

©2020. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/about/downloads>



The version of record of Walford, Nigel Stephen (2020) Demographic and social context of deaths during the 1854 cholera outbreak in Soho, London : a reappraisal of Dr John Snow's investigation. *Health & Place*, 65, p. 102402. can be found at <https://doi.org/10.1016/j.healthplace.2020.102402>

Demographic and social context of deaths during the 1854 cholera outbreak in Soho, London: a reappraisal of Dr John Snow's investigation

Nigel Stephen Walford

Department of Geography, Geology and the Environment,
School of Engineering and the Environment,
Kingston University,
Penrhyn Road,
Kingston upon Thames. KT1 2EE
UK

Tel. +44(0)20 8417 2512

Email. N.Walford@kingston.ac.uk

<http://orcid.org/0000-0002-9182-0252>

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

Deaths from cholera in Soho, London (August to September 1854) exposed the epidemiology of the disease and demonstrated applied geospatial analysis by highlighting the shortest path principle followed by local residents when they obtained drinking water from a contaminated pump. The present investigation explores if households and individuals with different demographic and socio-economic characteristics were more or less likely to obtain their water from the pump and succumb to the disease. It combines information from the 1851 Population Census and topographic databases with the digital deaths and water pump data to reveal the risk of exposure and the mortality rate were greater for certain occupations, age groups and people living at high residential density irrespective of proximity to the contaminated water pump.

Key words

Historical GIS, cholera, kernel density estimation, spatial analysis

Introduction

There can be few students or researchers of epidemiology or Geographical Information Systems and Science (GISSc) who have not come across the example of Dr John Snow's investigation of 578 deaths from cholera in Soho central London in August/September 1854. It is often regarded as one the earliest recorded instances of geospatial data analysis revealing underlying processes (e.g. Longley, et al., 2005: 317-319). However, the main focus of John Snow's inquiry was to discover the means whereby the disease was transferred: it is an example par excellence of the value of epidemiology in public health and demonstrates how analysis of spatial patterns can help to reveal and understand the operation of underlying physical processes. The prevailing view was that cholera was caught by inhaling putrid air, whereas John Snow believed it was the result of drinking contaminated water. His investigation confirmed the latter hypothesis

“demonstrating the water-borne origin of cholera” (Gilbert, 1958: 174) and made a major contribution to the pathology of the disease. Reassessing Snow’s work in conjunction with other historical data sources connects with contemporary debates about the role of environment, deprivation and neighbourhood characteristics on health, morbidity and mortality (e.g. Timmermans et al., 2020), urban sanitation in developing countries where cholera and similar diseases are a continuing threat to public health (Perez-Heydrich et al., 2013) and other instances of where lessons from the nineteenth century are relevant in this context (Konteh, 2009).

Self-evidently Snow was working at a time when notions of what might comprise GIScience were a long way in the future. Geography itself was in the early stages of becoming an established academic discipline in universities and was still very much viewed as an endeavour associated with expeditionary discovery and classification. Geographers’ interest in Snow’s work was quiescent for many years until the 1950s (May, 1958; Gilbert, 1958; and Stamp, 1964), although epidemiologists continued to refer to his work (Sedgwick, 1902; Frost, 1936). Interest in Snow’s investigation and the data generated from it has increased over recent decades, which has resulted in the capture of the geospatial data representing the point locations of the deaths and the water pumps used by residents in the neighbourhood and applying spatial and statistical techniques. These analyses have demonstrated how such procedures could be applied in order to reach similar conclusions to those in the original investigation by means of geocomputation rather than field experiment and to illustrate the potential of GIScience for contemporary epidemiological applications.

The essence of Snow’s investigation showed that the 1854 cholera deaths were clustered around one of the water pumps in Broad Street and that the intensity of deaths decreased with distance from it. In other words, it exemplifies the operation of a distance decay function. Most of the deaths occurred in St. James parish with a significant minority in St. Anne’s parish and one in St. George’s Hannover Square, all located in the Borough of Westminster. He used a manual process to draw areas (polygons) around the six water pumps within St. James parish (there were another five pumps just over its boundary in neighbouring wards) representing their catchment areas and thereby he revealed the spatial pattern. He hypothesised a link between cholera and contaminated water and sought to confirm this by removing the handle from the Broad Street pump, which resulted in a decrease in the rate at which people succumbed to the disease. Snow was not without detractors (Parkes, 1855), but with time the passage of time his argument prevailed. Recent re-analysis of geospatial data relating to the addresses of the deaths and the pump has focused on showing how a similar outcome could be achieved by applying spatial analytic techniques in a modern GIS framework (Koch, 2005). Shiode et al. (2015) extended such investigation by using data collected during visits to households in the area in order to estimate the total population at risk and to chart the spatio-temporal progression of the disease. However, although some details of people and households living at the addresses where the deaths occurred was collected during the contemporary inquiry into the outbreak, this population at risk seems to have been regarded as undifferentiated in terms of its socio-economic and demographic characteristics.

The mortality rate and space-time pattern revealed by the cholera data have been examined (Shiode et al. (2015), but a further aspect of the 1854 cholera outbreak remains to be explored: namely the extent to which there were differences in the socio-demographic background and living conditions of households and streets

with and without people who succumbed to the cholera pathogen. Exploration of this aspect will address a pertinent public health issue, namely whether some diseases differentially impact some sections of society, an issue that has been thrown into sharp focus by the COVID-19 pandemic of 2020. This article aims to address these issues by moving on from simply regarding the population at risk as an undifferentiated collection of individuals who had the misfortune to live in the vicinity of a contaminated water supply and obtained their drinking water from this source. It incorporates household and individual level data from the 1851 Population Census for all inhabited addresses in Snow's study area to characterise the residents taking into account changes in the residential population between 'census night' (30/31 March 1851) and the start of the cholera outbreak in Soho on 26 July 1854. Combining Snow's cholera data with 1851 household and individual census data adds to our understanding of the higher mortality rates and greater risk of the disease on certain social and demographic groups. The remainder of the paper is divided into four sections with the next two reviewing the development of Historical GIS (HGIS) and exploring the data and methods in the present analysis. These are followed by a presentation of the analytical results and discussion of their implications.

Historical GIS and the 1854 cholera outbreak

Significant obstacles must be navigated in order to unlock the spatial context of the past, but successful HGIS research can serve to enhance and challenge the historical narrative, thereby having the potential to prompt "questions that might otherwise go unasked" (White, 2010: 36). Researchers using modern data sources may benefit from their existence in a digital form and including standard georeferencing, in contrast "extracting geographical data from historical sources is analogous data mining ... with a pickaxe and shovel" (Knowles, 2008: 13), where these advantages may be absent. Flexible and innovative visualisation is one of the key features of GIS, but this must play 'second fiddle' in HGIS research until the essential but often laborious task of converting analogue data into a digital format has been completed. Many historical data sources are potentially but not inherently spatial and their geographical component has to be coaxed from them before visualisation and sophisticated analysis can be undertaken (Gregory and Geddes, 2014). There have been considerable advances in the digitising of printed text in recent years with three methods now dominating: scanning to create an indexed PDF image, optical character recognition and transcription (Hitchcock, 2013).

The fundamental principle that data held in a GIS has two key elements, geometry and attributes, applies as much to historical as contemporary sources. Lessons learnt from the early years of GIS development in the modern era relating to the capture of analogue data now also apply in respect of digitising historical sources, although with added complications (Knutzen, 2014). Historical data sources relating to human populations may contrast with contemporary ones by allowing the specific locations of people and the 'things' (e.g. houses, farms and workplaces) with which they were associated to be digitised, whereas modern sources usually hold people's data in aggregate form with information relating to them contained within discrete and often arbitrary boundaries in respect of where they live, work or otherwise spend their time (Gregory and Ell, 2005). Several researchers have shown that analysis of thematic variables for areas across a range of granularities can produce variation and complexity in the historical narrative (Gregory and Cooper, 2013; Fotheringham et al., 2013) and therefore disaggregated data records are preferable.

The potential advantage of building a GIS with historical data using disaggregated data for individuals, households and addresses introduces the possibility of working with point level geometry by digitising the XY coordinates of these phenomena and to analyze these as points or as clusters of entities on streets and thereby associate people with the places where they lived their lives. However, obtaining such digital coordinates is challenging when streets may have changed their name, buildings may have been demolished or destroyed and redevelopment may have obliterated the physical fabric of the past. A mixture of approaches may yield greatest success in achieving the aim of attaching accurate coordinates to historical records (Hitchcock et al., 2015; Navickas, 2016), although it is likely that a degree of uncertainty will remain over the extent to which completely accurate georeferencing has been achieved (Plewe, 2003). There is a wealth of historical data sources available in Great Britain, leading Hitchcock and Shoemaker (2014: 75) to describe it as the “most digitised when and where in the world”. The Great Britain Historical GIS Project (Gregory et al., 2002) had already addressed the challenges of ‘geo-enabling’ many of these sources some 15 years earlier for the purposes of visualisation and interrogation (e.g. Southall, 2003, 2006, 2014). However, their use for analytical purposes emerged more slowly but has accelerated in recent years (see for example Shiode et al., 2015; Brown, 2016).

