Epidemiol. Infect. (2009), **137**, 810–820. © 2008 Cambridge University Press
doi:10.1017/S0950268808001477 Printed in the United Kingdom

An ecological analysis of sociodemographic factors associated with the incidence of salmonellosis, shigellosis, and *E. coli* O157:H7 infections in US counties

M. CHANG[1]*, S. L. GROSECLOSE[2], A. A. ZAIDI[2] AND C. R. BRADEN[3]

[1] *Division of Integrated Surveillance Systems and Service, National Center for Public Health Informatics, Coordinating Center for Health Information and Service, Centers for Disease Control and Prevention, Atlanta, GA, USA*
[2] *Division of STD Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD and TB Prevention, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA*
[3] *Division of Foodborne, Bacterial and Mycotic Diseases, National Center for Zoonotic, Vector-Borne and Enteric Diseases, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA*

(*Accepted 15 September 2008; first published online 24 October 2008*)

SUMMARY

Identifying county-level sociodemographic and economic factors associated with the incidence of enteric disease may provide new insights concerning the dynamics of community transmission of these diseases as well as opportunities for prevention. We used data from the National Notifiable Diseases Surveillance System, the U.S. Census Bureau, and the Health Resources and Services Administration to conduct an ecological analysis of 26 sociodemographic and economic factors associated with the incidence of salmonellosis, shigellosis, and *E. coli* O157:H7 infections in US counties for the period 1993 to 2002. Our study indicates that race, ethnicity, place of residence, age, educational attainment, and poverty may affect the risk of acquiring one of these enteric bacterial diseases. The lack of specificity of information regarding salmonellae and shigellae serotypes may have led to less specific associations between community-level determinants and reported incidence of those diseases. Future ecological analyses should use serotype-specific data on incidence, which may be available from laboratory-based surveillance systems.

Key words: Demographic factors, ecology, *Salmonella* infections, shigellosis, surveillance.

INTRODUCTION

Salmonellosis, shigellosis, and *Escherichia coli* O157:H7 infection are the three most commonly reported nationally notifiable enteric bacterial diseases in the United States [1, 2]. Between 1993 and 2002, the annual incidence rates (cases per 100 000 persons) for salmonellosis, shigellosis, and *E. coli* O157:H7 infection in the US ranged from 14·5 to 17·7, from 6·4 to 12·5, and from 1·0 to 1·8, respectively [3]. These pathogens are typically transmitted via food or directly from an infected animal or human. Investigations of outbreaks have found that infections caused by these pathogens may be associated with poor personal hygiene, improper infection control practices within nursing homes or day-care centres, and inappropriate production or preparation of food (e.g. inadequate cooking or keeping food at the wrong holding temperature) [4–7].

* Author for correspondence: M. Chang, M.P.H., National Office of Public Health Genomics, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, 4770 Buford Highway, MS K-89, Atlanta, GA 30341, USA.
(Email: mchang@cdc.gov)

In recent years, several societal and behavioural factors that contribute to the epidemiology of enteric diseases have changed [8, 9]. For example, increasing numbers of people are eating raw or uncooked foods as they pursue healthier lifestyles. In addition, new methods of food production have been implemented and networks for food distribution have expanded. More day-care facilities and nursing homes have been established for the increasing number of children and older people requiring care in institutionalized settings. These changes in society and industry have increased the potential risk of exposure to the pathogens described – often by increasing the opportunity for contaminated food or for person-to-person exposure.

For the organisms of interest, only limited data has been collected and reported at the national level on demographic or other risk factors. Previous analyses of surveillance data on enteric diseases have demonstrated considerable variability in incidence by demographic and socioeconomic risk factors, such as age [10–12], race/ethnicity [1, 13–15], sex [1, 11, 16], educational attainment [17], poverty status [18, 19], household composition and size [18, 20–22], and geographic distribution [1, 23, 24]. Many of these analyses have used individual-level factors to identify risk factors for infection, but ecological analysis, which focuses on groups, can also be useful [25–28], as it may identify community-level factors that are associated with the risk of enteric illness. Identifying sociodemographic and economic factors associated with the incidence of enteric disease may lead to new hypotheses concerning vehicles and routes of disease transmission in the community and interventions that may prevent transmission. We conducted an ecological study to identify community-level sociodemographic factors associated with county-specific incidence rates for salmonellosis, shigellosis, and *E. coli* O157:H7 infection in the United States.

METHODS

Data and sources

We analysed data from the National Notifiable Diseases Surveillance System (NNDSS) that was voluntarily reported to the Centers for Disease Control and Prevention (CDC) from state health departments and the health departments of New York City and the District of Columbia (DC) from 1993 to 2002 for salmonellosis and shigellosis and from 1995 to 2002 for *E. coli* O157:H7 [29]. Salmonellosis and shigellosis were designated as nationally notifiable throughout the study period; *E. coli* O157:H7 became nationally notifiable in 1994. Although surveillance data was available for *E. coli* O157:H7 from 1993 and 1994, we excluded it to allow national reporting practices for this infection to stabilize. Our analysis included cases reported from states and other jurisdictions (New York City, DC) in which the disease was reportable by law or statute. For *E. coli* O157:H7 infection, data from states in which this disease was not reportable by law or statute (12 states in 1995 and six states in 1996) were excluded from the analysis. For salmonellosis and shigellosis, data were reportable by law or statute from all states during 1993–2002 and were included in the analysis. Surveillance data reported from the US territories were excluded.

