Public Health (2004) **118**, 387–394

ELSEVIER

PUBLIC HEALTH
JOURNAL OF THE ROYAL INSTITUTE OF PUBLIC HEALTH
www.elsevierhealth.com/journals/pubh

John Snow, William Farr and the 1849 outbreak of cholera that affected London: a reworking of the data highlights the importance of the water supply

P. Bingham[a,*], N.Q. Verlander[b], M.J. Cheal[a]

[a] *Isle of Wight Primary Care Trust, Whitecroft, Sandy Lane, Newport, I.O.W PO30 3ED, UK*
[b] *PHLS Statistics Unit, Colindale, UK*

Received 11 August 2003; received in revised form 8 January 2004; accepted 24 May 2004

KEYWORDS
John Snow; William Farr; Cholera; Logistic regression

Summary *Objectives*. This paper examines why Snow's contention that cholera was principally spread by water was not accepted in the 1850s by the medical elite. The consequence of rejection was that hundreds in the UK continued to die.

Methods. Logistic regression was used to re-analyse data, first published in 1852 by William Farr, consisting of the 1849 mortality rate from cholera and eight potential explanatory variables for the 38 registration districts of London.

Results. Logistic regression does not support Farr's original conclusion that a district's elevation above high water was the most important explanatory variable. Elevation above high water, water supply and poor rate each have an independent significant effect on district cholera mortality rate, but in terms of size of effect, it can be argued that water supply most strongly 'invited' further consideration.

Conclusions. The science of epidemiology, that Farr helped to found, has continued to advance. Had logistic regression been available to Farr, its application to his 1852 data set would have changed his conclusion.

Introduction

Snow's hypothesis that cholera could be spread in water, food or hand-to-mouth was not accepted in the 1850s, arguably because Snow was not a member of the medical elite. Although he gave chloroform to Queen Victoria for pain relief in child birth, he was not one of the medical signatories on the birth announcement for Prince Leopold.[1] Instead, the medical fraternity supported William Farr, the dominant epidemiologist of the day, who had concluded that the available data not only supported miasma (spread via atmospheric vapours) but also indicated that there was an underlying 'natural law' linking infection with cholera inversely to elevation above high water.[2] This paper re-analyses data that were contested between Snow and Farr.

John Snow

Snow first published his hypothesis on the causation of cholera in 1849, shortly after

*Corresponding author. Tel.: +44-1983-535409; fax: +44-1983-537140.
E-mail address: paul.bingham@iow.nhs.uk

Table 1 Deaths from cholera in London, registered from 23 September 1848 to 25 August 1849.

Sectors of London	Population in 1841	Deaths from cholera	Death rate per 1000 inhabitants
West	300,711	533	1.77
North	375,971	415	1.10
Central	373,605	920	2.46
East	392,444	1597	4.07
South	502,548	4001	7.96
Total	1,945,279	7466	3.84

Snow J. *On the Mode of the Communication of Cholera*. London: John Churchill; 1849.

the second outbreak had swept through England.[3,4] In support of his hypothesis, initially Snow only had anecdotal evidence and the data reproduced in Table 1.

Snow pointed out that more deaths had occurred in the south districts of London than all the other districts put together (3465 deaths, death rate 2.4 per 1000 inhabitants).[3] He suggested that this was because water for the south districts was taken from the River Thames below Vauxhall Bridge, where the river was polluted by sewage discharged into the Thames higher in its course. The other districts of London either abstracted from the cleaner, high reaches of the Thames or from its tributaries. Snow obtained his data from the Weekly Return of the Registrar-General that was published by William Farr, the Compiler of Abstracts to the General Register Office.

William Farr

In 1852, Farr published additional data on deaths that had occurred from cholera in London during 1849, including eight potential explanatory variables for the 38 registration districts of London (see Table 2).[5] Farr examined the association between death from cholera and the explanatory variables by 'grouping' the data (see Figs. 1–4), and by comparing the death rate for the 19 districts with the highest values of the explanatory variable with the 19 districts with the lowest values (see Table 3).

