel (IGRP) project number 1078.

Contributors JMM conducted the initial literature searches, conducted the analysis and completed the initial drafts of the manuscript with input from all authors. DC and NA extracted the data and BJ, JMM and CG formatted the data. UM, PMA and FK conducted literature searches. UM, PMA, HVW, DTB and FK reviewed the statistical methods. All authors (JMM, UM, PMA, DC, NA, BJ, CG, HVW, DTB and FK) read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Map disclaimer The depiction of boundaries on the map(s) in this article does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. The map(s) are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. This work uses non-identifiable data provided through use of the DoH NI app, COVIDCare NI (formerly known as 'COVID-19 NI'). Data may be obtained from the data controllers and are not publicly available. The app was produced on behalf of the DoH by Digital Health and Care Northern Ireland (DHCNI), working partnership with commercial partners Civica and BigMotive). Data may be obtained from the data controllers and are not publicly available.This work also uses data provided by participants of the COVID-19 Symptoms Study, developed by ZOE Global Limited with scientific and clinical input from King's College London. This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. Data may be obtained from the data controllers by application through https://www.healthdatagateway.org.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Jennifer M McKinley http://orcid.org/0000-0003-2327-5560
David Cutting http://orcid.org/0000-0002-1088-4749
Conor Graham http://orcid.org/0000-0003-3099-8222
Brian Johnston http://orcid.org/0000-0001-5864-160X
Ute Mueller http://orcid.org/0000-0002-8670-2120
Peter M Atkinson http://orcid.org/0000-0002-5489-6880
Hugo Van Woerden http://orcid.org/0000-0003-3382-1684
Declan T Bradley http://orcid.org/0000-0003-1468-1823

REFERENCES

1. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) situation report 52; 2020.
2. Anderson RM, Heesterbeek H, Klinkenberg D, *et al*. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020;395:931–4.
3. Arenas A, Cota W, Gomez-Gardenes J. A mathematical model for the spatiotemporal epidemic spreading of COVID19. *MedRXiv* 2020.
4. Ferguson NM, Laydon D, Nedjati-Gilani G, *et al*. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. Imperial College COVID-19 Response Team, 2020.
5. Metcalf CJE, Morris DH, Park SW. Mathematical models to guide pandemic response. *Science* 2020;369:368–9.
6. Verity R, Okell LC, Dorigatti I, *et al*. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 2020;20:669–77.
7. Drew DA, Nguyen LH, Steves CJ, *et al*. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020;368:1362–7.
8. Menni C, Valdes AM, Freidin MB, *et al*. Real-Time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26:1037–40.
9. Chan AT, Brownstein JS. Putting the public back in public health — surveying symptoms of Covid-19. *N Engl J Med Overseas Ed* 2020;383:e45.
10. Jones KH, Ford DV, Jones C, *et al*. A case study of the secure anonymous information linkage (Sail) gateway: a privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014;50:196–204.
11. Health and social care Northern Ireland public health agency COVID-19 monthly epidemiological Bulletins. Available: https://www.publichealth.hscni.net/publications/coronavirus-bulletin
12. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol* 2017;27:1–9.
13. NHS England. Definition of health inequalities. Available: https://www.england.nhs.uk/ [Accessed 2 Dec 2020].
14. Kwan M-P. *Geographies of health, disease and well-being. recent advances in theory and method. 1st edition ISBN 9780415870016*. Routledge, 2014.
15. Diez Roux AV. Investigating neighborhood and area effects on health. *Am J Public Health* 2001;91:1783–9.
16. Jones NL, Breen N, Das R, *et al*. Cross-Cutting themes to advance the science of minority health and health disparities. *Am J Public Health* 2019;109:S21–4.
17. Duran DG, Pérez-Stable EJ. Novel approaches to advance minority health and health disparities research. *Am J Public Health* 2019;109:S8–10.
18. Breen N, Jackson JS, Wood F. Translational health disparities research in a Data-Rich world. methods and measurement science. *AJPH* 2019;109.
19. Williamson E, Walker AJ, Bhaskaran KJ. OpenSAFELY: factors associated with COVID-19-related Hospital death in the linked