County-specific sociodemographic, economic, and occupational data collected by the U.S. Bureau of the Census for 2000 and data on the health-care workforce and indexes of capacity collected by the Health Resources and Services Administration were used as the independent variables in the analysis (Table 1) [30–34]. To examine the incidence of disease by geographic distribution, counties were categorized into four US regions (Northeast, Midwest, South, and West) [35].

Data analysis

County- and disease-specific mean incidence rates for salmonellosis, shigellosis, and *E. coli* O157:H7 infection were calculated as the sum of annual disease-specific case counts reported to NNDSS divided by the sum of the annual county-specific population estimates over the period evaluated in this analysis (10 years for salmonellosis and shigellosis; 8 years for *E. coli* O157:H7 infection). The county-specific mean incidence rates served as dependent variables in the disease-specific models. County-level bridged-race population estimates from the U.S. Bureau of the Census for 1993–2002 were used as denominators to calculate county incidence rates [36]. To avoid using rates that might be unstable for counties with small populations and extreme rates, we excluded data for counties with a population below 1000 or with incidence rates above the 99th percentile of ranked county incidence rates by disease and year (resulting in the exclusion of 1·3%, 1·2%, and 1·1% of US counties for salmonellosis, shigellosis, and *E. coli* O157:H7 infection, respectively).

Table 1. *County-specific sociodemographic, economic, and workforce characteristics used as independent variables in the analysis*

Variable	Data year	Data source (see notes)
% Population aged <5 years	2000	1
% Population aged 5–17 years	2000	1
% Population aged 18–44 years	2000	1
% Population aged 45–64 years	2000	1
% Population aged ⩾65 years	2000	1
% Population male	2000	1
% Population black or African American	2000	1
% Population Hispanic or Latino	2000	1
% Adults with less than a ninth-grade education	1990	1
% Population below poverty level	1997	1
Square root (nurses and employees of personal care facilities per 100 000 persons)	2000	1
Enrolment in Medicare programme per 100 000 persons	1999	1
% Households with one person	2000	1
% Households with one or more persons aged <18 years	2000	1
% Households with one or more persons aged ⩾65 years	2000	1
% Civilian labour force unemployed	2000	1
Log (reported violent crime rate per 100 000 persons)	1999	1
% Population living on a farm	1990	1
Log [local *per capita* expenditures for education services (including education and libraries)]	1997	2
Square root [local *per capita* expenditures for social services (including health, hospital, and public welfare)]	1997	2
% Population urban	2000	3
Food service employees per 100 000 persons	2000	4
Square root (day-care workers per 100 000 persons)	2000	4
Log (active, non-federal physicians per 100 000 persons +1)	2001	5
Square root (community hospital beds per 100 000 persons)	2001	5
US region (Northeast, Midwest, South, West)	2000	6

1. United States Bureau of the Census. County and City Data Book: 2000 (http://www.census.gov/prod/www/ccdb.html). Accessed 15 January 2004.
2. United States Department of Commerce, Economics and Statistics Administration, United States Bureau of the Census. 1997 Census of Governments, Volume 4, No. 5, Compendium of Government Finances (http://www.census.gov/prod/gc97/gc974-5.pdf). Accessed 25 June 2004.
3. United States Bureau of the Census. Census 2000 summary file 1 100-percent data, Detail tables (http://factfinder.census.gov/servlet/DTGeoSearchByListServlet?ds_name=DEC_2000_SF1_U&state=dt&mt_name=DEC_2000_SF1_U_P002&_lang=en&_ts=97683722680). Obtained using CDC WONDER.
4. United States Bureau of the Census. Census 2000 special equal employment opportunity (EEO) file (http://www.eeoc.gov/stats/census/index.html). Obtained using CDC WONDER.
5. Area Resource File Access System 2003. Health Resources and Services Administration, Bureau of Health Professions, National Center for Health Workforce Analysis. Rockville, Maryland. Prepared by: Quality Resource Systems, Inc, Fairfax, Virginia (http://www.arfsys.com).
6. United States Bureau of the Census. Geographic terms and definitions (http://www.census.gov/popest/geographic/estimates_geography.html). Accessed 18 August, 2006.

We calculated Spearman's rank correlation coefficients to examine the associations between the county-level sociodemographic variables and the county-specific 10-year mean annual incidence rates for each study condition. Simple linear regression analyses were performed between mean annual incidence rates for each study condition and each of the 26 independent sociodemographic variables. To stabilize the variance of the independent variables and to normalize their distribution, independent variables were transformed by taking either the square root or natural log of the value. If the value of the independent variable equalled 0, 1 was added to the value before taking the natural log. Extreme values were excluded to reduce the variance in the model. Multivariate linear regression analysis was

Table 2. *Distribution of US counties by mean annual number of reported cases of salmonellosis, shigellosis, and* E. coli *O157:H7 infections, National Notifiable Diseases Surveillance System, 1993–2002*

Mean reported cases per year	Salmonellosis Reporting counties No. (%)	Shigellosis Reporting counties No. (%)	*E. coli* O157:H7 infection* Reporting counties No. (%)
0 to <1	614 (19·5)	1258 (40·0)	1373 (43·7)
1 to <10	1768 (56·3)	1061 (33·8)	627 (19·9)
10 to <100	610 (19·4)	351 (11·2)	68 (2·2)
⩾100	78 (2·5)	47 (1·5)	0 (0)
No reports	73 (2·3)	426 (13·6)	1075 (34·2%)

* Includes data reported from 1995 to 2002.

also conducted for each selected condition by using a forward stepwise regression procedure with *P* values of 0·25 and 0·05 as thresholds for a variable to enter the model and stay in the model, respectively [37].