Farr found an association between death from cholera and all of the eight explanatory variables. However, on the basis of a more consistent relationship with elevation (see Fig. 1) and the relatively large differential between the 19 most-elevated and 19 least-elevated districts (1:3), he concluded that the association with elevation was the most important: 'The elevation of the soil in London has a more constant relation with mortality from cholera than any other known element'.[5] Farr accepted that there was an association between mortality from cholera and a district's water supply, and personally exhorted the water company to make improvements.[6] However, Farr also pointed out that if registration districts were sorted into three 'water supply groups' and then stratified according to elevation, an elevation effect was still present (see Table 4). 'While the effects of water of the districts are apparent, they do not, in this analysis, conceal the effects of elevation'.[5]

Farr's natural law

Farr progressed his concern with elevation by sorting registration districts into 'terraces', 0–19-feet elevation, 20–39-feet elevation etc. and calculated the average mortality for each terrace. Farr noted that if the 'observed' mortality of the 0–19-feet terrace (102 deaths per 10,000 population) was divided successively by 2, 3, 4, 5 and 6, the observed mortality of successive terraces was approximated (see Table 5). Farr concluded that this reflected a natural law and went on to derive a formula that, he demonstrated, predicted observed mortality at subdistrict level.[5]

Data and methods

The data re-analysed in this paper, first published by Farr in 1852,[5] relate to deaths from cholera that were registered in London during 1849 (a total of 14,137 deaths, see Table 2). This was part of the second outbreak of cholera to affect the UK and began on 22 September 1848 when a sailor was diagnosed with the condition in Southwark. In total, 652 residents died from cholera in London in 1848. The outbreak continued until the end of 1849.

Explanatory variables

Farr's original study included eight explanatory variables (see Table 2), as follows.

Elevation

The elevation above the Trinity high-water mark of each registration district was estimated by Captain Dawson RE of the Tithe Commission. Clearly a 'mean' elevation is of questionable

Table 2 Deaths from cholera in London, registered in 1849 by registration district together with eight possible explanatory variables.

District	Deaths from cholera in 1849 per 10,000 inhabitants	Elevation above high water (feet)	Annual deaths from all causes 1838–1844 per 10,000 inhabitants	Persons per acre	Persons per inhabited house	Average annual value of house (£)	Annual value of house per person (£)	Poor rate precept per pound of house value	Water supply[a]
Newington	144	−2	232	101	5.8	22	3.788	0.075	1
Rotherhithe	205	0	277	19	5.8	23	4.238	0.143	1
Bermondsey	161	0	264	66	6.2	18	3.077	0.134	1
St George Southwark	164	0	267	181	7.0	22	3.318	0.089	1
St Olave	181	2	281	114	7.9	35	4.559	0.079	1
St Saviour	153	2	292	141	7.1	36	5.291	0.079	1
Westminster	68	2	260	70	8.8	36	4.189	0.076	1
Lambeth	120	3	233	34	6.5	28	4.389	0.039	1
Camberwell	97	4	197	12	5.8	25	4.508	0.072	1
Greenwich	75	8	238	18	6.8	22	3.379	0.038	2
Poplar	71	10	241	15	6.2	44	7.360	0.081	2
Chelsea	46	12	287	62	7.1	29	4.210	0.060	1
Hammersmith, Brompton, Kensington and Fulham	33	12	228	11	6.6	33	5.070	0.067	3
St George in the East	42	15	289	195	6.9	32	4.753	0.039	2
Stepney	47	16	242	85	6.3	20	3.319	0.080	2
Belgrave	28	19	194	65	8.0	67	8.875	0.066	1
Wandsworth	100	22	198	4	6.2	29	4.839	0.018	1
West London	96	28	302	212	9.7	65	7.454	0.072	2
Whitechapel	64	28	290	194	8.1	26	3.388	0.067	2
Lewisham	30	28	173	2	5.8	27	4.824	0.075	2
St Martin-in-the-Fields	37	35	240	81	10.3	119	11.844	0.049	2
Bethnal Green	90	36	239	115	6.3	9	1.480	0.039	2
London city	38	38	214	129	7.1	117	17.676	0.136	2
East London	45	42	259	284	8.3	38	4.823	0.056	2
St James Westminster	16	43	212	222	10.4	128	12.669	0.088	3
Shoreditch	76	48	251	161	6.6	20	3.103	0.023	2
St Luke	34	48	276	242	7.8	28	3.731	0.082	2
Hanover Square and Mayfair	8	49	179	57	9.5	153	16.754	0.081	3
Strand	35	50	242	254	10.1	66	7.374	0.018	2
Holborn	35	53	266	235	9.7	52	5.883	0.047	2
Hackney	25	55	197	14	5.9	25	4.397	0.034	2
Clerkenwell	19	63	242	202	8.2	33	4.138	0.074	2
St Giles	53	68	269	221	11.0	60	5.635	0.057	2
Paddington	8	76	197	32	7.3	64	9.349	0.052	3
St Pancras	22	80	222	59	8.8	41	4.871	0.039	2
Islington	22	88	200	28	6.6	35	5.494	0.042	2
Marylebone	17	100	227	102	9.8	71	7.586	0.030	3
Hampstead	8	350	202	5	7.2	40	5.804	0.043	3