The 1854 cholera outbreak investigated by Dr John Snow offers the opportunity of moving beyond visualisation to explore the relevant historical data sources using spatial analytic techniques. One of the key conclusions of Snow’s research (Snow, 1849) was to challenge the prevailing view that cholera is an airborne pathogen. He demonstrated, by arranging for the handle from the Broad Street water pump to be removed, that a reduction in the number of new cholera cases occurred and from this he reached the conclusion that people caught the disease as a result of consuming contaminated water. Further investigation revealed that the supply of drinking water to this pump had become tainted by sewerage in groundwater associated with the use of the Thames and other rivers for sewerage disposal. The documents published in the immediate aftermath of the cholera outbreak and in the following year explored a range of environmental characteristics prevailing at the time focusing in particular on weather conditions, elevation, population density, age and sex (General Board of Health, 1855).

Subsequent GIS-based research focusing on the cartographic aspects of Snow’s map or spatial analysis of the pattern depicted on it has started with the documents and maps published by Dr Snow (Cliff and Haggett, 1988; Koch, 2011; Nakaya, 2001). Educational interest in using the spatial data contained in Snow’s map resulted in the creation of a dataset from these sources containing the latitude and longitude coordinates of each of the 578 deaths and the water pumps (Tobler, 1994). These ‘georeferenced deaths’ have formed a starting point for applying different types of spatial analytic technique that can be used with point feature data with the aim of exploring the area influenced by the Broad Street Pump and the point pattern of deaths surrounding it. These include the generation of Thiessen or voronoi polygons around the sites of the water pumps in the area (Koch, 2011) and use of network analysis in respect of pedestrian travel routes from the residential addresses where the deaths occurred to the nearest pumps (Cliff and Haggett, 1988).

Data and Methods

The present analysis extends previous work by adding data from the 1851 Population Census and contemporaneous topographic maps to previously analysed geospatial data from the plan showing where 578 cholera deaths occurred, the records of the house-to-house visits, the General Registrar Office's returns of deaths for weeks coincident with the outbreak and the Ordnance Survey map of parts of the parishes of St James and St Anne Westminster. Shiode et al (2015) matched information about cholera events in these sources and identified a further 45 deaths; the present analysis is based on this total of 623 deaths. Despite these researchers populating each house and street with the number of the residents at the time of the outbreak according to the visitation records, thus far no attempt has been made to link the archived records of the 1851 Population Census or the physical areas of the residential buildings to the cholera deaths data in order to examine the population's socio-demographic characteristics and residential occupation density. In order to create an internally consistent geospatial database that included these additional sources it was necessary to re-digitise some previously captured data. In summary the analysis presented here, carried out in ArcGIS 10.6, was based on the following:

- 1851 Population Census records as images of original documents and from the Integrated Census Microdata for addresses within Snow's study area
- Documents held in the freely accessible John Snow Archive and Research Companion (nd) and the Wellcome Foundation Archive published by Dr John Snow and others in respect of the inquiry into the 1854 cholera outbreak and other associated historical records
- Ordnance Survey historical topographic mapping from Digimap (Edina Historic Digimap Service)
- Polygon shapefile of building 'footprints' digitised from the topographic maps
- Polygon shapefiles of Snow's study area and the parishes of St Anne Westminster, St George Hannover Square and St James Westminster also digitised from the topographic maps
- Point shapefiles of addresses within Snow's study area with census data appended as attribute fields
- Polyline shapefile of thoroughfares (streets, alleys and yards) with census data added as attribute fields

The following subsections review these historical data sources and how they were assembled into a database.

1851 Population Census

Higgs et al (2013: 11) argue that British censuses of the mid-nineteenth century should be viewed in the context of developments in medical and epidemiological research pointing out that the main 'architect' of these enumerations, William Farr, the General Registrar Office's Superintendent of Statistics at the time, came from a medical background (Eyler, 1979). Farr's work using data obtained from death registration certificates to explore the progress of epidemic diseases and his strongly held opinion, in common with others of his time, that "human effluent in large cities" was the main source of noxious chemicals entering the body and causing disease (Higgs et al., 2013: 12), has close correspondence with Snow's own research on the transmission of cholera. This focus on generating data potentially of use in combating the spread of diseases linked to rapid urbanisation helps to account for the expanding scope of the nineteenth century British census (Higgs, 1991).

Three years before the 1854 cholera outbreak in Soho the British government had conducted the 1851 Population Census, “undoubtedly the most ambitious decennial enumeration of the Victorian period” (Higgs, et al., 2013: 24). It recorded a more extensive and detailed set of information than the preceding 1841 enumeration including each person present on census night (30/31 March 1851) in separate households at a residential address and assigned a numerical schedule number to each household. The data recorded on the householder’s schedule, apart from the address itself (street with building name or number), related to a person’s name, gender, age, marital condition, occupation, year and place of birth. Although the scope of this information is considerably less than is collected in a modern population census, it provided a basis for subsequent census enumerations through to the early twentieth century (Higgs et al., 2013). The 1851 census was also more comprehensive in comparison with previous enumerations to the extent that it sought to regulate the enumeration not only of people living in households and institutions, but also those aboard vessels on coastal, estuarine and inland water, travelling, undertaking night-working or otherwise away from home on census night (Higgs, 1989). The cholera outbreak in St. James and to a much lesser extent St. Anne’s parishes in the Soho area of Westminster started on 26 July and continued until 01 October 1854. Despite the 1851 Population Census having been held some three years and four months earlier, it is reasonable to argue that the census records combined with other data sources, such as information about the victims’ occupations and ages in the visitation survey, provide a reliable basis for examining the characteristics of the households and individuals that were or were not afflicted by the disease. The analysis carried out here used the 1851 census data from two sources: the original records from The National Archive available in partnerships with FindMyPast (part of BrightSolid) and from the Integrated Census Microdata (ICeM) project (Higgs et al., 2013; Schürer and Higgs, 2014)

Cholera Death and Related Data

One of the main items on the Snow Archive is a map or plan showing the building frontages with their address number on the streets in the study area (Snow, 1855; Wellcome Library, nd). The addresses where cholera deaths occurred were digitised in conjunction with the topographic maps (see below). The house-to-house inquiry was carried out “especially and primarily in the streets which [had] suffered most” (General Board of Health, 1855: 138) and included addresses where there had been a cholera death or case reported. The 1851 Census addresses were cross-referenced with the house-to-house visitation records, although the latter only accounted for 343 of the 2,409 addresses in the census records. The 623 cholera deaths used in the current analysis were determined from the ‘plan of deaths’ in conjunction with the Weekly Returns and the Appendix of the Report into the outbreak (see above).

Topographic mapping

The Ordnance Survey embarked on its national land survey in the 1840s and by the mid-1850s topographic maps at a scale of six inches to the mile (1:10,560) had started to be published. The Landmark Information Group with cooperation from the OS embarked on a programme to scan these maps at 300 dpi and to georeference the images to the British National Grid in 1995. Subsequently a seamless digital mosaic of these map tiles was created and made available to the UK higher education community for research and teaching purposes by Edina at the University of Edinburgh (<https://digimap.edina.ac.uk/historic>). The OS maps show details of building outlines, which are only partially evident on the scanned image of Snow’s plan of the cholera deaths.

Data capture and assembly

The map included in the report of the inquiry into the 1854 cholera outbreak shows an area within which the cholera deaths occurred overlapping the St. James and St. Anne parishes defined by a boundary running for the most part along the centre line of roads: in their entirety St James extended south and west, and St. Anne east of this area. Three main sets of geographical entities have been captured for inclusion in the geospatial database covering this area: point features representing the addresses where deaths occurred along the alleys, mews and streets, and the location of 13 water pumps; line features for these thoroughfares where residences were present within the study area (i.e. all thoroughfares where there were residential addresses according to the 1851 Census); and polygons of the building footprints of all residential addresses and of the enclosing study area boundary. Figure 1 shows this assembly of geospatial data used in the present analysis correcting for the discrepancy just noted between some of the previously digitized data sources.