Because no adequate county-specific data on rates of food service employees or of day-care workers was available for counties with population <50 000, these two variables were excluded in the multivariate analysis. The independent variables for which county-specific data was missing for >5% of the counties were excluded in the multivariate analysis: these included reported violent crime rate, local *per capita* expenditures for social services, local *per capita* expenditures for education services, and rates of food service employees and day-care workers. To avoid collinearity between independent variables in fitting reliable regression models, several variables were dropped from the models, including percentage of the population aged 18–44 years, percentage of the households with one or more persons aged <18 years, and percentage of the households with one or more persons aged ⩾65 years.

RESULTS

From 1993 to 2002, a total of 403 464 cases of salmonellosis were reported from 3070 US counties, 234 148 cases of shigellosis were reported from 2717 US counties, and 26 411 cases of *E. coli* O157:H7 infection were reported from 2068 US counties (1995–2002) (see Table 2 for the distribution of US counties by annual number of reported cases). Twenty percent of US counties reported a mean of <1 salmonellosis case per year, for shigellosis and *E. coli* O157:H7 infection the comparable figures were 40% and 44%, respectively (Table 2). One-third (1075) of US counties had no reports on *E. coli* O157:H7 infection during 1995–2002, far more than the 73 (2%) and 426 (14%) counties that did not report salmonellosis and shigellosis cases, respectively, during 1993–2002. For those counties reporting at least one case during the study period, the average county-specific annual incidence rates were 13·5 (median 12·0, range 0·8–49·0), 6·6 (median 4·0, range 0·2–57·5), and 2·1 (median 1·4, range 0·0–15·9) per 100 000 persons for salmonellosis, shigellosis, and *E. coli* O157:H7 infections, respectively.

The highest incidence rates for salmonellosis, shigellosis, and *E. coli* O157:H7 infection were observed in counties in the Northeast, South, and West, respectively (Fig. 1). Across all four regions, higher incidence rates for salmonellosis and shigellosis were seen in counties where >50% of the population lived in urban areas. In contrast, incidence of *E. coli* O157:H7 was highest in counties where <50% of the population lived in urban communities.

In the Spearman's rank correlation analysis, the incidence of salmonellosis was moderately correlated ($0·2 \leqslant r < 0·3$) with the percentage of the population that was black or African American ($r=0·2$), the physician rate per 100 000 persons ($r=0·2$), and the percentage of the population aged 45–64 years ($r=-0·2$). The three leading correlated factors for incidence of shigellosis were percentage of the population Hispanic or Latino ($r=0·3$), percentage of the population aged <5 years ($r=0·3$), and the percentage aged 45–64 years ($r=-0·3$). For *E. coli* O157:H7 infection, these factors were the percentage of the population that was black ($r=-0·5$), residence in the South ($r=-0·5$), and percentage of the population living on a farm ($r=0·4$).

In the simple linear regression analysis, salmonellosis and shigellosis were generally similar to each other in their patterns of positive and negative associations with selected sociodemographic factors (Table 3). Many sociodemographic and economic