Table sorted by district elevation above Trinity high-water mark.
[a] 1, Thames between Battersea Bridge and Waterloo Bridge; 2, New River/Rivers Lea and Ravensbourne; 3, Thames between Kew and Hammersmith.

validity, not least because districts varied in size between 136 and 17,224 acres. (Note: small corrections were made to the estimation of mean elevation in data subsequently published for the 1866 outbreak of cholera; however, this paper focuses on the validity of Farr's analysis that was questioned by Snow, rather than the validity of the data itself).

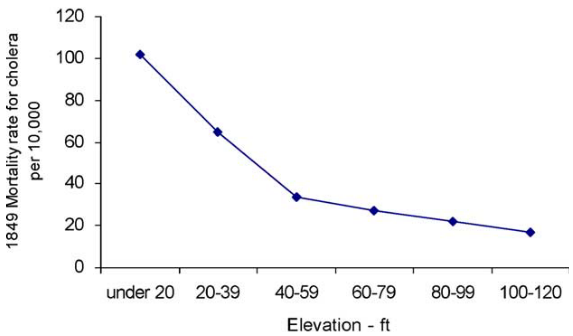

Figure 1 Death from cholera vs. elevation.

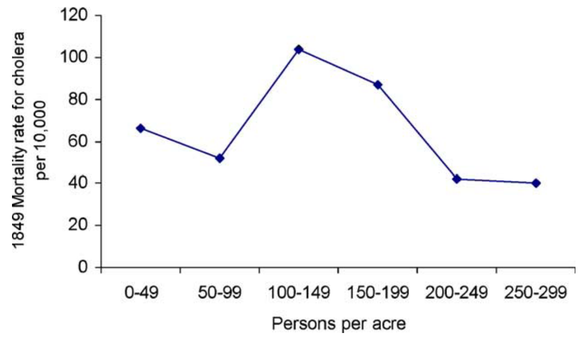

Figure 3 Death from cholera vs. persons per acre.

Crowding

Farr included two measures of crowding: persons per house and persons per acre.

Wealth

Farr calculated two measures of wealth: the average value of a house within a district and the average house value per head of the district population.