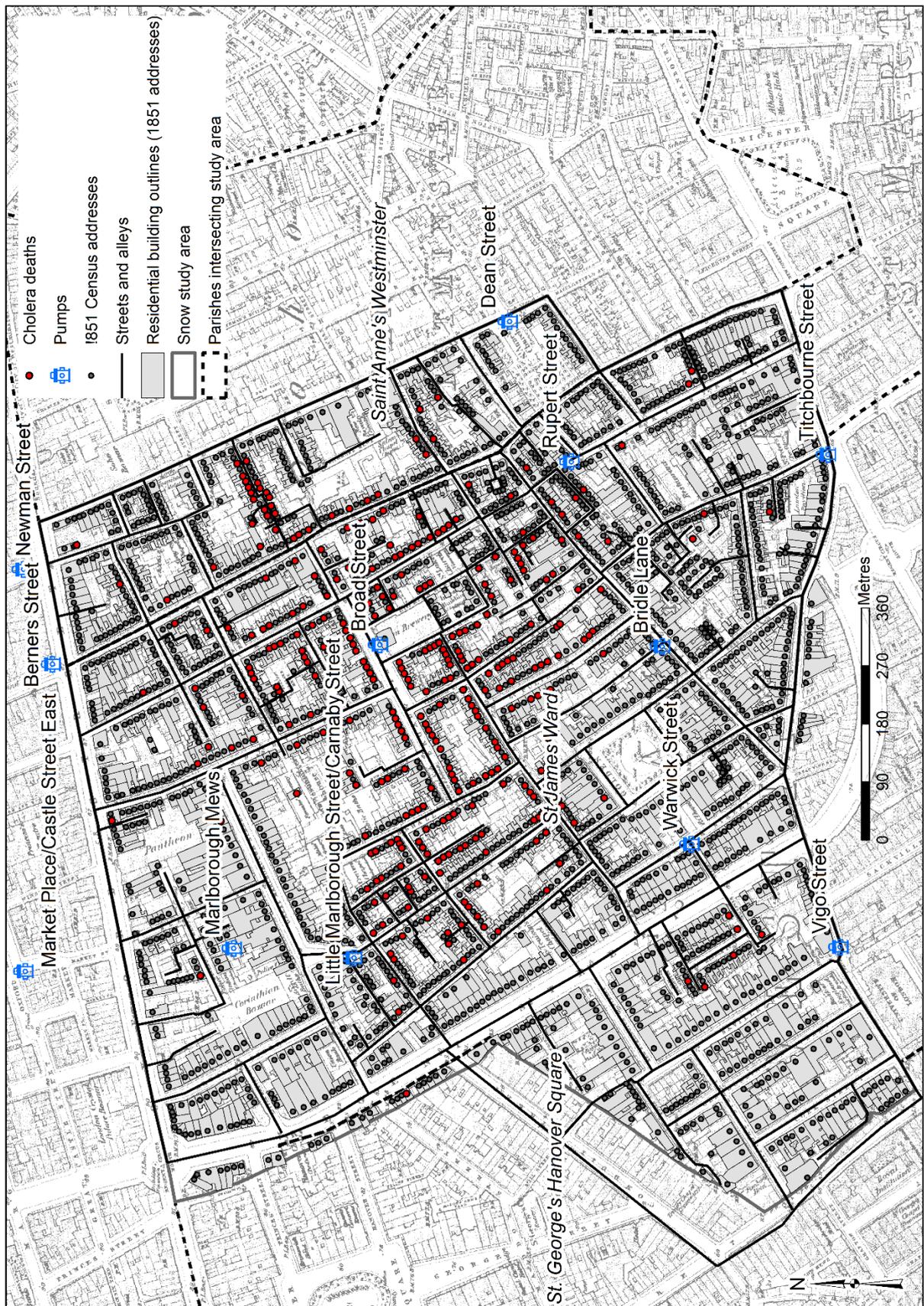


Figure 1: 1851 Census addresses and building outlines and deaths from 1854 Cholera outbreak in Soho, Central London.

Source: 1:2500 County Series 1st Edition [TIFF geospatial data], Scale 1:2500, Tiles: lond-tq2880-1,lond-tq2881-1,lond-tq2980-1,lond-tq2981-1, Updated: 30 November 2010, Historic, Using: EDINA Historic Digimap Service, <<https://digimap.edina.ac.uk>>, Downloaded: 2019-11-29 13:40:24.579

The 1851 Census data comprise a detailed record of who was present at each address on census night, which can be linked to the house-to-house visitation records to examine changes over the three years and four months between these events. Although the open access version of the ICeM does not enable users to examine the geographical location of individuals and households, combining the various data sources enables a link to their residential address to be made. For example, by using address numbers or property names shown on Snow's map of the cholera deaths in conjunction with the historical OS topographic mapping, it was possible by extension of techniques used in other research with historical census records (Walford, 2019) to geocode all addresses recorded as occupied in the 1851 Census. In addition, information contained in the visitation reports was added to the database. Addition of the household and individual attribute data from the 1851 Census represents a potentially important extension to previous investigations insofar as it allows the overall socio-demographic characteristics of the study area, individual streets and residential addresses to be compared with those where cholera-related deaths occurred. The area measurement of building footprints does not take into account the number of floors that were occupied residentially. It might have been possible to estimate the number of floors occupied in those streets included in the visitation survey, but this would have omitted a substantial number of streets and addresses across the entire study area. Using the simpler area measure (building outline) means that the underestimation of residentially occupied space has been applied consistently to all persons and households.

Connecting the house-to-house visitation records with the census data collected just over three years earlier offers a way of checking if the 1851 census statistics are a reasonable indicator of the socio-economic and demographic conditions in 1854. The visits were only carried out in selected streets, at some addresses only in relation to households where there had been case of cholera, and the records provide an incomplete and to some degree inconsistent account. For example, addresses on some streets record details of residents' occupations, whereas for others this level of detail is absent or imprecise. Nevertheless, apart from information relating to the sanitary conditions and disease, these records include, to varying extents, the number of rooms, inmates (residents), with males, females and children shown separately in some cases, the number of persons per floor, the number of cases and deaths from cholera and diarrhea. The occupation of some residents is also recorded, although this information may be somewhat imprecise using phrases such as 'tailor, etc.' or 'basketmakers etc.', whereas other entries were more specific (e.g. 'egg merchant' and 'cheesemonger'), although in both cases the terms are rarely associated with specific individuals. However, comparison of the textual details in both sources has enabled some assessment of continuity by recording whether at least one occupation occurred at an address in both the 1851 census and visitation survey. In some cases, where there is only one household present at an address and an uncommon occupation (e.g. 'surgeon' or 'gold lace shopman') is recorded in both data sources, it is feasible to conclude that some of the same individuals were present in 1851 and 1854. Using this occupation information, addresses have been categorized dichotomously on this basis as including at least some individuals who were present at both times.

Notwithstanding these challenges and with a gap of just over three years, it is unlikely that there would be exact matches between paired counts or values of variables for addresses in the 1851 Census and the 1854 house-to-house survey, nevertheless it seems realistic to argue that the 1851 Census data would provide a good overall basis for describing the demographic and socio-economic character of the study area. Table 1 shows the mean difference between comparable variables according to the 1851 Census and the 1854 visitation survey for those addresses where at least one cholera death occurred. The N values differ because of inconsistency in the house-to-house survey records: for example, the numbers of males and females were separately recorded for 47 surveyed addresses, whereas the number of persons was given for 167. Having applied the Kolmogorov-Smirnov and Shapiro-Wilk tests, there is no reason for thinking the data are significantly differently from normality (apart from the difference in count of females and persons per room). The T-tests results in Table 1 examine whether the difference is significantly different from zero, which would pertain if the data values per address were the same in 1851 and 1854. The test results produce a rather mixed outcome at the 0.01 significance level with some variables showing a significant difference between the two years and others the opposite. Persons per room and m² per person had P values less > 0.01 indicating they were not significantly different from zero, whereas the numbers of persons, males, females and children produced the opposite outcome. Noting these contrasts in the subsequent analyses, it seems plausible to argue that the 1851 Census data geocoded to the study area addresses provides an acceptable account of the spatial variation in its the demographic and socio-economic characteristics.

	N	Mean of difference in count	Std. of mean	t	p
Persons at address	167	3.9	0.64	6.03	0.00
Males at address	47	5.1	4.39	7.98	0.00
Females at address	47	3.5	5.62	4.31	0.00
Persons per room	128	-0.1	0.09	-1.39	0.17
Children (under 16 years)	46	3.2	4.43	5.56	0.00
M ² per person	45	-1.8	0.83	-2.19	0.03

Table 1: T test results for difference in mean of comparable variables in 1851 and 1854 for addresses with recorded cholera deaths.

