BMJ Open

The association between community-based self-reported COVID-19 symptoms and social deprivation explored using symptom tracker apps

Journal:	*BMJ Open*
Manuscript ID	Draft
Article Type:	Original research
Date Submitted by the Author:	n/a
Complete List of Authors:	McKinley, Jennifer; Queen's University Belfast, School of Natural and Built Environment Cutting, David; Queen's University Belfast, School of Natural and Built Environment Anderson, Neil; Queen's University Belfast, School of Natural and Built Environment Graham, Conor; Queen's University Belfast, School of Natural and Built Environment Johnston, Brian; Queen's University Belfast, School of Natural and Built Environment Mueller, Ute; Edith Cowan University Atkinson, Peter; Lancaster University Van Woerden, Hugo; Public Health Agency Northern Ireland Bradley, Declan; Public Health Agency, Public Health Medicine; Queen's University Belfast, Centre for Public Health Kee, Frank; Queen's University Belfast, Centre for Statistical Science and Operational Research (CenSSOR)
Keywords:	COVID-19, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

The association between community-based self-reported COVID-19 symptoms and social deprivation explored using symptom tracker apps

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 [6], Declan T. Bradley [5,6] and Frank Kee[5,6]

[1] School of Natural and Built Environment, Queen's University Belfast, Northern Ireland

[2] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Northern Ireland

[3] School of Science, Edith Cowan University, Perth, Western Australia

[4] Lancaster Environment Centre, Lancaster University, UK

[5] Centre for Public Health, Queen's University Belfast, Northern Ireland

[6] Public Health Agency, Northern Ireland

* Corresponding author: Jennifer 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 and area-level social deprivation.

Design: Spatial mapping, generalised linear models and spatial-lag models were used to explore the relation between self-reported COVID-19 symptom prevalence as recorded through two smartphone symptom tracker apps and a range of socio-economic factors.

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

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 and with those areas with the lowest employment levels. Higher rates of COVID-19 symptoms within the age groups, 18-24 and 25-34 yrs were found within the most deprived areas during the earliest stages of non-pharmaceutical interventions and societal restrictions ('lockdown'). A significant, positive correlation between self-reported prevalence rates and population housing density was observed, indicating higher prevalence rates for higher density housing areas.

Conclusions: Through spatial regression of self-reporting COVID-19 smartphone data in the community this research underlines the link between health and place and the potential negative impact on health disparities of social deprivation, housing density and age. The findings highlight higher prevalence of self-reported COVID-19 symptoms between 18-34 yrs within the most deprived areas. This has potential implications for targeted public health interventions, including messaging for future periods of societal restrictions or 'lockdown'.

Key words: COVID-19 self-reported symptoms; social deprivation; population housing density; (spatial) regression; societal restrictions or 'lockdown'.

Strengths and Limitations of this study

- The geographic spread of the self-reporting participants using the different smartphone apps was investigated through spatial mapping and regression.
- The use of two apps from different smartphone app providers enabled a broad sampling of the general population.
- 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.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Patient consent for publication: It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Data availability statement:

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.

In addition this work 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.

Word Count: 4,039

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 infecteds 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].

The importance of the link between health and place is widely recognised [11]. Health inequalities are defined as differences in health across the population, and between different groups within society [12]. An interplay of factors at multiple levels can influence health inequalities, including the physical and socio-economic environment [13-16]. Limitations in data sampling, data collection and analysis techniques have constrained our understanding of the causes of these disparities [17]. 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 socio-economic status and deprivation, vulnerable groups of society, or 'inclusion health' groups and geography (NHS). 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

Date (end of 14 day reporting period)	KCL ZOE (Total number of users)	COVIDCare NI (Total number of users)	KCL ZOE Classic PHE Reported symptoms	KCL ZOE Refined PHE Reported symptoms	COVIDCare NI Reported symptoms
2020.03.20	22159		3013	3381	
2020.04.06	24949		3658	4161	
2020.04.13	15990		2163	2457	
2020.04.20	16675	6403	1623	1849	2173
2020.04.27	17692	5882	1392	1584	1912
2020.05.04	17992	5068	1320	1503	1577
2020.05.11	17767	4487	1367	1476	1282

recognised that the COVID-19 crisis has disproportionately affected certain at-risk communities, based on their previous health, socio-economic position and ethnic characteristics [18-21]. 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.

Methods

In the UK, regional devolved governments in England, Scotland, Wales and Northern Ireland (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) [22].   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.joinzoe.com/) and is available to download throughout the UK [10]. The Northern Ireland Health and Social Care (HSC) service launched its own symptom tracker app, COVIDCare NI (formerly known as 'COVID-19 NI'), on 6 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. Both smartphone apps provided a way to track the spatial and temporal spread of the virus spread through Northern Ireland by self-reported health information from both asymptomatic and symptomatic individuals.

The study concentrates on the reporting period 24 March – 22 June 2020 at the earliest stages of non-pharmaceutical interventions and societal restrictions ('lockdown'). The UK KCL ZOE symptom tracker app provided data for Northern Ireland for the full study period (24 March – 22 June 2020) whereas the HSC NI Symptom checker feature (COVIDCare NI) provided data for the reporting period 6 April to 22 June 2020 (Table 1). Smartphone data will not be fully representative of the whole population in NI, although smartphone ownership does not vary significantly by urban or rural location in NI [23] and in 2017, 76% of adults in NI reported ownership of smartphones. While NI specific data on smartphone ownership are not available after 2017, NI shows strong alignment with UK on ownership up until 2017. In 2019, 76% of adults in the UK as a whole reported smartphone ownership [24].

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 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 case definitions [25]. Data containing invalid postcodes or postcodes outside of NI were removed during post-processing.

For both datasets, data 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 postcode to SOAs. There are 890 SOA administrative areas across Northern Ireland [26]. When the numbers of users or those reporting symptoms (from either app) were too small in any SOA ($n \leq 5$) the data providers suppressed these small cell counts to avoid disclosure risk. By "reporting symptoms" we mean that, on any given date, symptoms would have satisfied the PHE case definition [25].

2020.05.18	16340	3721	1224	1325	928
2020.05.25	15180	3116	1129	1228	804
2020.06.01	14380	1692	860	968	430
2020.06.15	12395	939	627	732	420
2020.06.20	11778	667	548	647	309

Table 1: Summary data for COVID-19 symptom mobile data platforms provided by two sources: KCL ZOE symptom tracker app data for NI (reporting period 24 March – 22 June 2020) and HSC NI Symptom checker feature (COVIDCare NI), (reporting period 6 April to 22 June 2020). The dates shown correspond to the end date of the 14 day symptom reporting sliding window (resulting in a one week overlap of data). The data represent a time series of the prevalence of self-reported symptoms. KCL ZOE symptom tracker app data for refined symptoms as defined by PHE COVID case definitions [25].

Measurement of social deprivation

Social deprivation was characterised using the Northern Ireland Multiple Deprivation Measure 2017 (NIMDM) provided by the Northern Ireland Statistics and Research Agency (NISRA) [26]. The NIMDM 2017 provided information on seven individual domains of deprivation and an overall MDM ranking for each SOA (Table 2; Fig.1)) [26]. The ranking scale was from 1 (most deprived) to 890 (least deprived). We used the NIMDM 2017 domains of income, employment and education as indicators of socio-economic factors [26, 27]. Following an initial analysis involving all MDM domains, NI census 2011 data were used to further investigate for population household density [28]. This was derived as the number of residents divided by number of households for each SOA.