Poor rate

The poor rate precept per pound of house value (if the precept was set at 0.1 and a person owned a house valued at £10, they paid £1 poor rate). Poor rate was determined by the number of district residents qualifying for relief, with the total house value of the district forming the denominator. Farr's tables omitted the poor rate precept for Hampstead. Records held at the House of Commons's Library do not give a precise figure but the precept is estimated to lie between 0.045 and 0.050 (Andrew Presland, personal communication).

Annual mortality

The average annual mortality from all causes per 10,000 inhabitants between 1838 and 1844.

Water supply

Farr divided the registration districts of London, based on their water supply, into three (see Table 4): districts that obtained their water from the Thames above Hammersmith Bridge; districts that obtained their water from tributaries of the Thames (Rivers New, Lea and Ravensbourne); and districts that obtained water from below Battersea Bridge. Vauxhall Bridge, suggested as a boundary by Snow, is located between Battersea Bridge and Waterloo Bridge.

Reworking of data and statistical methods

The outcome variable in Farr's analysis was the number of deaths from cholera in 1849. In the re-analysis, the outcome is assumed to follow a binomial distribution with the number of individuals at risk, i.e. the population totals, as the denominators. Logistic regression has been used to fit and compare a series of models to assess the form of associations between the possible explanatory variables and variation in the risk of death from cholera on a natural log (odds) scale. Effects are estimated as changes in log (odds), i.e. log (odds ratios).

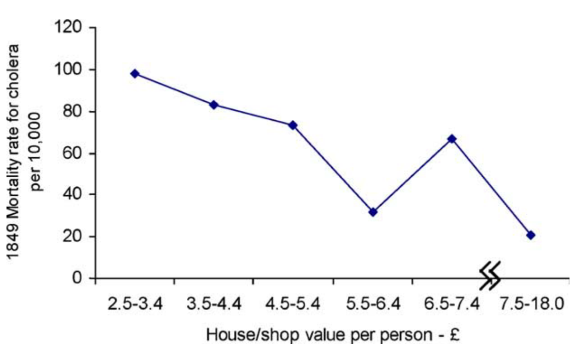

Figure 2 Death from cholera vs. house/shop value per person.

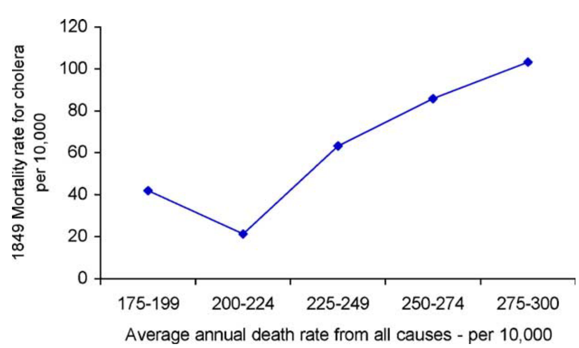

Figure 4 Death from cholera vs. annual mortality from all causes 1838–1844.

Table 3 Comparison of the cholera death rate for the 19 districts with the highest values of the explanatory variable with the 19 districts with the lowest values.

Explanatory variable	Mortality from cholera per 10,000 population for the 19 districts with the highest values of the explanatory variable compared with the 19 districts with the lowest values (ratio)
Elevation	33:100 (1:3.0)
Average house value per person	43:90 (1:2.1)
Persons per acre	71:61 (1.16:1)
Annual death rate 1838-1844	84:48 (1.75:1)

Table 5 Observed mortality compared with expected for 'terraces of London'.