Note: Differences in the N values arises from variability in demographic data recorded in the 1854 visitation records.

Sources: 1851 Population Census; General Board of Health 1855.

The following section presents results from analysis of this interconnected assemblage of datasets, firstly exploring the overall demographic and socio-economic character of the area and then examining how these characteristics varied from street-to-street and address-to-address. The second part of the results section extends and develops on Shiode et al's. (2015) visualisation and analysis of the spatial variation in mortality from cholera in two ways: first, using the 1851 Census data to obtain an address level mortality rate; and second to explore variations in this rate according to predominant occupation type, age structure and residential density at addresses. In both cases kernel density estimation (KDE) was used to model the

spatial distribution of mortality rate. KDE produces a density surface from recorded data values for a series of points (here number of cholera deaths or persons working domestic service per address). Although density surfaces can be estimated by other means, KDE has an established track record and may be used in effectively in an exploratory fashion (Shi, 2010).

Results

Demographic and socio-economic characteristics of Snow's study area

The study area, as defined originally in Snow's investigation, comprised 397.1 hectares and contained 6,740 households and 31,596 individuals at the time of the 1851 Census giving a high level of population density overall at 79.6 persons per hectare. In addition to the 2,413 inhabited addresses in the 1851 census, there were another 10 at which a cholera death occurred in the 1854 outbreak, but were recorded as uninhabited in 1851. Table 2 provides a summary of selected demographic and socio-economic characteristics of the population across the study area as a whole. Most addresses were inhabited by between 6 and 14 individuals and just under 50 per cent had just one household, although the distribution is fairly skewed with 10 per cent having 7 or more households. Similar percentages of households had 2-3 or 4-5 residents but over 20 per cent had 6-9 and 10 per cent just one person. Approximately a third of the population were children aged under 16 years and a slightly lower percentage were adults aged 25-39 years. Over 60 per cent of households comprised a married couple or widowed person with or without never married children. Households with extended families that included siblings, parents, grandchildren or nephews/nieces of the household head accounted for over 12 per cent of the total. There were 43 'household units' comprising residents in an institution, the largest of which was St James Workhouse in Poland Street with 628 residents, which has been excluded from some of the later analysis where the number of deaths, occupants and the extent of the building's 'footprint' would have distorted residential density calculations and results. Apart from the 4 per cent of the population who were born outside the UK including Ireland, the remainder was almost equally split between those born in London or Middlesex and those in other parts of the UK. The final row of Table 2 shows the top six occupations in which people were working, these occupations are Level 1 of the Historical International Standard Classification of Occupations (HISCO) (Leeuwen, 2002; Higgs et al., 2013). Occupations connected with dress (clothing) were the most common with nearly a third of people recording an occupation engaged in this sector. Domestic service came second with 22 per cent of the total. The other four occupations shown accounted for at least 5 per cent of recorded occupations, but it is noteworthy that professional occupations featured as the fifth most significant group.

Persons at address	1	2-3	4-5	6-9	10-14	15-19	20 and over
Whole study area	1.5	7.7	11.4	25.5	20.1	13.0	20.7
Streets with at least one death	1.1	5.5	9.2	23.1	19.5	14.8	26.7
Streets without any deaths	2.3	11.7	15.6	30.0	21.4	9.7	9.4

Households at address	1	2	3	4	5	6	7 and over
Whole study area	47.8	13.4	8.3	8.1	6.5	5.7	9.9
Streets with at least one death	39.3	13.2	9.4	9.7	8.3	7.4	12.7
Streets without any deaths	64.4	13.7	6.2	5.0	3.1	2.6	4.9
Persons at household	1	2-3	4-5	6-9	10-14	15-19	20 and over
Whole study area	9.6	26.9	25.4	21.4	4.9	1.2	0.1
Streets with at least one death	9.3	38.3	26.2	20.1	4.2	1.2	0.6
Streets without any deaths	10.4	32.7	22.6	25.2	7.0	1.3	0.8
Age (persons)	0-15	16-24	25-39	40-54	55-64	65 and over	
Whole study area	29.3	19.0	28.3	15.4	5.0	3.0	
Streets with at least one death	28.3	19.0	28.4	15.8	5.2	3.3	
Streets without any deaths	24.7	21.0	30.8	15.9	4.7	2.8	
Place of Birth (persons)	London or Middlesex	Other UK	Overseas				
Whole study area	48.8	47.2	4.1				
Streets with at least one death	49.5	47.1	3.4				
Streets without any deaths	46.5	47.2	6.1				
Household unit type	Solitary	Unrelated group	Married without children	Married with children	Generationally extended from head	Multiple groups	Institution
Whole study area	3.8	3.7	11.2	52.6	12.4	4.5	0.8
Streets with at least one death	3.8	3.3	11.4	53.3	6.4	4.3	0.8

Streets without any deaths	4.0	5.1	10.6	50.1	7.2	4.1	0.9
Occupation (persons)	Working in dress	Domestic	Conveying and transport	Food	Professional occupations in metals	Working in metals	
Whole study area	29.0	20.8	7.0	6.8	5.0	4.8	
Streets with at least one death	7.0	18.7	1.9	5.4	4.1	2.3	
Streets without any deaths	6.2	26.1	2.9	3.3	7.5	1.8	

Table 2: Selected characteristics of population and households in Snow's study area in 1851 (percentages). Notes: For definitions of household unit type see Higgs et al. 2013. Figures for separate indicators exclude missing data. General and undefined occupations were excluded from analysis. Whole study area refers to the area covered by Snow's report (see Figure 1).

Source: 1851 Population Census.

Figure 2 highlights address-level variation by indicating the frequency distributions of some of the characteristics tabulated in table 2. Making a simple visual comparison of the characteristics represented in Figure 2 with the distribution of cholera deaths shown in Figure 1 starts to provide some initial evidence for a connection between the occurrence of a death at an address and the social and demographic features of the people living there. There is evidence of higher population density in the central area where the majority of deaths occurred, with a noteworthy marginal case towards the south-west of the study area in Heddon Street and Heddon Court where a minor cluster of addresses with at least one death were located. Focusing on the three types of occupation that accounted for the highest numbers of deaths where this was recorded, reveals that people employed in the manufacture and sale of dress (clothing) (23.4 per cent of deaths) were also concentrated in the central area. Persons employed in building and construction (9.2 per cent of deaths) were more widely dispersed, whereas people working in domestic service (6.6 per cent of deaths) to a notable extent lived at addresses some distance from the central area. These results showing the uneven distribution of people in certain types of occupation suggest that further investigation of the spatial variation in the cholera mortality according to these social characteristics is worthwhile.



Figure 2: Spatial distribution of selected demographic and socio-economic indicators from 1851 Census across Snow's study area.

Source: 1851 Population Census.

Spatial variation in cholera mortality rate across Snow's study area

Before examining differences in mortality between addresses where certain types of occupation were present and where there were variations in age structure and residential density, the undifferentiated mortality rate for addresses has been calculated starting from the address level 1851 census data (Figure 3). It offers a comparison with similar analysis based on an average number of residents per street (Shiode et al, 2015: 5). These authors illustrate their approach to calculating the population at risk with the example of Husband Street, which had 120 residents in 10 houses giving an average of 12 persons per house based on data in the cholera inquiry report (Snow, 1855). According to the 1851 census there were 9 occupied addresses in Husband Street with 236 residents spread across 24 households. KDE has been applied to addresses along all thoroughfares within the boundary of Snow's study area (Figure 3). The KDE predicted mortality rates in Figure 3 (and Figures 4 and 5) are classified in a standardised way using the geometric interval method, which assigns approximately the same of values to each class and seeks reasonably consistent class intervals, and is well suited to continuous data where a moderate number of points have a value of zero. The density pattern is broadly similar to that produced by Shiode et al (2015) centred on Broad Street with some extension along thoroughfares to the north and east.