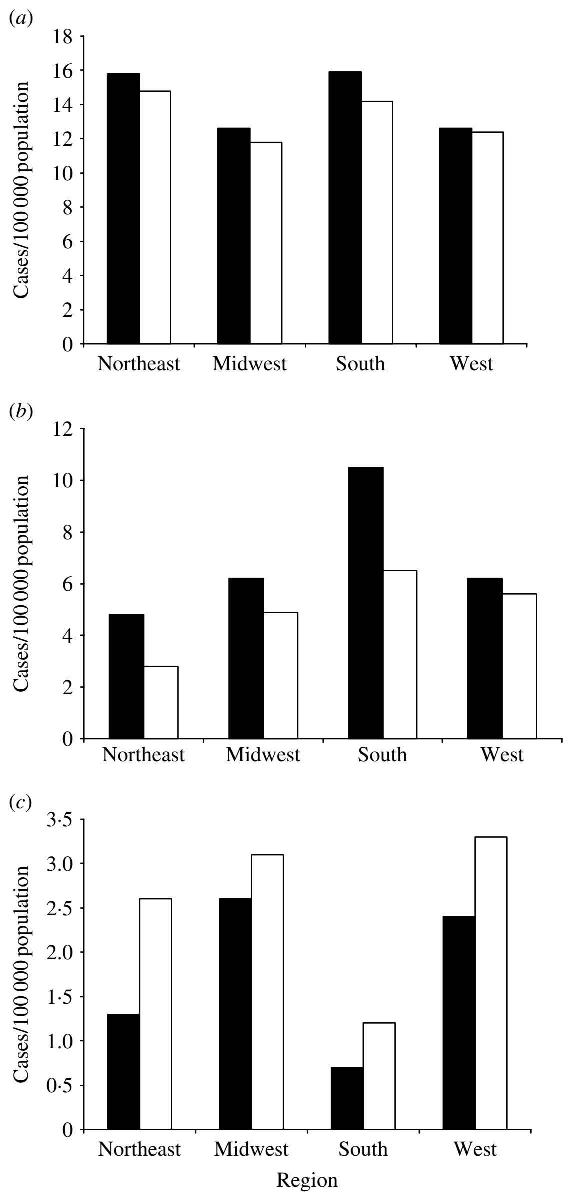

Fig. 1. Mean annual incidence rate of (*a*) salmonellosis, (*b*) shigellosis, and (*c*) *E. coli* O157:H7 infection by percent urban population and US region, National Notifiable Diseases Surveillance System, 1993–2002 (■, ⩾50% urban population; □, <50% urban population). The Northeast region includes: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont, New Jersey, New York, and Pennsylvania. The Midwest region includes: Illinois, Indiana, Michigan, Ohio, and Wisconsin, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota. The South region includes: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, and Tennessee, Arkansas, Louisiana, Oklahoma, and Texas. The West region includes: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming, Alaska, California, Hawaii, Oregon, and Washington. Panel (*c*) includes data from 1995 to 2002.

factors (e.g. population distribution by selected age groups, race, ethnicity, urbanization, poverty level, crime rate, and physician rate) were positively associated with the incidence of salmonellosis and shigellosis but negatively associated with the incidence of *E. coli* O157:H7 infection. In contrast, population distribution by education level, population living on a farm, local *per capita* expenditures for education, and Medicare enrolment rates showed inconsistent associations with these three diseases.

In the multivariate regression analysis, the sociodemographic and economic variables included in Table 4 accounted for 12%, 17%, and 33% of the variation in incidence of salmonellosis, shigellosis, and *E. coli* O157:H7 infection in US counties, respectively. Much of the attributed variation was due to the three leading factors for each condition. For salmonellosis, the percentage of the population that was black, the percentage unemployed (negative association), and percentage of the population that was Hispanic or Latino accounted for 7% of the total variation. For shigellosis, the three leading factors were percentage of the population aged <5 years, percentage of population below poverty level, and percentage unemployed (negative association), and these three factors accounted for 12% of the variation. For *E. coli* O157:H7 infection, the leading factors were percentage of population living on a farm, percentage of adults with less than a ninth-grade education (negative association), and residence in the South (negative association), which accounted for 33% of the variation.

DISCUSSION

In this analysis we found that variation in the incidence of salmonellosis, shigellosis, and *E. coli* O157:H7 infection in US counties was due in part to a diverse set of sociodemographic and economic factors, illustrating the complex relationship between community characteristics and the dynamics of disease transmission. During the study period, salmonellosis had a higher incidence and was more widely dispersed geographically than shigellosis or *E. coli* O157:H7 infection. The county-level characteristics most closely associated with incidence of these enteric diseases included measures of race, ethnicity, place of residence, age group, poverty, unemployment, and urbanization. The variation in incidence rates attributed to these county-level variables ranged from only 12% for salmonellosis to 33% for *E. coli* O157:H7.

https://doi.org/10.1017/S0950268808001477 Published online by Cambridge University Press

Table 3. *Statistical associations* of selected sociodemographic factors with incidence of salmonellosis, shigellosis, and* E. coli *O157:H7 infection for US counties, 1993–2002*

Variable	Salmonellosis	Shigellosis	*E. coli* O157:H7 infection
% Population aged <5 years	+	+	−
% Population aged 5–17 years	×	+	+
% Population aged 18–44 years	+	+	−
% Population aged 45–64 years	−	−	×
% Population aged ⩾65 years	×	−	+
% Population male	−	×	+
% Population black or African American	+	+	−
% Population Hispanic or Latino	+	+	−
% Adults with less than a ninth-grade education	×	+	−
% Population below poverty level	+	+	−
Square root (nurses and employees of personal care facilities per 100 000 persons)	×	−	+
Enrolment in Medicare programme per 100 000 persons	×	−	+
% Households with one person	×	×	+
% Households with one or more persons aged <18 years	+	+	−
% Households with one or more persons aged ⩾65 years	×	−	+
% Civilian labour force unemployed	−	×	−
Log (reported violent crime rate per 100 000 persons)	+	+	−
% Population living on a farm	−	−	+
Northeast region	+	−	×
Midwest region	−	−	+
South region	+	+	−
West region	−	×	+
% Population urban	+	+	−
Log (local *per capita* expenditure for education)	×	×	+
Square root (local *per capita* expenditure for social services)	×	×	×
Square root (day-care workers per 100 000 persons)	×	×	+
Food service employees per 100 000 persons	+	+	+
Log (active, non-federal physicians per 100 000 persons +1)	+	+	−
Square root (community hospital beds per 100 000 persons)	+	+	+

+, Positive association; −, negative association; ×, no association.
* From univariate regression models. Significance was assessed using a P value of $\leqslant 0{\cdot}05$.