NIMDM							
Domains of Deprivation	Income	Employment	Health deprivation and disability	Education, skills and training	Access to services	Living Environment	Crime and disorder
Relative contribution to overall MDM	25%	25%	15%	15%	10%	5%	5%

Table 2: Northern Ireland Multiple Deprivation Measures 2017 (NIMDM) provided by the Northern Ireland Statistics and Research Agency (NISRA) [26]. Composition and relative contribution of individual deprivation domains to the overall Multiple Deprivation Measure (MDM).

Generalised Linear and Spatial Regression analysis

KCL ZOE symptom tracker app data for NI with revised PHE case definitions and COVIDCare NI were used for regression analysis (Table 1). 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 the 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 [28]. 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, >65. Regression analysis was conducted using R version 4.0.0.

Regression analysis (Generalised linear model (GLM) using *glm* R package with log link) was undertaken using the data provided by both COVID-19 self-reporting symptom mobile platforms. The regression models were fitted between the dependent variable 'population standardised self-reported COVID-19 symptoms' and the independent variables or covariates, that is, social deprivation indices using overall MDM and then with individual MDM domains (adjusting p values for multiple comparisons using the Bonferroni correction). Census data were used to investigate the relationship between population standardised self-reported COVID-19 symptoms and population household density, derived as the mean number of residents per household for each SOA [28].

Regression models assume independence between the regression residuals [11, 29]. However, this assumption may not be valid for spatial data [30]. Several techniques are available which incorporate spatial parameters, including spatial regression [29] and geographically weighted regression (GWR), a spatially non-stationary regression technique used to characterise the spatially varying relation between the dependent and explanatory variables [31-33].

Understanding the interactions between interdependent areas (SOAs) may be critical to understanding the prevalence of COVID-19. A spatially lagged model incorporates spatial dependence explicitly into the regression equation and as such acknowledges that prevalence in neighbouring areas (SOAs in this research) may be an important predictor for the estimation of prevalence rates in the area of interest. To assess the nature of the impact of spatial autocorrelation Moran's I statistic was calculated to test for spatial autocorrelation in the residuals computed from the regression models. In the case where the Moran's I for the residuals indicated that spatial autocorrelation may be present, a spatial lag model using the *spatialreg* R package, was used to adjust for spatial autocorrelation. Where the Moran's I for the residuals was found to be significantly different from random, the GLM regression results where compared with a spatial lag model and the model fit compared using an Akaike Information Criterion (AIC).

Results

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 to 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 and 6 April 2020 followed by a sharp decrease after 6 April 2020 (Table 1). The COVIDCare NI app shows a decrease in active users reporting COVID-19 symptoms from the start of reporting period 21 April. The COVIDCare NI app data shows an increase in the percentage of App users self-reporting COVID-19 symptoms from the 30 May (Fig. 2). Although the overall rate (per 100,000 population) of self-reported symptoms is comparable, the geographic coverage varies across the time periods (Fig 3a and b). The participants using the KCL ZOE COVID-19 symptom tracker app are encouraged and reminded to update their status with regards to whether they have active symptoms. In this way, the ZOE app records are more comparable to a longitudinal data than those from the COVIDCare NI app, which was designed to provide a triage system. Both apps show a decrease in reported COVID-19 symptoms over time which mirrors the reported peak and subsequent decline in COVID-19 cases.

Investigating the influence of area-level social deprivation and population housing density

Regression analysis (GLM log link) of population standardised self-reported prevalence of COVID-19 symptoms reveals a statistically significant negative correlation between active users of both mobile platforms reporting symptoms and the overall MDMs ($p < 0.001$) across the reporting dates (Table 3). These findings indicate that throughout the reporting period, from initial lockdown when restrictions were most stringent, the most deprived SOAs (lowest MDM rankings) were associated with higher population standardised prevalence rates of self-reported COVID-19 symptoms.

Time period	KCLZOE Population Standardised rates					COVIDCare NI Population Standarised rates			
		Estimate	Std. Error	Pr(>|t|)	Signif. Codes	Estimate	Std. Error	Pr(>|t|)	Signif. Codes
30.03.2020	Intercept	2.592	0.042	<2e-16	<0.001				
	MDM	-0.001	0.000	<2e-16	<0.001				
	Intercept	2.714	0.142	<2e-16	<0.001				

	Housing Density	-0.223	0.054	0.00004	<0.001				
06.04.2020	Intercept	2.564	0.040	<2e-16	<0.001				
	MDM	-0.001	0.000	<2e-16	<0.001				
	Intercept	2.681	0.135	<2e-16	<0.001				
	Housing Density	-0.193	0.051	0.00015	<0.001				
13.04.2020	Intercept	2.911	0.049	<2e-16	<0.001				
	MDM	-0.001	0.000	<2e-16	<0.001				
	Intercept	2.588	0.170	<2e-16	<0.001				
	Housing Density	-0.119	0.064	0.06370	0.05				
20.04.2020	Intercept	2.902	0.060	<2e-16	<0.001	2.986	0.044	<2E-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.00001	<0.001
	Intercept	2.679	0.238	<2e-16	<0.001	3.257	0.166	<2e-16	<0.001
	Housing Density	-0.209	0.091	0.02240	0.001	-0.169	0.063	0.00782	0.01
27.04.2020	Intercept	2.847	0.066	<2e-16	<0.001	2.942	0.046	<2e-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.00021	<0.001
	Intercept	2.621	0.260	<2e-16	<0.001	3.196	0.175	<2e-16	<0.001
	Housing Density	-0.226	0.099	0.02260	0.001	-0.154	0.066	0.02040	0.001
04.05.2020	Intercept	2.912	0.064	<2e-16	<0.001	2.984	0.050	<2e-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.00001	<0.001
	Intercept	2.617	0.270	<2e-16	<0.001	3.324	0.186	<2e-16	<0.001
	Housing Density	-0.231	0.103	0.02440	0.001	-0.204	0.071	0.00398	0.001
11.05.2020	Intercept	2.937	0.067	<2e-16	<0.001	2.950	0.053	<2e-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.00183	0.001
	Intercept	2.706	0.285	<2e-16	<0.001	3.278	0.200	<2e-16	<0.001
	Housing Density	-0.261	0.108	0.01620	0.001	-0.181	0.076	0.01780	0.001
18.05.2020	Intercept	3.163	0.073	<2e-16	<0.001	2.950	0.053	<2e-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.00183	0.001
	Intercept	2.801	0.320	<2e-16	<0.001	3.278	0.200	<2e-16	<0.001
	Housing Density	-0.271	0.122	0.02740	0.01	-0.181	0.076	0.01780	0.01
25.05.2020	Intercept	3.149	0.073	<2e-16	<0.001	3.077	0.058	<2e-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.00039	<0.001
	Intercept	2.883	0.308	<2e-16	<0.001	3.749	0.228	<2e-16	<0.001
	Housing Density	-0.289	0.118	0.01480	0.01	-0.327	0.087	0.00020	<0.001
01.06.2020	Intercept	3.156	0.084	<2e-16	<0.001	3.355	0.065	<2e-16	<0.001
	MDM	-0.002	0.000	<2e-16	<0.001	0.000	0.000	0.02970	0.01
	Intercept	2.681	0.349	0.00000	<0.001	3.859	0.242	<2e-16	<0.001
	Housing Density	-0.201	0.135	0.13600		-0.242	0.093	0.00959	0.001

Table 3: Regression analysis (GLM log link) results for covariates overall MDMs and population housing density using population standardised self-reported prevalence of COVID-19 symptom provided by two sources: KCL ZOE symptom tracker app (30.3.20 – 01.06.20) and COVIDCare NI (20.04.20- 01.06.20) The dates correspond to the end date of 14 day symptom reporting sliding window.