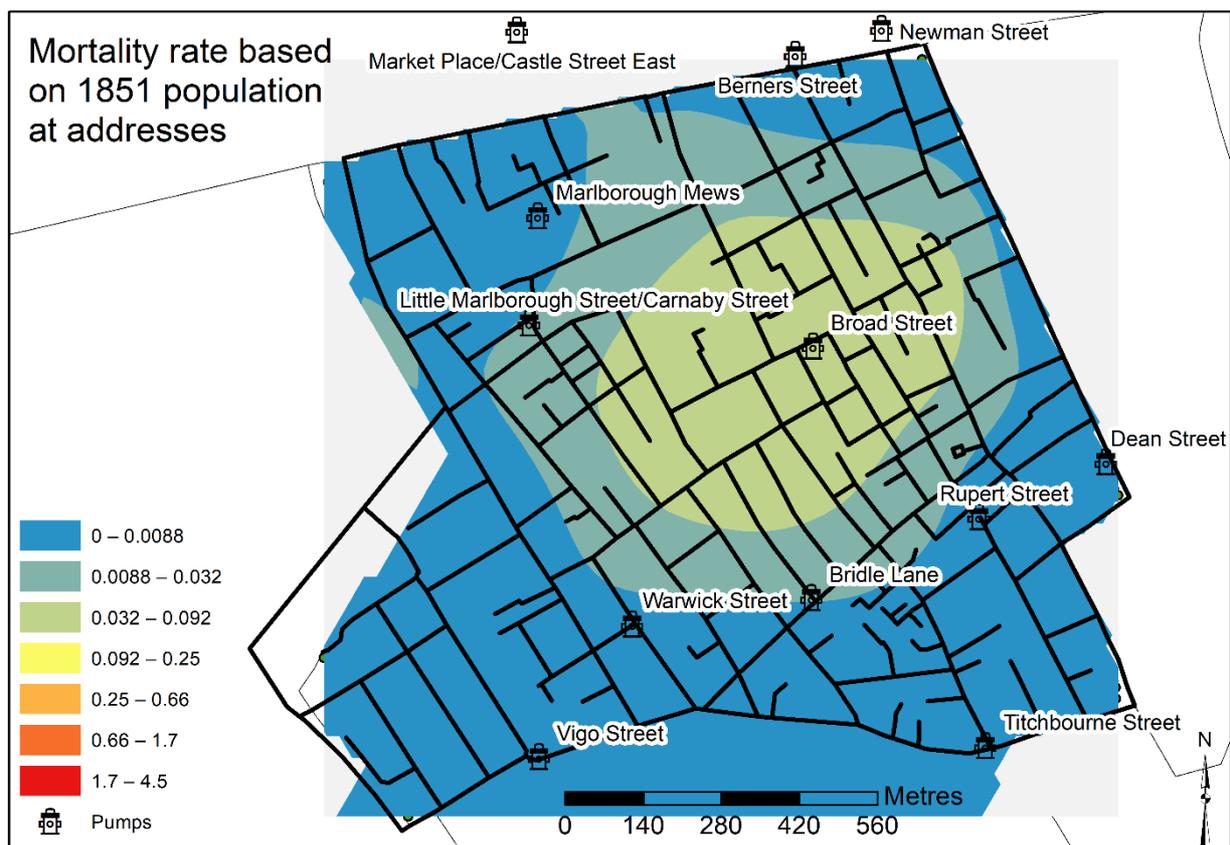


Figure 3: Kernel density estimation of cholera mortality using (N=623).

Table 3 indicates the prevalence of the HISCO occupational categories at the addresses where cholera deaths occurred. Occupation categories have been allocated to addresses in two ways: first by reference to the largest household in 1851 (if only one household present its category was used); and secondly by calculating the total number of persons at each address in the HISCO categories present and then assigning to each address the occupation of the largest number of people. Table 3 also shows the percentage of persons at an address who were categorised using the second method (omitting the small numbers in 0-

24.9% group). Addresses where people were working in dress (clothing) or in food accounted for the highest percentages: using the second allocation method 34.5 and 12.9 per cent of address were placed in these categories. The second method of allocating an occupation to an address related to at least 50 per cent of persons in present in fifteen of the categories. This analysis of occupations at addresses indicated that the second method of assignment would be a suitable starting point for investigating the spatial distribution of mortality rates for certain occupations.

HISCO occupation of head of household (HoH)	Address occupation determined by largest household	Address occupation determined by number of persons	Persons at address in specified occupation	25-49%	50-74%	75-99%	100%
				of persons at address in occupation category			
Persons working in agriculture	0.6	0.6	0.4	0.0	50.0	0.0	50.0
Persons working in building and construction	4.9	4.3	3.6	50.0	21.4	7.1	7.1
Persons working in commercial occupations	2.2	1.8	1.6	33.3	16.7	0.0	50.0
Persons working in conveying and transport	6.2	6.5	5.9	23.8	38.1	0.0	38.1
Persons working domestic service	4.6	4.9	6.1	43.8	25.0	0.0	31.3
Persons working in utilities	0.3	0.3	0.3	100.0	0.0	0.0	0.0
Persons working in government	4.3	4.3	3.6	21.4	14.3	7.1	57.1
Persons working professional occupations	6.5	5.8	4.1	21.1	21.1	5.3	52.6
Persons working defence	0.6	0.6	0.5	0.0	50.0	0.0	50.0
Persons working in wood and brick	0.6	0.6	0.4	0.0	0.0	0.0	100.0
Persons working chemicals	0.6	0.6	0.6	0.0	0.0	0.0	100.0
Persons working dress	25.8	34.5	43.0	29.5	40.2	6.3	23.2
Persons working in food	14.5	12.9	10.2	16.7	23.8	2.4	57.1
Persons working in metals	7.7	5.8	5.5	31.6	36.8	0.0	31.6
Persons working in paper	1.5	1.2	1.1	25.0	0.0	25.0	50.0
Persons working precious metals	2.5	1.8	1.2	0.0	16.7	16.7	66.7
Persons working in skins, hides, etc.	0.9	0.9	0.6	33.3	0.0	0.0	66.7
Persons working in textiles	2.2	1.8	1.7	0.0	50.0	0.0	50.0
Persons working in wood sector	6.5	5.5	4.3	27.8	16.7	0.0	55.6