In general, salmonellosis and shigellosis had similar group-level associations in the county sociodemographic, economic, and workforce characteristics evaluated. Geographically, the incidence of salmonellosis was higher in the Northeast and South, and the incidence of shigellosis was higher in the South. In contrast, the incidence of *E. coli* O157:H7 was higher in the West and Midwest regions and lowest in the South. These findings suggest that the incidence of both salmonellosis and shigellosis was higher in counties with higher urban populations in the Eastern coast region in which communities have more health-care facilities and more physicians available, and have better access to medical care. In contrast, the incidence of *E. coli* O157:H7 was consistently higher in counties with a higher percentage of the population living on a farm or in non-urban settings in the US Mountain region. This last association (at least in terms of farms) may be due to more direct or indirect contact with cattle or other ruminant animals, the primary reservoir for *E. coli* O157:H57 [38]. The physician rate per 100 000 population, a surrogate measure of access to health care accounted for <2% of the variation in the incidence of salmonellosis and <0·5% of the variation in the incidence of shigellosis or *E. coli* O157:H7 (Table 4). Factors associated with greater health-care resources (such as rates of physicians and community hospital beds) may result in higher rates of diagnosis and case reporting, but in our study the explanatory value of the physician rate was quite small, as noted.

We found that the incidence of salmonellosis was higher in communities with a higher percentage of children aged <5 years or a greater percentage of

https://doi.org/10.1017/S0950268808001477 Published online by Cambridge University Press

Table 4. *Multivariate regression models* of selected sociodemographic factors with incidence of salmonellosis, shigellosis, and* E. coli *O157:H7 infection for US counties, 1993–2002*

Salmonellosis				Shigellosis				*E. coli* O157:H7 Infection			
Variable	Parameter estimate	Standard error	Partial R^2	Variable	Parameter estimate	Standard error	Partial R^2	Variables	Parameter estimate	Standard error	Partial R^2
% Population black or African American	0·0850	0·0112	0·0324	% Population aged <5 years	1·4170	0·2176	0·0583	% Population living on a farm	0·1205	0·0100	0·2078
% Civilian labour force unemployed	−0·3637	0·0612	0·0194	% Population below poverty level	0·2978	0·0423	0·0331	% Adults with less than a ninth-grade education	−0·0260	0·0077	0·0734
% Population Hispanic or Latino	0·0624	0·0133	0·0171	% Civilian labour force unemployed	−0·4069	0·0715	0·0263	South region	−0·3729	0·1263	0·0179
Log (active, non-federal physicians per 100 000 persons +1)	0·8167	0·1629	0·0140	% Population Hispanic or Latino	0·0930	0·0135	0·0077	% Population urban	−0·0118	0·0023	0·0060
% population aged ⩾65 years	0·2552	0·0468	0·0134	% Households with one person	0·2471	0·0580	0·0053	Log (active, non-federal physicians per 100 000 persons +1)	0·2546	0·0623	0·0046
Midwest region	−3·0156	0·3776	0·0079	South region	3·9085	0·4985	0·0044	West region	0·6555	0·1388	0·0045
% Population aged 45–64 years	−0·4052	0·0731	0·0047	Midwest region	1·8085	0·4436	0·0043	% Population aged 5–17 years	0·0969	0·0265	0·0030
% Adults with less than a ninth-grade education	−0·0687	0·0205	0·0027	% Adults with less than a ninth-grade education	−0·0951	0·0251	0·0041	% Households with one person	0·0862	0·0172	0·0021
% Population urban	−0·0212	0·0066	0·0022	Log (active, non-federal physicians per 100 000 persons +1)	0·8175	0·1670	0·0025	% Population below poverty level	−0·0396	0·0126	0·0020
West region	−2·0402	0·5155	0·0017	% Population aged 5–17 years	0·3315	0·1014	0·0024	% Population aged 45–64 years	−0·0655	0·0207	0·0017
% Population aged <5 years	0·4442	0·1908	0·0017	% Population male	0·2355	0·0087	0·0024	% Population black or African American	−0·0113	0·0045	0·0016
Square root (community hospital beds per 100 000 persons)	0·0334	0·0157	0·0014								

* Models were built using a forward selection stepwise regression procedure with a P value of ⩽0·05 as a significant threshold to retain a variable in the model. Variables were excluded from the analysis if: (1) more than 5% of the county-specific data on variables in Table 1 were not available including reported violent crime rate, local *per capita* expenditures for social services, local *per capita* expenditures for education services, and rates of food service employees and day-care workers. (2) To avoid collinearity between variables, percentage of the population aged 18–44 years, percentage of the households with one or more persons aged <18 years, and percentage of the households with one or more persons aged ⩾65 years were also excluded from the analysis.

persons aged $\geqslant 65$ years. The higher incidence of salmonellosis at the extremes of age may be due to these groups tending to get more severe *Salmonella* infections. Patients with more severe infections may be more likely to seek medical care and be diagnosed and reported. In addition, parents may be more likely to seek medical care for their young child with a diarrhoeal disease than they would for themselves [39]. Furthermore, the elderly and children may have better access to care than other age groups due to higher insurance coverage rates [40]. Last, the relatively higher incidence of salmonellosis at the extremes of age may also be due to factors not evaluated in this analysis, such as greater susceptibility of the host or certain environmental exposures.