To assess the impact of spatial autocorrelation on the regression models, Moran's *I* test statistic was used to test the residuals computed from the regression models (Fig. 3c and d; results shown for time period ending 20 April 2020 for population standardised self-reported prevalence of COVID-19 symptom using both mobile app sources: KCL ZOE symptom tracker app and COVIDCare NI). For the time period ending 20 April, for COVIDCare NI, the GLM produced an AIC of 7159.7 whereas the spatial lag model produced an AIC of 7161.5, indicating that the GLM regression model provided a better fit (Moran's *I* statistic 0.00895; *p*-value = 0.6048). Using the KCLZOE app data, the GLM produced an AIC of 6787.4 whereas the spatial lag model produced an AIC of 6767.6, indicating that the regression model provided a better fit when a spatial lag was included (Moran's *I* statistic 0.0842, *p*-value = 1.892e-05). Where the spatial lag model was shown to provide a better fit, such as for the 20.04.20 time period for KCLZOE App data, the results remained consistent and indicated a statistically significant negative correlation between active users of both mobile platforms reporting symptoms and the overall MDMs ($p < 0.001$).

Varying relationships with the different MDM domains were observed across the analysis period. However, a negative statistically significant correlation was found between users reporting COVID-19 symptoms and the social deprivation measures of *employment* ($p < 0.001$) and *living environment* ($p = 0.01$) recorded for data from both mobile platforms, (Table 4). The deprivation domain of *employment* is defined as the proportion of working age population who are either in receipt of at least one of five employment-related benefits, or who are not in receipt of any of these benefits and have not received income from employment [26]. The findings suggest that during the initial lockdown when restrictions were most stringent, the most deprived SOAs with lowest ranking for employment (and living environment) were associated with higher population standardised prevalence rates of self-reported COVID-19 symptoms.

Time period	KCLZOE Population Standardised rates					COVIDCare NI Population Standardised rates			
		Estimate	Std. Error	Pr(>|t|)	Signif. Codes	Estimate	Std. Error	Pr(>|t|)	Signif. Codes
30.03.2020	(Intercept	2.592	0.04172	<2e-16	<0.001				
	MDM	-0.00101	7.95E-05	<2e-16	<0.001				
	(Intercept)	2.601	0.07903	<2e-16	<0.001				
	Income	0.000446	0.000123	0.000296	<0.001				
	Employment	-0.00174	0.000294	4.35E-09	<0.001				
	Health	0.000458	0.000306	0.134074					
	Education	-9.7E-05	0.000168	0.564402					
	Service	0.000142	9.51E-05	0.136053					
	Living	-0.00023	7.93E-05	0.004038	0.001				
	Intercept	2.71397	0.1424	<2e-16	<0.001				
	Housing Density	-0.22301	0.05365	3.57E-05	<0.001				
06.04.2020	(Intercept	2.564	0.03953	<2e-16	<0.001				
	MDM	-0.00086	7.6E-05	<2e-16	<0.001				
	(Intercept)	2.545	0.07483	<2e-16	<0.001				
	Income	0.000422	0.000117	0.000309	<0.001				
	Employment	-0.00132	0.000282	3.22E-06	<0.001				
	Health	0.00028	0.000294	0.341318					
	Education	-0.00016	0.000161	0.326395					
	Service	0.000202	9.1E-05	0.026278	0.01				

	Living	-0.00024	7.61E-05	0.001964	0.001				
	Intercept	2.68148	0.13516	<2e-16	<0.001				
	Housing Density	-0.19342	0.05087	0.000154	<0.001				
13.04.2020	(Intercept	2.911	0.04925	<2e-16	<0.001				
	MDM	-0.00135	9.24E-05	<2e-16	<0.001				
	(Intercept)	3.128	0.09155	<2e-16	<0.001				
	Income	0.000525	0.00014	0.00019	<0.001				
	Employment	-0.00211	0.000338	7.33E-10	<0.001				
	Health	0.000405	0.000356	0.2561					
	Education	-0.00012	0.000195	0.52768					
	Service	-0.00012	0.00011	0.27473					
	Living	-0.00041	9.23E-05	9.82E-06	<0.001				
	Intercept	2.58756	0.16975	<2e-16	<0.001				
	Housing Density	-0.11894	0.06404	0.0637	0.05				
20.04.2020	(Intercept	2.901861	0.059686	<2e-16	<0.001	2.985542	0.043537	<2E-16	<0.001
	MDM	-0.00158	0.00011	<2e-16	<0.001	-0.00038	8.53E-05	9.03E-06	<0.001
	(Intercept)	3.152	0.1131	<2e-16	<0.001	2.944	0.08769	<2e-16	<0.001
	Income	0.000309	0.000169	0.06806	0.05	0.000117	0.000138	0.39375	
	Employment	-0.00165	0.000408	5.95E-05	<0.001	-0.00088	0.000339	0.0093	0.001
	Health	-0.00037	0.000436	0.39026		0.000513	0.000352	0.14558	
	Education	6.87E-05	0.000234	0.76857		2.06E-07	0.000194	0.99915	
	Service	-0.00015	0.000135	0.27296		0.000214	0.000109	0.04972	0.01
	Living	-0.00034	0.000111	0.00242	0.001	-0.00024	8.92E-05	0.00717	0.001
	Intercept	2.67865	0.23847	<2e-16	<0.001	3.25668	0.16625	<2e-16	<0.001
	Housing Density	-0.20852	0.09104	0.0224	0.01	-0.16853	0.06319	0.00782	0.001
27.04.2020	(Intercept	2.847141	0.065584	<2e-16	<0.001	2.942	0.04633	<2e-16	<0.001
	MDM	-0.00168	0.00012	<2e-16	<0.001	-0.00034	9.15E-05	0.000213	<0.001
	(Intercept)	3.155316	0.119768	<2e-16	<0.001	2.921	0.09315	<2e-16	<0.001
	Income	0.00025	0.000191	0.191589		0.00024	0.000146	0.101279	
	Employment	-0.00159	0.000458	0.000553	<0.001	-0.001	0.000361	0.005599	0.001
	Health	-0.00036	0.000477	0.453552		0.000346	0.000376	0.357288	
	Education	-2.9E-05	0.000249	0.906554		0.000302	0.000204	0.139552	
	Service	-0.00019	0.000144	0.191344		0.000186	0.000114	0.10324	
	Living	-0.00042	0.000118	0.000366	<0.001	-0.00035	9.49E-05	0.000255	<0.001
	Intercept	2.62074	0.2599	<2e-16	<0.001	3.19612	0.17485	<2e-16	<0.001
	Housing Density	-0.22596	0.09882	0.0226	0.01	-0.15419	0.06633	0.0204	0.01
04.05.2020	(Intercept	2.912	0.064255	<2e-16	<0.001	2.984	0.04974	<2e-16	<0.001
	MDM	-0.00186	0.000118	<2e-16	<0.001	-0.00043	9.7E-05	1.11E-05	<0.001
	(Intercept)	3.223	0.1117	<2e-16	<0.001	2.913	0.09993	<2e-16	<0.001