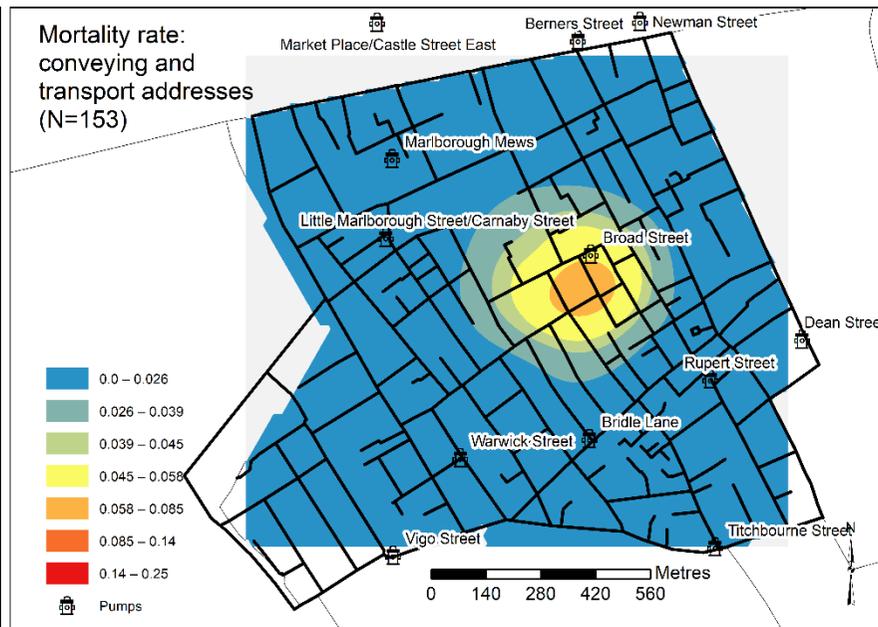
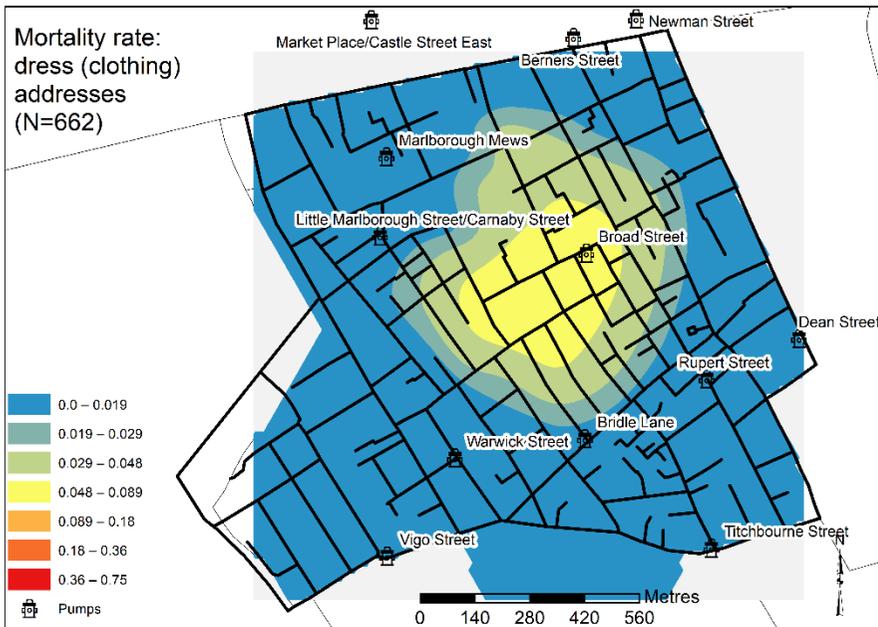
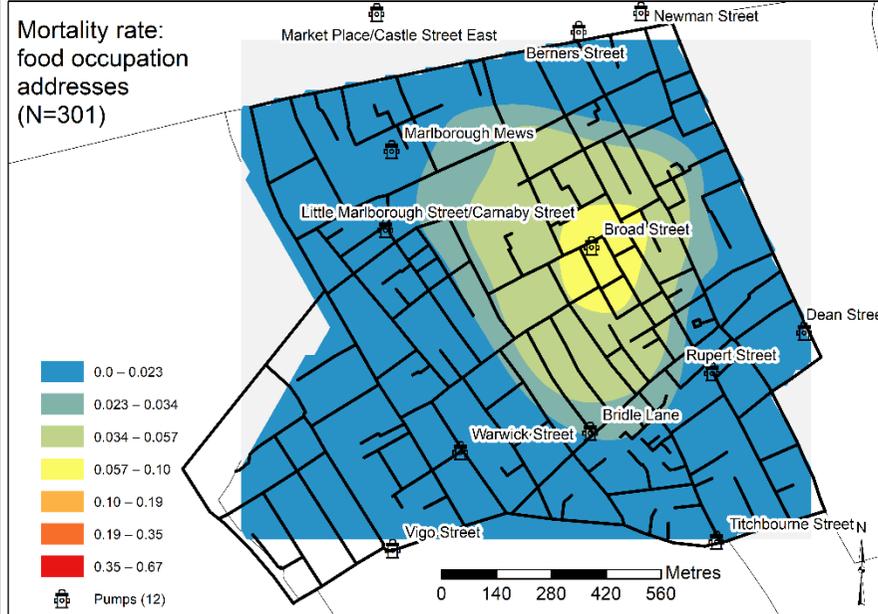
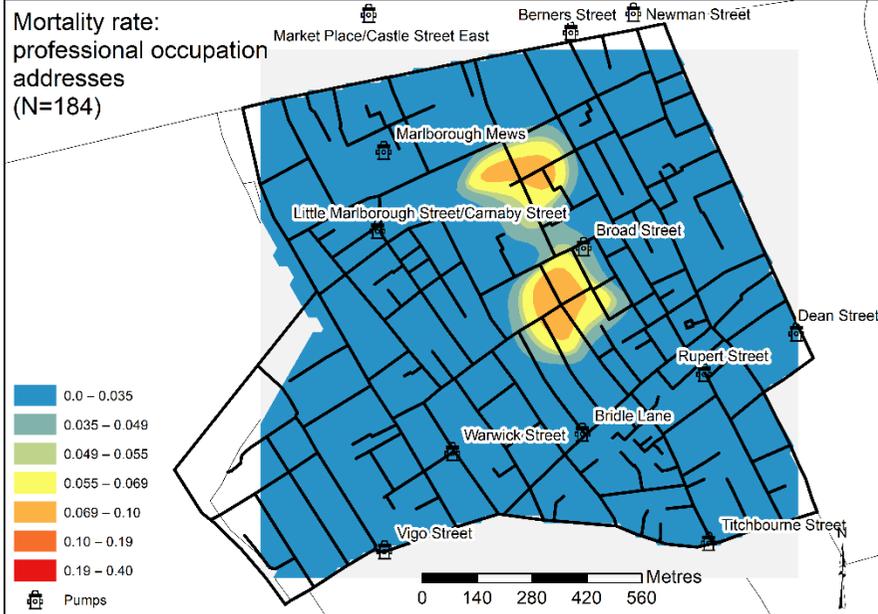
Persons working in or around mines	1.2	0.6	0.4	50.0	0.0	0.0	50.0
Persons in general or undefined work	4.3	3.1	3.2	20.0	20.0	20.0	40.0
Persons with unspecified or no occupation	1.5	1.2	1.7	0.0	0.0	25.0	50.0
Total	100	100	100				

Table 3: Occupation prevalence at addresses where deaths from cholera occurred (percentages).

Note: Addresses where no cholera deaths occurred and the Poland Street workhouse excluded from analysis. When presenting the percentage of persons at address in occupation category the lowest range (<25%) has not been included, which results in the four values presented not summing to 100%.

Source: 1851 Population Census.

The mortality rates used in the following analysis were calculated as outlined previously in respect of all addresses with persons present according to the 1851 census. The first part of Figure 4 relates to addresses where the majority of residents were in the following employment: professional occupations, food, dress (clothing), conveying (transport), domestic service and working or dealing in metals respectively these accounted for 4.1, 10.2, 43.0, 5.9, 6.1 and 5.5 per cent of persons with an occupation recorded (see Table 3). Furthermore, 57.9, 59.5, 29.5, 38.1, 31.3 and 31.6 per cent of people respectively employed in the six types of occupation accounted for between 75 and 100 per cent of residents at addresses. The six maps showing the kernel density estimated mortality for the occupation categories reveal differences, although clearly with a general focus on the Broad Street pump. Higher predicted cholera mortality for addresses categorized as professional occupations (N=184) were concentrated in two areas north and south of Broad Street. Addresses categorized as food (N=301) and dress (N=662) occupation types were more numerous and spread diffusely across a wider area. Addresses where the majority of residents worked in conveying of goods, people, etc. (N=153) also displayed some dispersion, but less than the previous two types. The focus of the kernel density prediction surface for domestic service addresses (N=142) lies somewhat to the south west of Broad Street, possibly reflecting a difference in the social standing households in that direction. The sixth occupation group examined in this analysis, people working or dealing in metals (non-precious) (N=97) displays high density along Broad Street itself with a minor peak on the eastern boundary of the study area.



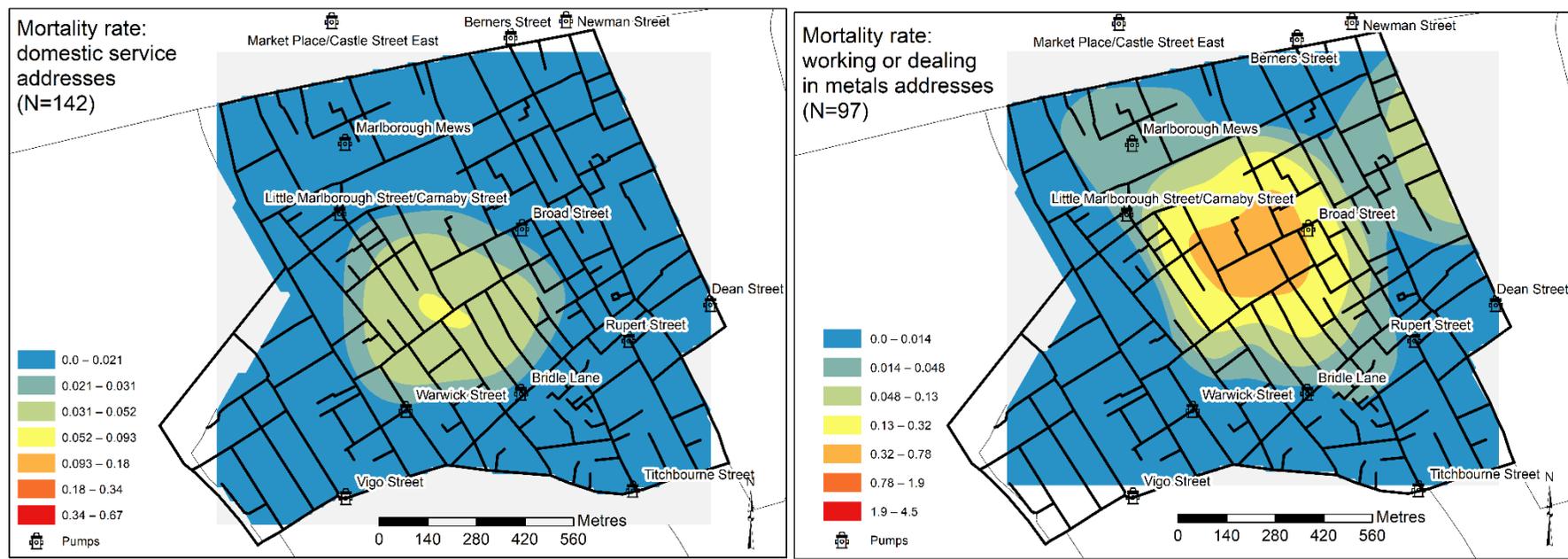


Figure 4: Kernel density estimation of cholera mortality rate for addresses categorised in selected occupational types.