Elevation of the 'terrace' above the Trinity high-water mark (feet)	Number of registration districts	Observed mortality from cholera per 10,000 inhabitants	102 divided by 1, 2, 3, 4, 5 and 6 ('expected' mortality)
0-19	16	102	102
20-39	7	65	51
40-59	8	34	34
60-79	3	27	26
80-99	2	22	20
100-120	1	17	17
340-359	1	8	-

Results

Estimated odds ratios and *P* values from a logistic regression model with a simple linear trend with the poor rate for Hampstead set at 0.045 are shown in Table 6. The results with the poor rate for Hampstead set at 0.05 are very similar. No interactions between the explanatory variables were found to be significant at $P < 0.05$. In particular, the interaction between water source and elevation was not significant ($P = 0.22$). A simpler model than that shown in Table 6 could be produced by removing the most insignificant variable with either a forwards or backwards stepwise procedure. The main purpose of such a procedure would be to increase the number of observations for model estimation. However, in this paper, the rather unusual situation exists where there is no missing data and so the model selection strategy referred to above would not lead to an increase in the number of available observations. Hence, the model as shown in Table 6 is presented, as this table also shows useful null information.

Table 6 shows that there is a significant association between water supply and risk of death from cholera, with water from the Thames between Battersea Bridge and Waterloo Bridge being significantly worse than either water from New River/Rivers Lea and Ravensbourne or from the Thames above Hammersmith Bridge with an estimated decrease in risk of 41 and 60%, respectively. There is also a significantly decreased risk with increasing elevation and decrease in the poor rate. None of the other variables appear to be associated with death from cholera.

Figs. 5 and 6 show the estimated relationships between death rate and poor rate and death rate and elevation, respectively, for the three water supply areas, illustrating the results given in Table 6. The left-hand plots are on a log scale and the right-hand plots are on the original scale. The data are plotted as squares, triangles and circles and the fitted models are shown as curves.

In order to put the apparent effect of elevation and poor rate into context, Table 7 shows the change in estimated risk over the interquartile range of these variables. Between district elevation of 8 and 50 ft, the risk of death from cholera decreases by 32.7%. Between a district poor rate of £0.039 and £0.079, the risk of death from cholera decreases by 31.4%.

It is suggested that, based on the size of effect, water supply most strongly invited further consideration.

Table 4 Cholera mortality rates by registration district sorted according to their source of water supply and stratified by elevation.

Source of water supply	Mortality from cholera per 10 000 inhabitants	Elevation of the 'stratum' above the Trinity high-water mark (feet)
Thames between Kew and Hammersmith		
Three highest districts	11.0	175.0
Three lowest districts	19.0	35.0
New River/Rivers Lea and Ravensbourne		
Ten highest districts	36.6	59.5
Ten lowest districts	59.0	24.2
Thames between Battersea Bridge and Waterloo Bridge		
Six highest districts	77.0	10.0
Six lowest districts	168.0	0.3

Table 6 Odds ratios, 95% confidence limits (CL) and *P* values for logistic regression model.

Explanatory variable	Low 95% CL	Odds ratio	High 95% CL	*P*
Constant	6.006×10^{-4}	0.002626	0.01149	-
Water from Thames between Battersea Bridge and Waterloo Bridge[a]		1.00		<0.001
Water from New River and Rivers Lea and Ravensbourne	0.44	0.59	0.79	
Water from Thames between Kew and Hammersmith	0.22	0.40	0.72	
Increase in elevation above high water (10 feet)	0.85	0.91	0.98	<0.01
Decrease in poor rate (£/100)	0.87	0.91	0.96	<0.001
Average annual death rate 1838–1844	1.00	1.00	1.01	0.48
Persons per inhabited house	0.89	1.03	1.19	0.71
Persons per acre	1.00	1.00	1.00	0.67
Average house value per person (£)	1.00	1.00	1.00	0.35
Average house value within district (£)	1.00	1.00	1.00	0.79

[a] Baseline.

Discussion

This paper details why Farr concluded that mortality from cholera was explained by a registration district's elevation above high water: the consistent trend with elevation (Fig. 1); the large difference in mean mortality between the 19 least-elevated districts compared with the 19 most-elevated districts (Table 3); the persistence of an elevation effect when districts were sorted into water supply groups (Table 4); and, finally, a relationship that Farr suggested reflected an underlying 'natural law' (Table 5).