Slightly higher incidence of salmonellosis was reported from communities with more black or Hispanic residents, a finding that may be due to socioeconomic and cultural differences, knowledge and practices of food safety, and personal hygiene in population subgroups. Higher incidence of salmonellosis was reported in blacks than in whites in a state registry-based study [14]; however, this association was not significant at the geographic block group level [17]. For shigellosis, incidence was higher in communities with more children aged <5 years, more residents living below the poverty level, and more Hispanic residents. Reasons for the association of the incidence of shigellosis and salmonellosis with the proportion of racial/ethnic subpopulations in counties are not known [1, 13], but may, in part, relate to higher poverty rates and lower education rates in Hispanic populations compared with other racial/ethnic groups [41, 42]. High shigellosis rates in young children may be attributable to difficulties in teaching and maintaining good hygiene practices (e.g. effective hand washing), lack of acquired immunity to *Shigella* infection, or exposure to congregate settings such as day-care facilities [1, 43, 44]. Finally, because the associations identified were attributed to county populations and not to individuals of certain racial or ethnic groups, other characteristics common to racially or ethnically diverse counties may have been responsible for higher rates of illnesses, but were not included in this analysis.

Previous epidemiological studies have demonstrated a higher incidence of enteric disease in demographic groups with lower socioeconomic status [45–48]. Our study identified a lower incidence of salmonellosis and shigellosis in communities with higher unemployment. Unemployment may limit access to health care and lead to under-diagnosis of these conditions in unemployed persons.

We found that the incidence of all three enteric diseases we investigated was higher in communities with more people educated at or above the ninth-grade level. These findings seem counterintuitive, but other studies examining risk factors for foodborne disease and the prevalence of practices for consuming or handling food have found that people with a college or university degree beyond a bachelor's were more likely to consume undercooked hamburger and to handle raw meat in an unsafe manner than were persons reporting less education [49, 50]. In addition, a meta-analysis of 20 studies assessing the association between consumers' knowledge and practices regarding food safety and their demographics noted that higher-income and more educated persons reported greater consumption of raw foods, less knowledge of hygiene, and poorer practices in terms of cross contamination of food [51]. It seems also possible that the association between lower socioeconomic status and lower incidence of salmonellosis and shigellosis may be due to less access to health-care services and to stool culture in this group (i.e. under-detection, a surveillance artifact) [52]. Individuals of higher education who may also have more discretionary income may eat outside the home more frequently and be more likely to own pets, both of which are previously identified risk factors for salmonellosis [17, 53].

There are several limitations to this study. First, while the epidemiology of salmonellosis and shigellosis may vary by serotype [54], the NNDSS does not differentiate *Salmonella* or *Shigella* species by serotype. Accordingly, we were unable to identify community-level determinants that may account for serotype-specific variation in the incidence of *Salmonella* or *Shigella*. In contrast, the NNDSS monitors a single serotype of *E. coli*, *E. coli* O157:H7, a significant cause of diarrhoea, bloody diarrhoea, and haemolytic uraemic syndrome, thus allowing for a more direct measurement of the variation related to specific exposure factors of this bacterial pathogen. Variation in the specificity of information on the pathogen may have led to less specific associations between the community-level determinants and reported enteric disease incidence for salmonellosis and shigellosis.

Another concern is that the burden of enteric bacterial disease is underreported [55] through the NNDSS [56], a passive surveillance system that relies on physicians and laboratories to report to state and local health departments. Although enteric illnesses can be severe or even fatal, many infected persons may have mild clinical illness and thus not seek care; these persons would be neither diagnosed nor reported through routine surveillance. On the other hand, increased county-specific incidence rates may reflect a true disease outbreak which occurred during the study period or merely reflect surveillance artifact. For example, NNDSS-based disease incidence reported from the counties participating in FoodNet (http://www.cdc.gov/FoodNet/), the Foodborne Diseases Active Surveillance Network of CDC's Emerging Infections Program, may be higher than the incidence reported from other counties due to greater completeness of reporting as a result of FoodNet's active surveillance methodology.

Finally, as is common with ecological analyses, we were not able to adequately assess confounding or bias due to misclassification [57–60]. Theoretically, using the results of ecological studies to make inferences about individual health risks can be problematic. The group-level data approach can address many of the sources of bias due to misspecification of confounders, confounder measurement errors, and the lack of information about the within-group distribution of exposures and potential confounders. Not all community-level determinants were available for all county comparisons in this analysis, which to some degree weakened our assessment of the associations between these determinants and incidence of the diseases of interest. Additionally, the time periods during which the independent community-level variables and the county incidence data were collected were not concordant for several variables. For example, data was collected in 1990 for the percentage of the population living on a farm, in 1997 for local *per capita* expenditures for social services, and in 1999 for the reported violent crime rate per 100 000 persons. In each case, more recent data was not available for the study period.

The perspective of group-level analysis acknowledges the contribution of both individual (indirectly) and community (directly) factors in determining population health status, although not to the extent possible in individual-level analyses (e.g. case-control or cohort studies). Enteric disease risk factor effects are commonly manifest upon contact between infected and susceptible individuals or following exposure to contaminated food. However, some of the exposure–infection relationships missed at the individual level may be demonstrated in ecological analysis [25]. As opposed to the individual level, group-level analysis accounts for community factors in addition to geographic location and population distribution. This study identified several county-level sociodemographic and economic factors associated with the risk of enteric illness which may help identify effective interventions. For example, a strong positive association between a county with a higher percentage of the population living on a farm and *E. coli* O157:H7 incidence could lead to recommendations at the county level, such as health departments in counties with a higher percentage of population living on a farm should provide education about prevention of *E. coli* O157:H7 transmission in farm settings. Interventions aimed at controlling the contamination of foods by various pathogens and improving hygienic conditions in certain subpopulations or specific occupations may reduce the risk of enteric diseases. The variable associations noted between county sociodemographic factors and the incidence of salmonellosis and shigellosis may be due to the lack of specificity with respect to information on serotype for salmonellae and shigellae. Future ecological analyses should use serotype-specific incidence data, which may be available from laboratory-based surveillance systems. Further investigation of the significance of these factors and the mechanisms by which they account for variation in the incidence of these diseases in community-based studies is needed. Counties with high incidence of enteric disease are also likely to be overburdened by other infectious and chronic diseases. Over the long term, addressing larger community issues related to social, legal, economic, and political factors may be necessary to reduce enteric bacterial disease incidence.