Source: 1851 Population Census.

Shiode et al (2015: 6) commented that “elderly [people] were reported to have less cholera incidence than younger persons, as they lived on upper floors, thus having poor access to water the pump”. Figure 5 uses the kernel density estimated mortality rate in relation to differences in age structure and residential density of persons and households to explore these aspects of the cholera deaths. The upper two maps relate addresses where the percentages of residents aged 50 or over and aged 15 and under were in the upper quartile (N=474 and N=490 respectively). Mortality at addresses where there were substantial percentages of older people were focused around Broad Street with some extension eastwards. The predicted density of the mortality rate in households where young people were abundant covered a more compact area and reached a higher level, suggesting that older people were less inclined to obtain water from the infected pump. The kernel estimated predicted mortality of those addresses where m^2 per person was in the lowest quartile of the range (i.e. high residential density) was highest in an area spread around Broad Street. Kernel density estimation of the mortality rate where m^2 per household was in the lowest quartile of the range has a more fragmented distribution and highlights the high mortality that occurred in a relatively discrete area toward the south west of the study area.

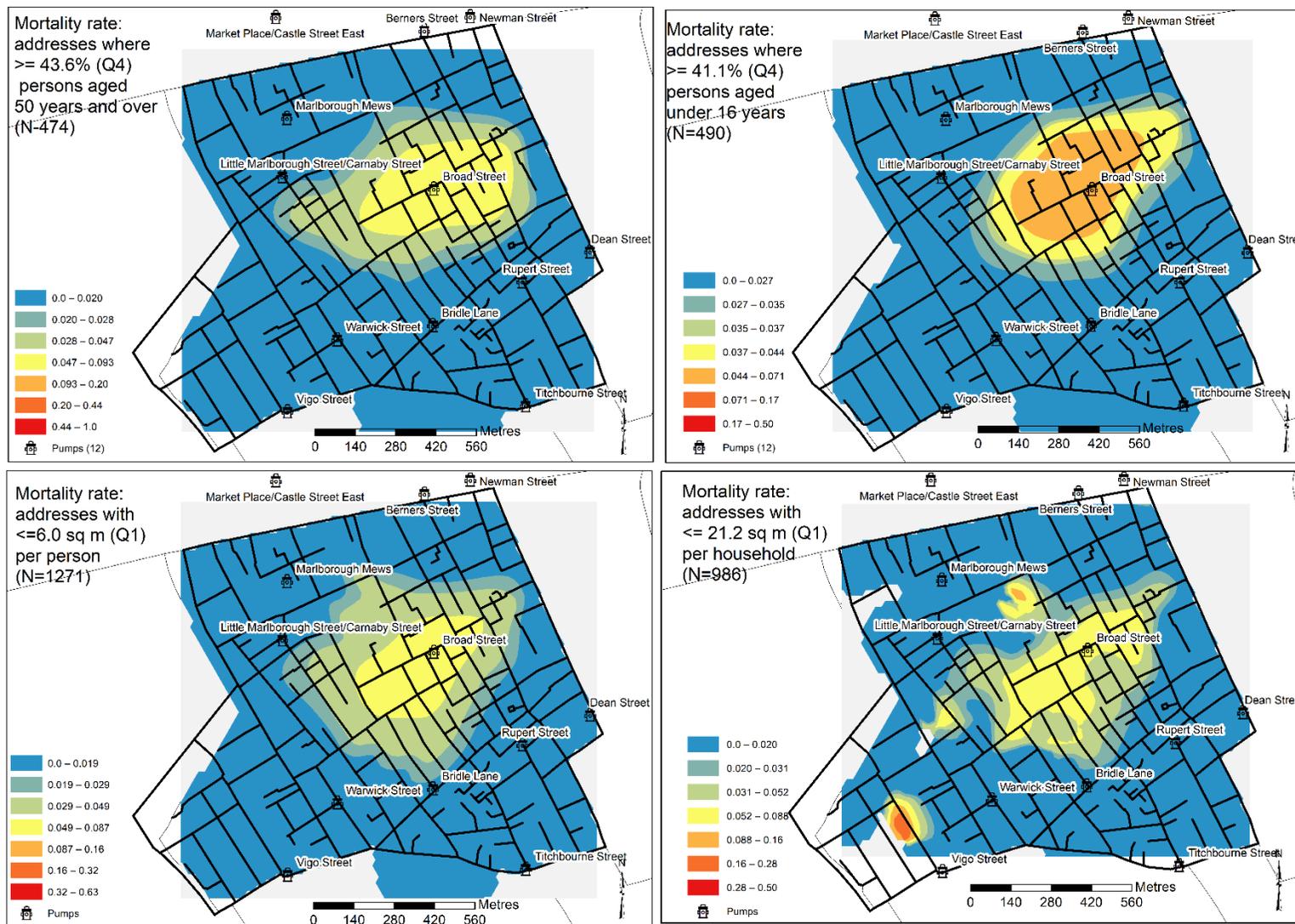


Figure 5: Kernel density estimation of cholera mortality rate for addresses with the percentage of younger or older residents in the upper quartile or residential density, m² per person and household in the lower quartile of the range.

Source: 1851 Population Census.

Conclusion

The analysis presented here has further demonstrated the new insights that can be obtained by combining historical data sources to cast new light on what might otherwise be regarded as a closed topic. Snow's original analysis of the deaths and cases of cholera during the 1854 outbreak in Soho is well known as a landmark in spatial analysis and subsequently replicated and augmented analyses of the geospatial data recorded on his map have revealed new aspects of the specific outbreak and the epidemiology of the disease overall. The present analysis moved in a new direction by seeking to discover if the demographic and socio-economic characteristics of the people inhabiting the streets of Soho in 1854 in some way predisposed them to succumbing to the cholera bacterium, now accepted as being transferred via water obtained from the Broad Street pump. In other words, given the variety of residents living on the streets around the pump who were equidistant to the point source of infection, were those with certain demographic or socio-economic characteristics or living in particular physical conditions more or less likely to become infected than others. The findings confirm that higher mortality rates were associated with age, occupation and residential density and resonates with research connecting higher incidence of the COVID-19 infection amongst Black, Asian and minority ethnic groups (Public Health England, 2020).

The statistical analysis performed on comparable variables obtained from the 1851 Population Census and the cholera inquiry visitation records indicates that while obviously there were some changes in the residents of these Soho streets between 30/31 March 1851 and the end of July 1854, the census counts are capable of spatially differentiating the demographic and socio-economic characteristics of the study area. Clearly the deaths arising from the cholera outbreak itself produced changes in the local population at that time, but the evidence presented here suggests that the census statistics can be regarded as a reasonable representation of the aggregate demographic and socio-economic characteristics of the population of the area and at the residential addresses at the start of the outbreak, even if some individuals were not present at both times. The findings indicate that in aggregate terms the streets where at least one cholera death occurred and those where the disease was absent were significantly different from the study area as a whole in statistical terms. Examination of the kernel density estimated mortality rate for addresses where six occupation types were prevalent, where the percentages of older and younger residents were in the upper quartile and where m² per person and household were in the lowest quartile reveal differences in mortality. Such findings connect with contemporary research concerning differential exposure to disease risk and the likelihood of succumbing to infection. The occurrence of deaths from cholera during the 1854 outbreak were not only related to distance from the Broad Street pump but were mitigated by occupation, age structure and residential density.

References

- Brown, D., 2016. Supplying London's Workhouses in the Mid-Nineteenth Century. *The London Journal* 41, 36-59. <https://doi.org/10.1080/03058034.2015.1127696>
- Cliff, A., Haggett, P., 1988. *Atlas of Disease Distributions: Analytic Approaches to Epidemiological Data*. Blackwell, Oxford.
- Eyler, J., 1979. *Victorian Social Medicine: The Ideas and Methods of William Farr*. Johns Hopkins University Press, London.

Fotheringham, A., Kelly, M., Charlton, M., 2013. The demographic impacts of the Irish famine: towards a greater geographical understanding. *Transactions of the Institute of British Geographers* 38, 221-237. <https://doi.org/10.1111/j.1475-5661.2012.00517.x>.

Frost, W., 1936. *Introduction to Snow on cholera*. Commonwealth Fund, New York.

General Board of Health, 1855. *Appendix to Report of the Committee for Scientific Inquiries in Relation to the Cholera-Epidemic of 1854*. HMSO, London.