electronic health records of 17 million adult NHS patients. *medRxiv* 2020.
20 NHS Confederation. The impact of COVID-19 on BME communities and health and care staff, BME leadership network, members Briefing April 2020. Available: https://www.nhsconfed.org/resources/2020/04/the-impact-of-covid19-on-bme-communities-and-staff [Accessed 9 Jun 2020].
21 Wadhera RK, Wadhera P, Gaba P, *et al*. Variation in COVID-19 hospitalizations and deaths across New York City boroughs. *JAMA* 2020;323:2192–5.
22 Yancy CW. COVID-19 and African Americans. *JAMA* 2020;323:1891–2.
23 Lewis NM, Friedrichs M, Wagstaff S, *et al*. Disparities in COVID-19 Incidence, Hospitalizations, and Testing, by Area-Level Deprivation - Utah, March 3-July 9, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:1369–73.
24 Brainard J, Rushton S, Winters T. Spatial risk factors for Pillar 1 COVID-19 case counts and mortality in rural eastern England, UK. *medRxiv* 2020.
25 Akanbi MO, Rivera AS, Akanbi FO, *et al*. An ecologic study of disparities in COVID-19 incidence and case fatality in Oakland County, MI, USA, during a State-Mandated shutdown. *J Racial Ethn Health Disparities* 2020. doi:10.1007/s40615-020-00909-1. [Epub ahead of print: 29 Oct 2020].
26 Northern Ireland Statistics and Research Agency (NISRA). *2021 census Northern Ireland proposals document*, 2019: 52.
27 Ofcom. Communications Market Report 2017 - Northern Ireland. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/104927/northern-ireland-internet-online.pdf [Accessed 25 Aug 2020].
28 Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Available: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [Accessed 25 Aug 2020].
29 Public health England coronavirus (COVID-19) guidance and support, 2020. Available: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-initial-investigation-of-possible-cases/investigation-and-initial-clinical-management-of-possible-cases-of-wuhan-novel-coronavirus-wn-cov-infection#criteria [Accessed 24 May 2020].
30 Northern Ireland Statistics and Research Agency (NISRA). *Ni multiple deprivation measures 2017 –Summary booklet*, 2017: 28.
31 Northern Ireland Statistics and Research Agency (NISRA). *Northern Ireland census 2011 key statistics summary report*, 2014: 247.
32 Ministry of Housing, Communities & Local Government, UK. *The English indices of deprivation 2019 technical report*, 2019.
33 Layte R, Whelan CT. Explaining social class inequalities in smoking: the role of education, self-efficacy, and deprivation. *Eur Sociol Rev* 2009;25:399–410.
34 Anselin L. Under the hood issues in the specification and interpretation of spatial regression models. *Agric Econ* 2002;27:247e67:247–67.
35 Anselin L. Spatial regression. In: *The SAGE Handbook of spatial analysis*, 2009: 255e76.
36 Karaye IM, Horney JA. The impact of social vulnerability on COVID-19 in the U.S.: an analysis of spatially varying relationships. *Am J Prev Med* 2020;59:317–25.
37 Palaiologou P, Ager AA, Nielsen-Pincus M, *et al*. Social vulnerability to large wildfires in the Western USA. *Landsc Urban Plan* 2019;189:99–116.
38 Rufat S, Tate E, Burton CG, *et al*. Social vulnerability to floods: review of case studies and implications for measurement. *Int J Disaster Risk Reduct* 2015;14:470–86.
39 Gilbert KL, Ray R, Siddiqi A, *et al*. Visible and invisible trends in black men's health: pitfalls and promises for addressing racial, ethnic, and gender inequities in health. *Annu Rev Public Health* 2016;37:295–311.
40 Galea S. Health haves, health have nots, and heterogeneity in population health. *Lancet Public Health* 2017;2:e388–9.
41 Griffith DM, Jaeger EC, Sherman LD. Patterns and causes of health inequities during a pivotal period in the life course. In: *Men's Health Equity: A Handbook*. New York, NY: Routlege, 2019: 72.
42 Bambra C, Riordan R, Ford J, *et al*. The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health* 2020;74:964–968.
43 Griffith GJ, Morris TT, Tudball MJ, *et al*. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020;11:5749.
44 Kroenke K. Studying symptoms: sampling and measurement issues. *Ann Intern Med* 2001;134:844–53.