	Income	0.000257	0.000179	0.153		0.000134	0.000156	0.39255	
	Employment	-0.00194	0.000456	2.64E-05	<0.001	-0.00052	0.000382	0.17245	
	Health	-0.00022	0.000475	0.642		-0.00015	0.000397	0.70212	
	Education	3.1E-05	0.000255	0.903		0.000333	0.00021	0.11447	
	Service	-0.00015	0.000136	0.27		0.000356	0.000121	0.00332	0.001
	Living	-0.00052	0.000115	6.38E-06	<0.001	-0.00041	9.99E-05	0.000054	<0.001
	Intercept	2.6171	0.2702	<2e-16	<0.001	3.32441	0.18577	<2e-16	<0.001
	Housing Density	-0.2314	0.1025	0.0244	0.01	-0.2038	0.07052	0.00398	0.001
11.05.2020	(Intercept	2.937279	0.066814	<2e-16	<0.001	2.950484	0.052882	<2e-16	<0.001
	MDM	-0.00186	0.000121	<2e-16	<0.001	-0.00033	0.000105	0.00183	0.001
	(Intercept)	3.195056	0.120929	<2e-16	<0.001	2.727365	0.110361	<2e-16	<0.001
	Income	0.000251	0.000187	0.181		0.000198	0.000169	0.242164	
	Employment	-0.00203	0.000467	1.73E-05	<0.001	-0.00092	0.000421	0.028666	0.01
	Health	0.000238	0.000487	0.625		0.000227	0.000441	0.60584	
	Education	-0.00032	0.000254	0.206		0.000465	0.000236	0.048956	0.01
	Service	-8.1E-05	0.000144	0.574		0.000496	0.000133	0.000213	<0.001
	Living	-0.00049	0.000119	4.35E-05	<0.001	-0.00029	0.000108	0.007755	0.001
	Intercept	2.7058	0.2848	<2e-16	<0.001	3.2781	0.1998	<2e-16	<0.001
	Housing Density	-0.2612	0.1083	0.0162	0.01	-0.1813	0.0763	0.0178	0.01
18.05.2020	(Intercept	3.16297	0.073018	<2e-16	<0.001	2.950484	0.052882	<2e-16	<0.001
	MDM	-0.00213	0.000131	<2e-16	<0.001	-0.00033	0.000105	0.00183	0.001
	(Intercept)	3.502711	0.136056	<2e-16	<0.001	2.727365	0.110361	<2e-16	<0.001
	Income	-8.2E-05	0.000211	0.69876		0.000198	0.000169	0.242164	
	Employment	-0.00149	0.00049	0.00249	0.001	-0.00092	0.000421	0.028666	0.01
	Health	-0.00042	0.000511	0.40992		0.000227	0.000441	0.60584	
	Education	-0.0002	0.000276	0.4598		0.000465	0.000236	0.048956	0.01
	Service	-0.00011	0.000161	0.49485		0.000496	0.000133	0.000213	<0.001
	Living	-0.00058	0.000129	9.32E-06	<0.001	-0.00029	0.000108	0.007755	0.001
	Intercept	2.8014	0.3197	<2e-16	<0.001	3.2781	0.1998	<2e-16	<0.001
	Housing Density	-0.2706	0.1223	0.0274	0.01	-0.1813	0.0763	0.0178	0.01
25.05.2020	(Intercept	3.148813	0.073451	<2e-16	<0.001	3.077436	0.05844	<2e-16	<0.001
	MDM	-0.00203	0.000131	<2e-16	<0.001	-0.00042	0.000118	0.000386	<0.001
	(Intercept)	3.415	0.1349	<2e-16	<0.001	2.843871	0.12651	<2e-16	<0.001
	Income	-0.00035	0.000208	0.09307	0.015	0.000237	0.000194	0.222938	
	Employment	-0.0013	0.000474	0.0066	0.001	-0.00107	0.000475	0.024396	0.01
	Health	-0.0004	0.000506	0.42472		0.000589	0.000491	0.230583	
	Education	-0.00011	0.000281	0.6945		0.000178	0.000287	0.534702	
	Service	5.23E-06	0.000159	0.97381		0.000522	0.000155	0.000798	<0.001

	Living	-0.00045	0.000132	0.00066	<0.001	-0.00033	0.000125	0.007764	0.001
	Intercept	2.8831	0.308	<2e-16	<0.001	3.7487	0.2283	<2e-16	<0.001
	Housing Density	-0.2889	0.1181	0.0148	0.01	-0.3266	0.087	0.000195	<0.001
01.06.2020	(Intercept	3.156103	0.084329	<2e-16	<0.001	3.355442	0.064846	<2e-16	<0.001
	MDM	-0.00201	0.000152	<2e-16	<0.001	-0.00028	0.00013	0.0297	0.01
	(Intercept)	3.281	0.1579	<2e-16	<0.001	3.201214	0.131909	<2e-16	<0.001
	Income	-0.00023	0.000236	0.32374		0.000343	0.000211	0.104	
	Employment	-0.00184	0.000555	0.00105	0.001	-0.00115	0.000493	0.0204	0.01
	Health	1.42E-05	0.00062	0.98177		0.000595	0.000485	0.2207	
	Education	-6.7E-05	0.000347	0.84715		0.000429	0.00028	0.1261	
	Service	8.87E-05	0.000185	0.63187		0.000532	0.000161	0.001	0.001
	Living	-0.00025	0.000156	0.10542		-0.00066	0.000133	1.25E-06	<0.001
	Intercept	2.681	0.3491	1.95E-13	<0.001	3.859	0.24244	<2e-16	<0.001
	Housing Density	-0.2013	0.1347	0.136		-0.24218	0.09304	0.00959	0.001

Table 4: Regression analysis (GLM log link) results for covariates overall MDMs and individual MDM domains using population standardised self-reported prevalence of COVID-19 symptom provided by two sources: KCL ZOE symptom tracker app (30.3.20 – 01.06.20) and COVIDCare NI (20.04.20- 01.06.20) The dates correspond to the end date of 14 day symptom reporting sliding window.

Using the mean number of residents per household for each SOA as a proxy for population housing density, a consistent negative statistically significant correlation was found between self-reported prevalence rates of COVID-19 symptoms and mean number of residents per household across all time periods (Table 3). 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 given the premise of contagion risk models, with the expectation that higher density housing would increase the risk of transmission. On a NI regional scale clusters of SOAs with high levels of self-reported symptom rates were observed, for both COVID-19 symptom mobile trackers, in rural SOAs (Fig 3a and b). The results suggest that a more in depth analysis by location is required to examine the influence of rural and urban geography on the impacts of social deprivation and population housing density on prevalence rates of COVID-19 symptoms.