General Registrar Office, 1852. *Report on the morality of cholera in England 1848-49*. HMSO, London.

Gilbert, E., 1958. Pioneer maps and health and disease in England. *Geographical Journal* 124, 172-183. <https://doi.org/10.2307/1790244>.

Gregory, I., Bennett, C., Gilham, V., Southall, H., 2014. The Great Britain Historical GIS project: from maps to changing human geography. *The Cartographic Journal* 39, 37-49. <https://doi.org/10.1179/caj.2002.39.1.37>

Gregory, I., Cooper, D., 2013. Geographical technologies and the interdisciplinary study of peoples and cultures of the past. *Journal of Victorian Culture* 18, 265-272. <https://doi.org/10.1080/13555502.2013.797686>.

Gregory, I., Ell, P., 2005. Breaking the boundaries: geographical approaches to integrating 200 years of the census. *Journal of the Royal Statistical Society* 168, 419-437. <https://doi.org/10.1111/j.1467-985X.2005.00356.x>.

Gregory, I., Geddes, A., 2014. Introduction: from historical GIS to spatial humanities: deepening scholarship and broadening technology. In: Gregory, I., Geddes, A., (Eds) *Towards Spatial Humanities: Historical GIS and Spatial History*. Indiana University Press, Bloomington and Indianapolis

Higgs, E., 1989. *Making Sense of the Census. The Manuscript Returns for England and Wales, 1801-1901*. HMSO, London.

Higgs, E., 1991. Disease, febrile poisons, and statistics: the census as a medical survey, 1841-1911. *Social History of Medicine* 4, 465-478. <https://doi.org/10.1093/shm/4.3.465>.

Higgs, E., Jones, C., Schürer, K., Wilkinson, A., 2013. *Integrated Census Microdata (I-CEM) Guide*. Department of History, University of Essex, Colchester, Essex, UK.

Hitchcock, T., 2013. Confronting the digital: or how academic history writing lost the plot. *Cultural and Social History* 10, 9-23. <https://doi.org/10.2752/147800413X13515292098070>.

Hitchcock, T., Shoemaker, R., 2014. *Making History Online*. Colin Matthew Memorial Lecture, Royal Historical Society/Gresham College. Available at: <http://www.gresham.ac.uk/lectures-and-events/making-history-online> [Accessed 12/03/20]

Hitchcock, T., Shoemaker, R., Emsley, C., Howard, S., McLaughlin, J et al., 2015. *The Old Bailey Proceedings Online, 1674-1913*. Available at: www.oldbaileyonline.org, version 7.2 [Accessed: 16/03/16]

John Snow Archive and Research Companion (nd) <http://johnsnow.matrix.msu.edu/>.

Knowles, A., 2008. GIS and history. In: Knowles, A., (Ed.) *Placing History: How maps, Spatial Data, and GIS are Changing Historical Scholarship*. ESRI Press, California.

Knutzen, M., 2014. Framing the vision and conference agenda. Moving Historical Geodata to the Web: Harvard University Center for Geographical Analysis. Available at: <http://gis.harvard.edu/services/blog/moving-historical-geodata-web> [Accessed 16/03/16]

Koch, T., 2005. Cartographies of Disease: Maps, Mapping and Medicine. Esri Press, Redlands, CA.

Koch, T., 2011. Disease Maps: Epidemics on the Ground. University of Chicago Press, Chicago.

Konteh, F., 2009. Urban sanitation and health in the developing world: reminiscing the nineteenth century industrial nations. *Health and Place* 15, 69-78. <https://doi.org/10.1016/j.healthplace.2008.02.003>

Leeuwen, M., van, Maas, I., Miles, A., 2002. HISCO: Historical International Standard Classification of Occupations. Leuven University Press, Leuven.

Longley, P., Goodchild, M., Maguire, D., Rhind, D., 2005. Geographic Information Systems and Science. John Wiley, Chichester.

May, J., 1958. The Ecology of Human Disease. MD Publications, New York.

Navickas, K., 2016. Political Meetings Mapper with British Library Labs: mapping the origins of British democratic movements with text-mining, NLP, geo-parsing and crowd-sourcing. Institute of Historical Research Digital History Seminar. Available at <https://www.youtube.com/watch?v=zgaUkTn-T-E> [Accessed: 21/02/20]

Parkes, E., 1855. Mode of communication of cholera. *The British and Foreign Medico-Chirurgical Review* 15, 449-463. PMID: [30164367](https://pubmed.ncbi.nlm.nih.gov/30164367/)

Perez-Heydrich, C., Furgurson, J., Giebultowicz, S., Winston, J., Emch, M., 2013. Social and spatial processes associated with childhood diarrheal disease in Matla, Bangladesh. *Health and Place* 19, 54-52. <https://doi.org/10.1016/j.healthplace.2012.10.002>

Plewe, B., 2003. Representing datum-level uncertainty in historical GIS. *Cartography and Geography Information Science* 30, 319-334. <https://doi.org/10.1559/152304003322606229>.

Public Health England, 2020. Beyond the data: understanding the impact of COVID-19 on BAME groups. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892376/COVID_stakeholder_engagement_synthesis_beyond_the_data.pdf [Accessed: 25/06/20]

Schürer, K., Higgs, E., 2014. Integrated Census Microdata (I-CeM), 1851-1911. [data collection]. UK Data Service. SN: 7481. <https://doi.org/10.5255/UKDA-SN-7481-1>

Sedgwick, W., 1902. Principles of Sanitary Science and the Public Health: With Special Reference to the Causation and Prevention of Infectious Diseases. McMillan, New York.

Shi, X., 2010. Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. *International Journal of Geographical Information Science* 24, 643-660. <https://doi.org/10.1080/13658810902950625>

Shiode, N., Shiode, S., Rod-Thatcher, E., Rana, S., Vinten-Johansen, P., 2015. The mortality rates and the space-time patterns of John Snow's cholera epidemic map. *International Journal of Health Geographics* 14, 1-16. <https://doi.org/10.1186/s12942-015-0011-y>.

Snow, J., 1849. On the pathology and mode of communication of cholera. *London Medical Gazette* 2 November 1849, 745-752 Part 1. John Snow Archive and Research Companion available at <http://johnsnow.matrix.msu.edu/work.php?id=15-78-2A> [Accessed 12/12/19]

Snow, J., 1854. Plan shewing the ascertained deaths from cholera in part of the parishes of St James, Westminster, and St Anne, Soho, during the Summer and Autumn of 1854. Available at <https://archive.org/details/b24759429/mode/2up> [Accessed: 11/12/16]

Snow, J., 1854. The Cholera near Golden-square, and at Deptford. *Medical Times Gazette*, 321-322.

Snow, J., 1855. Report on the Cholera Outbreak In the parish of St. James, Westminster, during the Autumn of 1854. Presented to the Vestry by the Cholera Inquiry Committee, John Churchill, London.

Southall, H. 2014. Rebuilding the Great Britain Historical GIS, Part 3: integrating qualitative content for a sense of place. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 47, 31–44. <https://doi.org/10.1080/01615440.2013.847774>.

Southall, H., 2003. A vision of Britain through time: making long-run statistics of inequality accessible to all. *Radical Statistics* 82, 26-43. <http://www.radstats.org.uk>.

Southall, H., 2006. Electronic resources for local population studies: a vision of Britain through time: making sense of 200 years of census reports. *Local Population Studies* 76, 76–89.

Stamp, D., 1964. *Some Aspects of Medical Geography*. Oxford University Press, London.

Timmermans, J., Reinhard, E., Ruijsbroek, A., Huisman, M., Avendano, M., 2020 Regeneration of deprived neighbourhoods and indicators of functioning older adults: a quasi-experimental evaluation of the Dutch District Approach. *Health and Place* 64, 102359. <https://doi.org/10.1016/j.healthplace.2020.102359>

Tobler, W., 1994. Snow's Cholera Map. Available at <http://ncgia.ucsb.edu/pubs/snow/snow.html> [Accessed: 13/03/2020]

Walford, N., 2019. Bringing historical British Population Census records into the twenty-first century: a method for geocoding households and individuals at their early twentieth century addresses. *Population, Place and Space*. <https://doi.org/10.1002/psp.2227>

Wellcome Library, (nd). Plan shewing the ascertained deaths from cholera in part of the parishes of St James, Westminster, and St Anne, Soho, during the Summer and Autumn of 1854. https://search.wellcomelibrary.org/iii/encore/record/C_Rb2475941_Splan%20of%20deaths_Orightresult_U_X2?lang=eng&suite=cobalt

White, R., 2010. What is spatial history? Available at <https://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29> [Accessed: 13/03/2020]