Modern logistic regression that makes best use of all the data, however, shows that three variables are independently associated with mortality from cholera. On the basis of the size of effect, it is suggested that water supply most strongly invited further consideration.

The lack of specificity in defining the water supply to districts (more than three companies were involved) may partly explain the observed association between elevation and mortality. The explanation for the association with the poor rate is less clear, but it is possible that mortality was very high among those claiming poor relief.

Farr's conviction that mortality was more closely linked to elevation than water supply was highly influential, and was used by the public health establishment of the day to justify the miasma theory. Even in 1855, the Scientific Committee that investigated the previous year's visitation of cholera that included the Broad Street pump

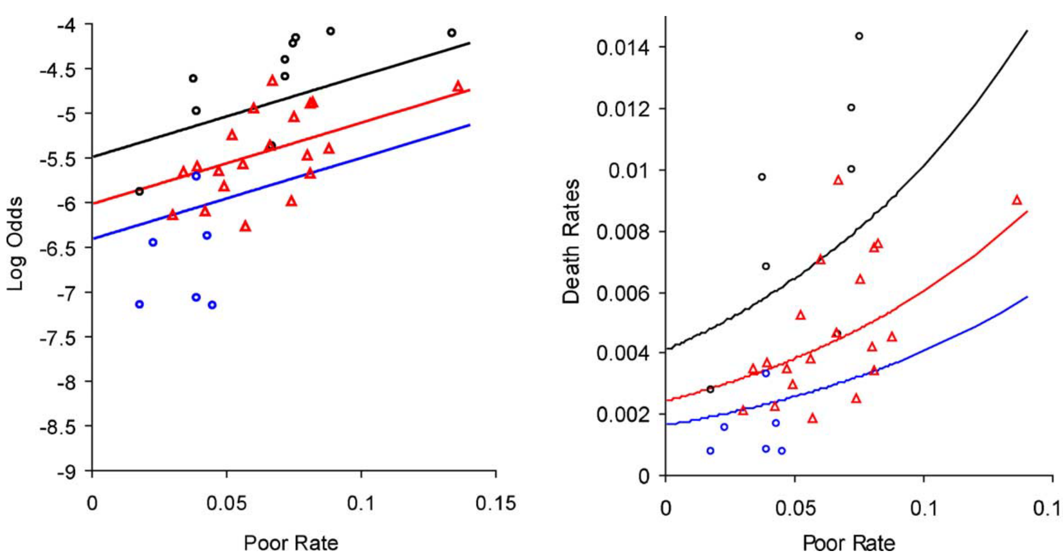

Figure 5 Estimated relationship of log (odds of death) and death rates with poor rate for three water supplies in a registration district with the mean annual death rate, housing density, elevation, area, number of houses and house value. See Fig. 6 for key.