Age standardised results

Further analysis to explore the relationship between self-reported prevalence rates of COVID-19 symptoms and the measures of social deprivation was completed using age standardised rates of self-reported COVID-19 symptom data (using COVIDCare NI). Two 14 day time periods (ending 20.04.20 and 11.05.20) were used as these time periods provided sufficient age standardised data within all age brackets (Table 5). A statistically significant negative correlation was found between self-reported prevalence rates of COVID-19 symptoms and overall MDMs for the age groups 18-24 and 25-34 (p <0.001 for both time periods for age group 25-34 yrs). This suggests that there is higher self-reported prevalence of COVID-19 symptoms within the age groups 18-24 yrs and 25-34 yrs within the most deprived SOAs with lowest MDMs ranking. In contrast, a statistically significant positive correlation with overall MDMs was found for the age groups 50-64 yrs and >65yrs (Table 5). The results reveal a statistically significant positive relationship between self-reported prevalence rates of COVID-19 symptoms and mean number of residents per household (housing density) was found for the age groups <18 yrs (for both time periods), 35-49 yrs and 50-64 yrs (shown for time period ending 11th May; Table 5). This suggests that population density may be important for age groups. In contrast, a statistically significant negative correlation with population housing density was found for the age group 25-34 yrs (p <0.001 for both time periods for age group 25-34 yrs Table 5). This may indicate that other factors such as social deprivation are more important for the prevalence rates of COVID-19 symptoms in this age group. As shown above a statistically significant negative correlation was found between self-reported prevalence rates of COVID-19 symptoms and overall MDMs for the age group 25-34 (*p <0.001* for both time periods for age group 25-34 yrs).

20.04.20		Estimate	Std. Error	Pr(>|t|)	Signif. Codes
Age group <18	Intercept	9.48E-02	1.08E-02	2.76E-15	<0.001
159 SOAs	MDM	-1.65E-05	2.11E-05	0.437	
	Intercept	-0.05261	0.04439	0.23769	
	Housing Density	0.05352	0.01684	0.00178	0.001
Age group 18-24	Intercept	4.76E-02	4.01E-03	<2e-16	<0.001
178 SOAs	MDM	-2.12E-05	8.25E-06	0.0111	0.1
	Intercept	0.028072	0.016148	0.0839	0.05
	Housing Density	0.004178	0.006175	0.4995	
Age group 25-34	Intercept	4.52E-02	2.43E-03	<2e-16	<0.001
349 SOAs	MDM	-2.31E-05	4.95E-06	4.60E-06	<0.001
	Intercept	0.061368	0.00958	4.86E-10	<0.001
	Housing Density	-0.010085	0.003696	0.00669	0.001
Age group 35-49	Intercept	1.26E-01	5.81E-03	<2e-16	<0.001
477 SOAs	MDM	-9.27E-06	1.13E-05	0.411	
	Intercept	0.082366	0.021898	0.00019	<0.001
	Housing Density	0.015207	0.008348	0.06916	0.05
Age group 50-64	(Intercept	9.39E-02	9.81E-03	<2e-16	<0.001
296 SOAs	MDM	5.45E-05	1.81E-05	0.00281	0.001
	Intercept	0.06483	0.03445	0.0608	0.05
	Housing Density	0.02115	0.01305	0.106	
Age group >65	Intercept	3.88E-02	1.15E-02	0.00112	0.001
77 SOAs	MDM	4.46E-05	2.13E-05	0.03954	0.1
	Intercept	0.11575	0.05233	0.03	0.1
	Housing Density	-0.02159	0.02	0.284	
11.05.20					
Age group <18	Intercept	1.34E-01	1.85E-02	2.56E-10	<0.001
79 SOAs	MDM	-4.11E-05	3.46E-05	0.239	
	Intercept	-0.04444	0.07677	0.5644	
	Housing Density	0.06134	0.02932	0.0397	0.1
Age group 18-24	(Intercept	7.59E-02	1.21E-02	6.69E-09	<0.001
111 SOAs	MDM	-4.02E-05	2.59E-05	0.124	
	Intercept	0.064809	0.056327	0.252	
	Housing Density	-0.001829	0.021937	0.934	
Age group 25-34	(Intercept	5.48E-02	2.87E-03	<2e-16	<0.001
279 SOAs	MDM	-2.09E-05	5.76E-06	0.000353	<0.001
	Intercept	0.076144	0.01068	8.76E-12	<0.001
	Housing Density	-0.011859	0.004139	0.00448	0.001
Age group 35-49	(Intercept	1.30E-01	7.62E-03	<2e-16	<0.001
326 SOAs	MDM	1.38E-05	1.51E-05	0.359	
	Intercept	0.06974	0.02784	0.0127	0.1
	Housing Density	0.02536	0.01063	0.0176	0.1
Age group 50-64	(Intercept	1.30E-01	1.43E-02	3.91E-16	<0.001
158 SOAs	MDM	4.46E-05	2.68E-05	0.0979	0.05
	Intercept	0.06643	0.04796	0.168	

	Housing Density	0.03209	0.01809	0.0781	0.05
Age group >65	(Intercept	6.32E-02	1.64E-02	0.000369	<0.001
46 SOAs	MDM	2.26E-05	3.28E-05	0.495668	
	Intercept	0.13116	0.06819	0.0609	0.05
	Housing Density	-0.02254	0.02618	0.3939	

Table 5: Results of regression analysis (GLM with log link) with covariates MDM and population housing density for age standardised rates of self-reported COVID-19 symptom data (using COVIDCare NI) for two time periods ending 20.04.20 and 11.05.20). The table shows the number of SOAs with reported symptoms for each of the time periods. The dates correspond to the end date of 14 day symptom reporting sliding window.

Urbanisation

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 [26]). A statistically significant negative correlation is found between self-reported prevalence rates and overall MDMs and the deprivation domain of employment (*p* value 0.01 for overall MDMs and *p* value 0.001 for employment – 14 day time period ending 25 May 2020; Table 6). The findings for the urban area of Belfast are consistent with that for the NI and indicate that during lockdown restrictions, the most deprived SOAs with lowest ranking for employment were associated with higher population standardised prevalence rates of self-reported COVID-19 symptoms. 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. A statistically significant positive correlation is observed for the 14 day time period ending 1 June 2020 (*p* value <0.001; Table 6).

Date	COVIDCare NI	Estimate	Std. Error	Pr(>|t|)	Signif. Codes
20.04.20	Intercept	1.09	0.0273	<2e-16	<0.001
	MDM	-7.34E-05	5.91E-05	0.217	
	Intercept	1.093	0.02681	<2e-16	<0.001
	Employment	-9.19E-05	6.34E-05	0.149	
	Intercept	0.91574	0.1443	2.66E-09	<0.001
	Housing Density	0.06189	0.05975	0.302	
27.04.20	Intercept	3.235622	0.068707	<2e-16	<0.001
	MDM	-0.000381	0.000174	0.0302	0.01
	Intercept	3.243687	0.067851	<2e-16	<0.001
	Employment	-0.000454	0.000193	0.0201	0.01
	Intercept	2.4437	0.3699	7.34E-10	<0.001
	Housing Density	0.2792	0.1508	0.0661	0.05
04.05.20	Intercept	3.29013	0.073431	<2e-16	<0.001
	MDM	-0.000298	0.000172	0.0862	0.05
	Intercept	3.319996	0.071143	<2e-16	<0.001
	Employment	-0.000424	0.000189	0.0268	0.01
	Intercept	2.99708	0.41664	4.98E-11	<0.001
	Housing Density	0.07906	0.17333	0.649	
11.05.20	Intercept	3.172993	0.081825	<2e-16	<0.001
	MDM	-0.000126	0.000187	0.502	
	Intercept	3.196285	0.079154	<2e-16	<0.001
	Employment	-0.000211	0.000199	0.291	