ACKNOWLEDGEMENTS

The authors are grateful to the staff of the US state and territorial health departments who support surveillance activities for notifiable diseases. Additionally, we thank the staff of the Division of Integrated Surveillance Systems and Service, National Center for Public Health Informatics, CDC, for maintaining and disseminating state-based data reported to the National Notifiable Diseases Surveillance System. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Gupta A, *et al.*** Laboratory-confirmed shigellosis in the United States, 1989–2002: epidemiologic trends and patterns. *Clinical Infectious Diseases* 2004; **38**: 1372–1377.
2. **Olsen SJ, *et al.*** Surveillance for foodborne-disease outbreaks – United States 1993–1997. *MMWR CDC Surveillance Summaries* 2000; **49**: 1–62.
3. **Centers for Disease Control and Prevention.** Summary of notifiable diseases – United States, 2003. *Morbidity and Mortality Weekly Report* 2005; **52**: 1–85.
4. **Mead PS, *et al.*** Food-related illness and death in the United States. *Emerging Infectious Diseases* 1999; **5**: 607–625.
5. **Centers for Disease Control and Prevention.** Preliminary FoodNet data on the incidence of foodborne illnesses – selected sites, United States 2000. *Morbidity and Mortality Weekly Report* 2001; **50**: 241–246.
6. **Slutsker L, *et al.*** Foodborne disease prevention in healthcare facilities. In: Bennett J, Brachman PS, eds. *Hospital Infections*, 4th edn. Philadelphia: Lippincott-Raven Publishers, 1998, pp. 333–341.
7. **Warren BR, Parish ME, Schneider KR.** *Shigella* as a foodborne pathogen and current methods for detection in food. *Critical Reviews in Food Science and Nutrition* 2006; **46**: 551–567.
8. **Hedberg CW, MacDonald KL, Osterholm MT.** Changing epidemiology of food-borne disease: a Minnesota perspective. *Clinical Infectious Diseases* 1994; **18**: 671–680.
9. **Kaferstein F, Abdussalam M.** Food safety in the 21st century. *Bulletin of the World Health Organization* 1999; **77**: 347–351.
10. **Wang XY, *et al.*** Trend and disease burden of bacillary dysentery in China (1991–2000). *Bulletin of the World Health Organization* 2006; **84**: 561–568.
11. **Ethelberg S, *et al.*** The significance of the number of submitted samples and patient-related factors for faecal bacterial diagnostics. *Clinical Microbiology and Infection* 2007; **13**: 1095–1099.
12. **Koutsotoli AD, *et al.*** Comparing *Shigella* waterborne outbreaks in four different areas in Greece: common features and differences. *Epidemiology and Infection* 2006; **134**: 157–162.
13. **Shiferaw B, *et al.*** Trends in population-based active surveillance for shigellosis and demographic variability in FoodNet sites, 1996–1999. *Clinicial Infectious Diseases* 2004; **38** (Suppl. 3): 175–180.
14. **Arshad MM, *et al.*** A registry-based study on the association between human salmonellosis and routinely collected parameters in Michigan, 1995–2001. *Foodborne Pathogens and Disease* 2007; **4**: 16–25.
15. **Robins-Browne RM.** Seasonal and racial incidence of infantile gastroenteritis in South Africa. *American Journal of Epidemiology* 1984; **119**: 350–355.
16. **Hasin T, *et al.*** Socioeconomic correlates of antibody leavels to enteric pathogens among Israeli adolescents. *Epidemiology and Infection* 2007; **135**: 118–125.
17. **Younus M, *et al.*** The role of neighborhood level socioeconomic characteristics in *Salmonella* infections in Michigan (1997–2007): assessment using geographic information system. *International Journal of Health Geographics* 2007; **6**: 56.
18. **Simonsen J, Frisch M, Ethelberg S.** Socioeconomic risk factors for bacterial gastrointestinal infections. *Epidemiology* 2008; **19**: 282–290.
19. **Kelly-Hope LA, *et al.*** Geographical distribution and risk factors associated with enteric diseases in Vietnam. *American Journal of Tropical Medicine and Hygiene* 2007; **76**: 706–712.
20. **Ethelberg S, *et al.*** Household outbreaks among culture-confirmed cases of bacterial gastrointestinal disease. *American Journal of Epidemiology* 2004; **159**: 406–412.
21. **Lietzau S, *et al.*** Household contacts were key factor for children's colonization with resistant *Escherichia coli* in community setting. *Journal of Clinical Epidemiology* 2007; **60**: 1149–1155.
22. **Johnson JR, *et al.*** *Escherichia coli* colonization patterns among human household members and pets, with attention to acute urinary tract infection. *Journal of Infectious Diseases* 2008; **197**: 218–224.
23. **Olsen SJ, *et al.*** The changing epidemiology of salmonella: trends in serotypes isolated from humans in the United States, 1987–1997. *Journal of Infectious Diseases* 2001; **183**: 753–761.
24. **Bender JB, *et al.*** Factors affecting surveillance data on *Escherichia coli* O157 infections collected from FoodNet sites, 1996–1999. *Clinical Infectious Diseases* 2004; **38** (Suppl. 3): S157–164.
25. **Koopman JS, Longini Jr. IM.** The ecological effects of individual exposures and nonlinear disease dynamics in populations. *American Journal of Public Health* 1994; **84**: 836–842.
26. **Susser M.** The logic in ecological: I. The logic of analysis. *American Journal of Public Health* 1994; **84**: 825–829.
27. **Susser M.** The logic in ecological: II. The logic of design. *American Journal of Public Health* 1994: **84**: 830–835.
28. **Morgenstern H.** Ecologic studies in epidemiology: concepts, principles, and methods. *Annual Review of Public Health* 1995; **16**: 61–81.
29. **Centers for Disease Control and Prevention.** National Notifiable Disease Surveillance System (http://www.cdc.gov/epo/dphsi/nndsshis.htm). Accessed 24 June 2004.
30. **United States Bureau of the Census.** County and city data book, 2000 (http://www.census.gov/prod/www/ccdb.html). Accessed 15 January 2004.
31. **United States Department of Commerce, Economics and Statistics Administration, United States Bureau of the Census.** 1997 Census of Governments, Volume 4, No. 5, Compendium of Government Finances (http://www.census.gov/prod/gc97/gc974-5.pdf). Accessed 25 June 2004.