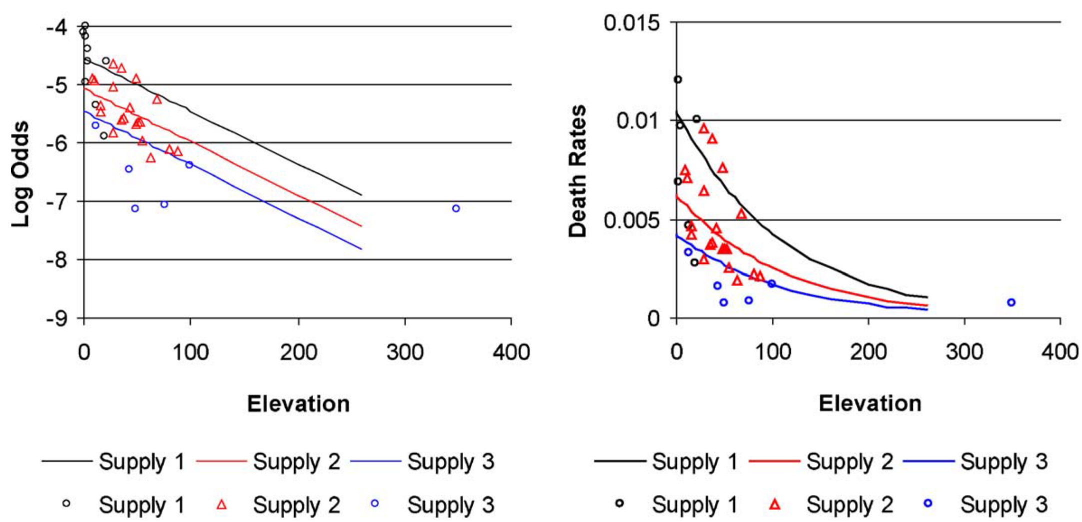

Figure 6 Estimated relationship of log (odds of death) and death rates with elevation for three water supplies in a registration district with the mean annual death rate, housing density, poor rate, area, number of houses and house value.

Table 7 Estimated change in the risk of cholera over the interquartile range for 'elevation' and 'poor rate'.

	Lower quartile	Median	Upper quartile	Interquartile range	Change in risk of cholera over interquartile range %
Elevation above high water (feet)	8	28	50	42	32.7
Poor rate (£/100)	3.9	6.7	7.9	4	31.4

outbreak concluded that "On the whole evidence, it seems impossible to doubt that the influences, which determine in mass the geographical distribution of cholera in London, belong less to the water than to air... mortality has more nearly followed the degrees of elevation of soil than been proportionate to any other general influence we could measure."[7]

An earlier understanding of the predominant role of a contaminated water supply in cholera transmission might well have prevented cases, particularly during the 1854 outbreak. It should, however, be noted that changes to London's water supply were already in hand. The Water Supply (Metropolis) Act of 1852 required companies to move their intakes from the River Thames by 31 August 1856 to above the tidal flow at Teddington lock.

By 1866, when England was affected by a further outbreak of cholera, Farr had realized that a contaminated water supply could be a direct cause of cholera and was quick to associate a sudden increase in cases in East London with contamination of a reservoir.[8]

Farr, however, continued to be impressed by the influence of elevation. While he accepted that local outbreaks could be waterborne, Farr continued to believe that the 'background rate of cholera' was accounted for by personal contact and through airborne transmission. In addition, Farr hypothesized that the agent (cholera flux) was heavier than water and therefore might be at higher concentrations in water pipes at low elevations than in pipes at high elevations.[9]

Farr was a great statistician but application of a modern technique (logistic regression) to his 1852 cholera data set demonstrates that, in this instance, his analysis led to the wrong conclusion being reached.

Acknowledgements

Dr Tony Swan of the PHLS Statistics Unit undertook initial work on a logistic model for the data. Andrew Presland, House of Commons's Library.

References

1. Mackintosh JM. Snow—the man and his times. *Proc R Soc Med* 1955;**48**:1004—7.
2. Farr W. Influence of elevation on the fatality of cholera. *J Stat Soc Lond* 1852;**15**:155—83.
3. Snow J. *On the mode of the communication of cholera*. London: John Churchill; 1849.

4. Snow J. On the pathology and mode of the communication of cholera. *Lond Med Gazette* 1849;745—52. see also p. 923-9.
5. Registrar-General. *Report on the mortality of cholera in England 1848—49*. Her Majesty's Stationery Office; 1852.
6. Farr, W. Weekly return of the Registrar-General of births and deaths in London. 1853;**14**:429—32.
7. General Board of Health. *Report of the committee for scientific inquiries in relation to the cholera epidemic of 1854*. Her Majesty's Stationery Office; 1855. p. 48.
8. Registrar-General. *Report on the cholera epidemic of 1866*. Her Majesty's Stationery Office; 1868.
9. Eyler JM, Farr W. *Willaim Farr (1807—1883): An intellectual biography of a social pathologist*. University of Wisconsin; 1971.