	Intercept	2.2234	0.4151	4.25E-07	<0.001
	Housing Density	0.3786	0.1682	0.0262	0.01
18.05.20	Intercept	3.324225	0.086483	<2e-16	<0.001
	MDM	-0.000497	0.000237	0.0381	0.01
	Intercept	3.349214	0.083416	<2e-16	<0.001
	Employment	-0.000648	0.000262	0.0148	0.01
	Intercept	2.372	0.4996	6.34E-06	<0.001
	Housing Density	0.3339	0.2034	0.103	
25.05.20	Intercept	3.501257	0.087625	<2e-16	<0.001
	MDM	-0.000668	0.000263	0.0127	0.01
	Intercept	3.535837	0.084178	<2e-16	<0.001
	Employment	-0.000901	0.000302	0.00363	0.001
	Intercept	2.8554	0.5209	3.23E-07	<0.001
	Housing Density	0.188	0.2143	0.383	
01.06.20	Intercept	3.692048	0.08379	<2e-16	<0.001
	MDM	-0.000317	0.000211	0.136	
	Intercept	3.716563	0.082196	<2e-16	<0.001
	Employment	-0.00043	0.000231	0.0662	.
	Intercept	2.0701	0.4482	1.48E-05	<0.001
	Housing Density	0.6437	0.1822	0.000689	<0.001

Table 6: Regression analysis (GLM log link) results for Belfast Urban area for the covariates of overall MDMs, the deprivation domain of employment and population housing density using COVID-19 symptom mobile data provided by COVIDCare NI (20.04.20- 01.06.20) The dates correspond to the end date of 14 day symptom reporting sliding window.

Discussion

Social Deprivation

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 [34-36]. The pandemic has magnified the heterogeneity in society's health burden with a disproportionately higher impact on socially vulnerable communities, including older age groups, those with lower educational attainment and ethnic minorities [19-21]. These socio-economic inequalities are linked directly to area-level deprivation indices including income, education, employment, housing and environment, which contribute to greater risk of poor health [37-39]. The COVID-19 pandemic occurred against this existing backdrop of social and economic inequalities [40]. This research suggests an impact of social and economic disparities, in the form of social deprivation measures and housing density, on the prevalence of self-reported COVID-19 symptoms in the local community of Northern Ireland. The findings reveal that the highest self-reported prevalence rates of COVID-19 symptoms were found to be associated with the most deprived SOAs (lowest MDM rankings) and the most deprived SOAs with lowest ranking for employment for all SOAs. Our findings suggest the potential for COVID-19 to exacerbate socio-economic inequalities with higher prevalence of self-reported COVID-19 symptoms associated with social deprivation, housing density and age.

Weaknesses and limitations

This study deals with symptom reporting, which is a combination of (i) COVID-19-induced symptoms and (ii) 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: (i) having the requisite symptoms, (ii) the propensity to report symptoms, (iii) the likelihood to participate in one or other survey, (iv) ownership of a smartphone and (v) being part of the at-risk population.

Symptom-based surveillance and the use of self-reporting participants may, thus, introduce bias in the analysis. This 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 of generalisability of the predictors [41]. 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 socio-economic groups [42].

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 geographic spread of the self-reporting participants using the different smartphone apps was investigated through spatial mapping. The greatest geographic coverage was reported for the 14 day period ending 21 April 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 lockdown, when restrictions were more severe across all social deprivation categories.

The findings reveal differing relationships with the domains of MDM across the age groups. A larger proportion of self-reported prevalence rates of COVID-19 symptoms in the age groups 18-24 and 25-34 occurred in the most deprived SOAs. These age groups may not be the most at-risk groups for the consequences of infection, but may spread the virus through the community. In this research, census data for other confounders including adult obesity and respiratory disease were not available at SOA level. However, our findings indicate a relationship between self-reported prevalence rates in the community, social deprivation, high density housing and age. These findings provide evidence to inform government, and health and social care services, on the nature and magnitude of COVID health disparities and suggest potential targets for interventions to reduce health them.

Implications for socio-economic position and ethnicity

Northern Ireland is the least ethnically diverse region in the UK. In the 2011 census 1.8 % (32,400) of the resident population belonged to minority ethnic groups, but more than double the percentage in 2001 [28]. An analysis of census data shows that Asian, black and mixed households are generally larger than those of other ethnic groups [26]. These groups have younger age profiles than those of white ethnicity (87 % of residents with black ethnicity were aged under 45 years, compared with 61 % of whites in the last recorded census). These statistics are relevant to the findings of the current research on the association between self-reported prevalence of COVID-19 in the community and housing density. The 2011 Census showed that members of the black ethnic group (30 %) were three times more likely to live in overcrowded households than those of the white population (9.3 %), with rates for the remaining ethnic groups shown as Asian 24 %, other 19 % and mixed ethnic groups 13 % [28]. This research underlines the impact of the prevalence of COVID-19 in at-risk communities, related to existing inequalities due to socio-economic position, amongst others.

Conclusions

COVID-19 symptom prevalence estimates obtained from self-reporting COVID-19 smartphone data were regressed on a range of socio-economic variables in Northern Ireland. Significant associations were found between reported COVID-19 prevalence and both social deprivation and housing density for a range of age groups. The findings underline the link between health and place and the potential of the COVID-19 pandemic to exacerbate socio-economic inequalities. Specifically, the results 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 social deprivation and the lowest deprivation rankings for employment, particularly within the age group 18-34. This increased reporting rate may signal increased prevalence of the virus, which is likely to have a negative impact on at-risk communities. These findings, therefore, have the potential to inform longer-term public health policy for effective interventions to reduce health disparities and improve targeted control of COVID-19.

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 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. 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 de-identified 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 Limited 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.

Author Contributors

JMK conducted the initial literature searches, conducted the analysis and completed the initial drafts of the manuscript with input from all authors.

DC, NA, extracted the data and BJ, JMK and CG formatted the data.

UM, PMA and FK conducted literature searches.

UM, PMA, HVW, DB and FK reviewed the statistical methods.

All authors (JMK, UM, PMA, DC, NA, BJ, CG, HVW, DB and FK) read and approved the final manuscript.

References

1. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) Situation Report 52, 12th March 2020.
2. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* 2020;395(1022821)931-934.
3. Arenas A, Cota W, Gomez-Gardenes J, et al. A mathematical model for the spatiotemporal epidemic spreading of COVID19. 2020; MedRXiv https://doi.org/10.1101/2020.03.21.20040022
4. Ferguson et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College COVID-19 Response Team Report; 2020.
5. Metclaf, CJE., Morris, DH, Park SW. Mathematical models to guide pandemic response. *Science* 2020; 369(6502); 368-369. DOI: 10.1126/science.abd1668
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. DOI: https://doi.org/10.1016/S1473-3099(20)30243-7
7. Drew D, Nguyen LH, Steves CJ. et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020 https://science.sciencemag.org/content/early/2020/05/04/science.abc0473/tab-article-info accessed 27 August 2020)
8. Menni C, Valdes AM, Freidin MB et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020. https://doi.org/10.1038/s41591-020-0916-2
9. Chan AT, Brownstein JS, Putting the Public Back in Public Health — Surveying Symptoms of Covid-19. *The New England Journal of Medicine* 2020. DOI: 10.1056/NEJMp2016259
10. Jones KH, Ford DV, Jones C, Dsilva R, Thompson S, Brooks CJ, Heaven Ml, Thayer DS, McNerney CL, Lyons RA. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy protecting remote access system for health related research and evaluation. Journal of Biomedical Informatics: special issue on medical data privacy 2014.
11. Kirby R, Delmelle E, Eberth J. Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology* 2016;27;1-9. http://dx.doi.org/10.1016/j.annepidem.2016.12.001

12. NHS England. Definition of Health Inequalities https://www.england.nhs.uk/ accessed 02/12/2020
13. Kwan M-P. Geographies of Health, Disease and Well-being. Recent Advances in Theory and Method. 1st Edition ISBN 9780415870016. Routledge 2014.
14. Diez Roux AV. Investigating neighborhood and area effects on health. *American Journal of Public Health* 2001;91(11);1783-1789. DOI: 10.2105/AJPH.91.11.1783
15. Jones NL, Breen N, Das R, Farhat T, Palmer R, Cross-Cutting Themes to Advance the Science of Minority Health and Health Disparities. *AJPH* 2019;109(Suppl 1): No. S1
16. Duran DG, Pérez-Stable EJ. Novel Approaches to Advance Minority Health and Health Disparities Research *AJPH* 2019;109(Suppl 1):No. S1.
17. Breen N, Jackson JS, Wood F, Wong DWS, Zhang X et al. Translational Health Disparities Research in a Data-Rich World. Methods and Measurement Science. *AJPH* 2019;109(Suppl 1): No. S1.
18. Williamson E, Walker AJ, Bhaskaran KJ, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. *medRxiv* 2020.05.06.20092999; doi: https://doi.org/10.1101/2020.05.06.20092999
19. NHS Confederation. The impact of COVID-19 on BME communities and health and care staff, BME Leadership Network, Members Briefing April 2020, https://www.nhsconfed.org/resources/2020/04/the-impact-of-covid19-on-bme-communities-and-staff (accessed 09.06.2020).
20. Wadhera RK, Wadhera P, Gaba P, et al. Variation in COVID-19 hospitalizations and deaths across New York City boroughs. *JAMA*.2020;323(21):2192–2195. https://doi.org/10.1001/jama.2020.7197.
21. Yancy CW. COVID-19 and African Americans. JAMA. 2020;323(19):1891–1892. https://doi.org/10.1001/jama.2020.6548
22. Northern Ireland Statistics and Research Agency (NISRA). 2021 Census Northern Ireland Proposals Document [2019] P52.
23. Ofcom. Communications Market Report 2017 - Northern Ireland. Available online at https://www.ofcom.org.uk/__data/assets/pdf_file/0023/104927/northern-ireland-internet-online.pdf (accessed 25.08.20)
24. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ (accessed 25.08.20)
25. Public Health England Coronavirus (COVID-19) guidance and support. 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 Updated 22 May 2020. (accessed 24th May 2020)
26. Northern Ireland Statistics and Research Agency (NISRA). NI Multiple Deprivation Measures 2017 – Summary Booklet [2017] P28.
27. Layte R, Whelan CT. Explaining Social Class Inequalities in Smoking: The Role of Education, Self-Efficacy, and Deprivation. European Sociological Review, 2009;25(4):399–410, https://doi.org/10.1093/esr/jcn022
28. Northern Ireland Statistics and Research Agency (NISRA). Northern Ireland Census 2011 Key Statistics Summary Report [2014] P247.
29. Anselin L. Under the hood issues in the specification and interpretation of spatial regression models. *Agric Econ* 2002;27(3):247e67.
30. Pardo-Iguzquiza E, Dowd PA. Normality test for spatially correlated data. *Math Geol* 2004;36(6):659–681.
31. Anselin L. Spatial regression. In: The SAGE handbook of spatial analysis, 2009;1:255e76.
32. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships. West Sussex, England: John Wiley & Sons. 2003:268.
33. McKinley JM, Ofterdinger U, Young M, et al. Investigating local relationships between trace elements in soils and cancer data. *Spatial Statistics* 2013;5: 25-41. http://dx.doi.org/10.1016/j.spasta.2013.05.001
34. 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(3):317−325
35. Palaiologou P, Ager AA, Nielsen-Pincus M, Evers CR, Day MA. Social vulnerability to large wildfires in the western USA. *Landsc Urban Plan* 2019;189:99–116. https://doi.org/10.1016/j.landurbplan.2019.04.006.

36. Rufat S, Tate E, Burton CG, Maroof AS. Social vulnerability to floods: review of case studies and implications for measurement. *Int J Disaster Risk Reduct*. 2015;14(4):470–486. https://doi.org/10.1016/j.ijdrr.2015.09.013.
37. Gilbert KL, Ray R, Siddiqi A, et al. (2016). 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. https://doi.org/10.1146/annurev-publhealth-032315-021556.
38. Galea S. Health haves, health have nots, and heterogeneity in population health. *Lancet Public Health*: 2017:2(9):e388–e389. https://doi.org/10.1016/S2468-2667(17)30160-3.
39. Griffith DM, Jaeger EC, Sherman LD, Moore HJ. Patterns and causes of health inequities during a pivotal period in the life course. In: Griffith DM, Bruce MA, Thorpe RJ Jr, eds. Men's Health Equity: A Handbook. New York, NY : Routlege, 2019:72. https://doi.org/10.4324/9781315167428-6.
40. Bambra C, Riordan R, Ford J, et al. The COVID-19 pandemic and health inequalities. J Epidemiol Community Health 2020;0:1–5. doi:10.1136/jech-2020-214401
41. Griffith GJ, Morris TT, Tudball M, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity *medRxiv* 2020; preprint doi: https://doi.org/10.1101/2020.05.04.20090506
42. Kroenke K. Studying Symptoms: Sampling and Measurement Issues. *Annals of Intern Med* 2001 May 1;134(9 Pt 2):844-53. doi: 10.7326/0003-4819-134-9_part_2-200105011-00008.

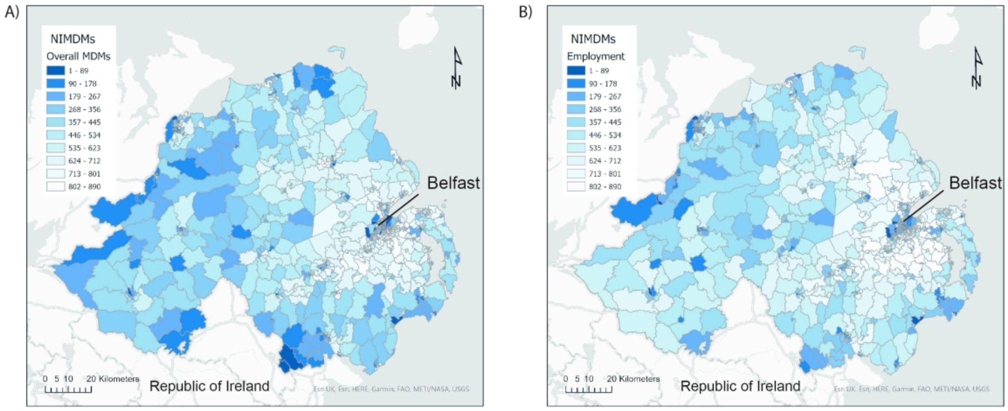

Figure 1: Northern Ireland Multiple Deprivation Measures 2017 (NIMDM) provided by the Northern Ireland Statistics and Research Agency (NISRA) [26] including information on A) an overall MDM ranking; and B) individual deprivation domain for employment. A low ranking is the most deprived.

244x98mm (300 x 300 DPI)

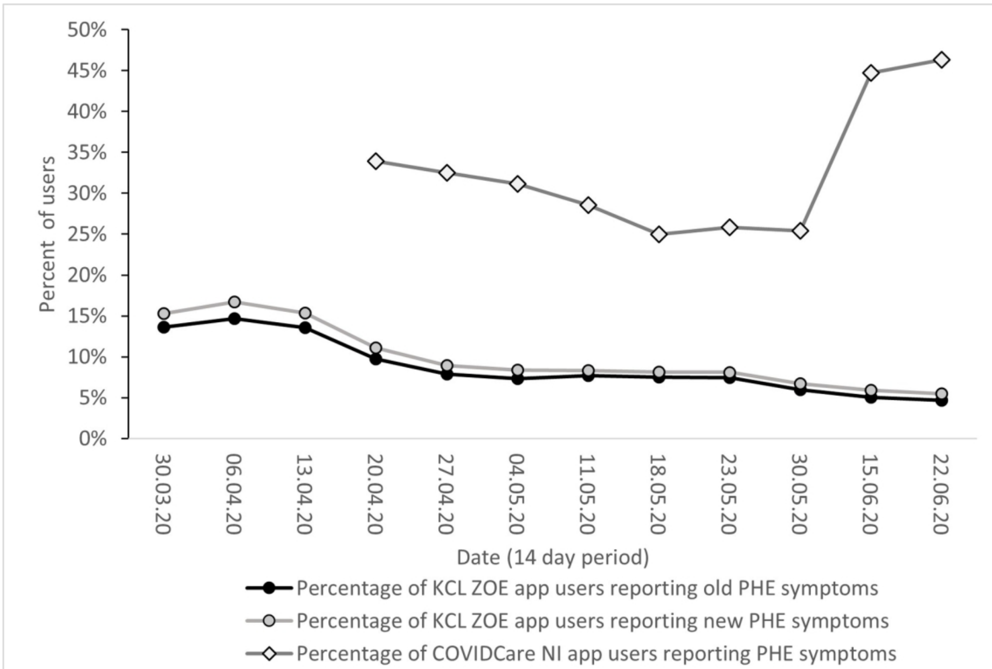

Figure 2: Comparison of percentage of users self-reporting COVID-19 symptom data (using data from Table 1) as provided by the KCL ZOE symptom tracker app data for NI (reporting period 24 March – 22 June 2020) and COVIDCare NI Symptom checker feature, (reporting period 6 April to 29 June 2020). The dates correspond to the end date of 14 day symptom reporting sliding window.

152x102mm (300 x 300 DPI)

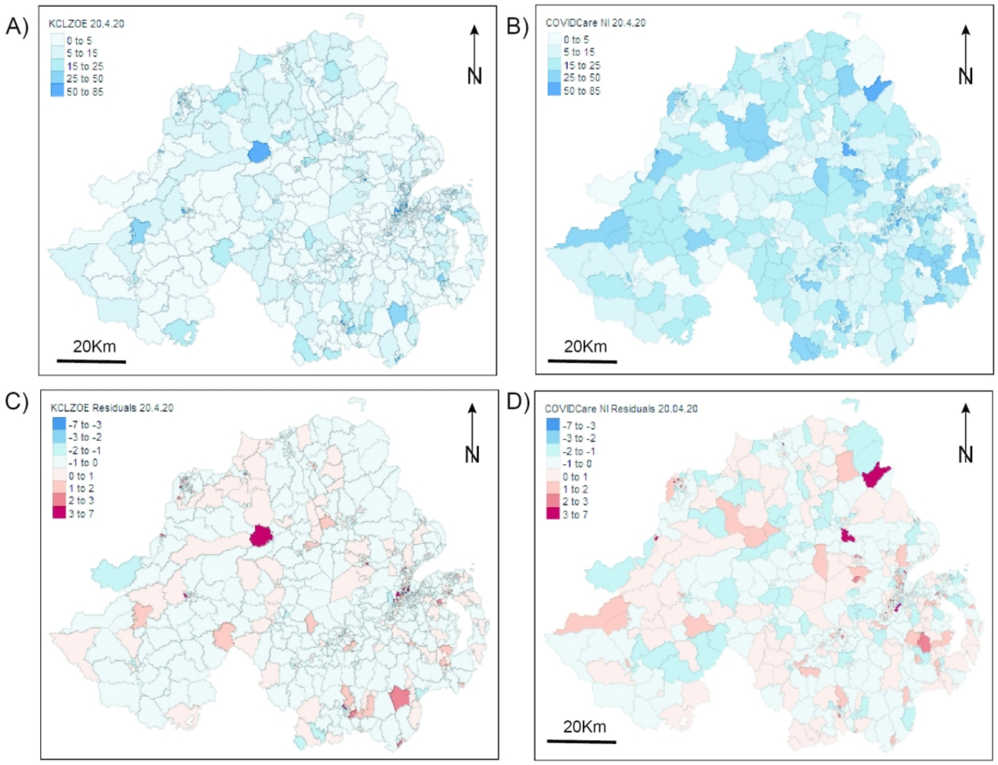

Figure 3: Maps of COVID-19 symptom mobile data for reporting period ending 20.04.20, provided by two sources: A) KCL ZOE symptom tracker app data for new PHE symptoms for Northern Ireland (reported symptoms in 592 SOAs) and B) COVIDCare NI Symptom checker feature (reported symptoms in 758 SOAs). The date 20 April corresponds to the end date of 14 day symptom reporting sliding window. Self-reported prevalence rates are standardised for 100,000 population. Residuals computed from the GLM regression model versus MDM for C) KCL ZOE app data and D) COVIDCare NI data.

212x163mm (300 x 300 DPI)

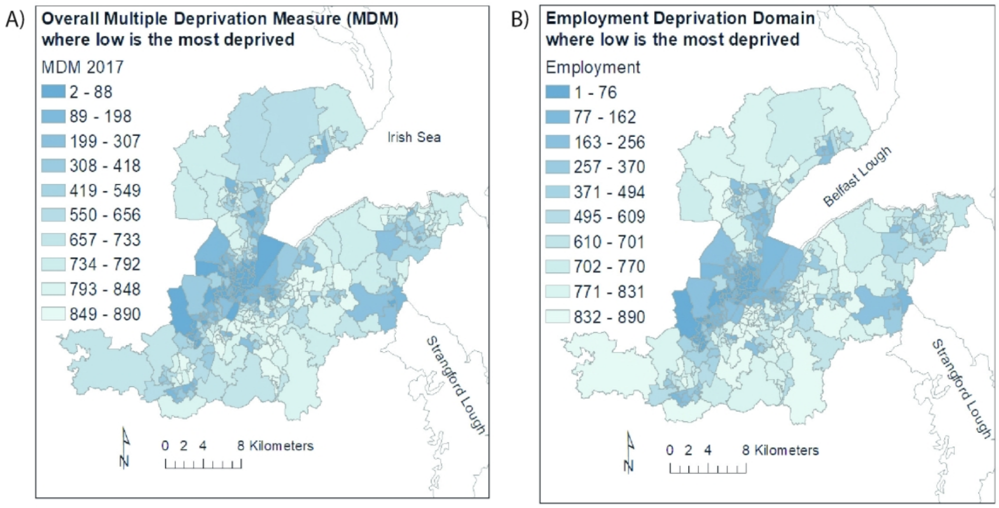

Figure 4 : Northern Ireland Multiple Deprivation Measures 2017 (NIMDM) provided by the Northern Ireland Statistics and Research Agency (NISRA) [26] including information for Belfast urban area on A) an overall MDM ranking; and B) individual deprivation domain for employment.

181x93mm (300 x 300 DPI)