32. **United States Bureau of the Census.** Census 2000 summary file 1 100-percent data, Detail tables (http://factfinder.census.gov/servlet/DTGeoSearchByListServlet?ds_name=DEC_2000_SF1_U&state=dt&mt_name=DEC_2000_SF1_U_P002&_lang=en&_ts=97683722680). Obtained using CDC WONDER.
33. **United States Bureau of the Census.** Census 2000 special equal employment opportunity (EEO) file (http://www.eeoc.gov/stats/census/index.html). Obtained using CDC WONDER.
34. **Area Resource File Access System 2003.** Health Resources and Services Administration, Bureau of Health Professions, National Center for Health Workforce Analysis. Rockville, Maryland. Prepared by: Quality Resource Systems, Inc., Fairfax, Virginia (http://www.arfsys.com).
35. **United States Bureau of the Census.** Geographic terms and definitions (http://www.census.gov/popest/geographic/estimates_geography.html). Accessed 18 August 2006.
36. **United States Department of Health and Human Services (US DHHS), Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS).** Bridged-Race Population Estimates, United States, 1990–2002, by age groups. Compiled from the 1 April 2000 resident population developed by the Bureau of the Census in collaboration with the NCHS on CDC WONDER Online database.
37. **Kleinbaum DG, Kupper LL, Muller KE.** *Applied Regression Analysis and Other Multivariable Methods*, 2nd edn. Boston: PWS-KENT Publishing Company, 1988.
38. **Crump JA, *et al.*** An outbreak of *Escherichia coli* O157:H7 infections among visitors to a dairy farm. *New England Journal of Medicine* 2002; **347**: 555–560.
39. **Centers for Disease Control and Prevention.** Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food – 10 states, 2007. *Morbidity and Mortality Weekly Report* 2008; **57**: 366–370.
40. **Heyman KM, Schiller JS, Barnes P.** Lack of health insurance coverage and type of coverage: early release of selected estimates based on data from the 2007 National Health Interview Survey. National Center for Health Statistics. June 2008 (http://www.cdc.gov/nchs/nhis.htm).
41. **Hayes-Bautista DE, Baezconde-Garbanati L, Hayes-Bautista M.** Latino health in Los Angeles: family medicine in a changing minority context. *Family Practice* 1994; **11**: 318–324.
42. **Hayes-Bautista DE, *et al.*** Latino health in California, 1985–1990: implications for family practice. *Family Practice* 1994; **26**: 556–562.
43. **Mohle-Boetani JC, *et al.*** Communitywide shigellosis: control of an outbreak and risk factors in child day-care centers. *American Journal of Public Health* 1995; **85**: 812–816.
44. **Shane AL, *et al.*** Sharing *Shigella*: risk factors of a multi-community outbreak of shigellosis. *Archives of Pediatrics & Adolescent Medicine* 2003; **157**: 601–603.
45. **Olowokure B, *et al.*** Deprivation and hospital admission for infectious intestinal diseases. *Lancet* 1999; **353**: 807–808.
46. **Lee LA, *et al.*** Hyperendemic shigellosis in the United States: a review of surveillance data for 1967–1988. *Journal of Infectious Diseases* 1991; **164**: 894–900.
47. **Ellencweig AY, Slater PE.** Demographic and socioeconomic patterns of hospitalization for infectious diseases in Israel. *European Journal of Epidemiology* 1986; **2**: 83–89.
48. **Cifuentes E, *et al.*** The risk of enteric diseases in young children and environmental indicators in sentinel areas of Mexico City. *International Journal of Environmental Health Research* 2002; **12**: 53–62.
49. **Shiferaw B, *et al.*** Prevalence of high-risk food consumption and food-handling practices among adults: a multistate survey, 1996 to 1997. The Foodnet Working Group. *Journal of Food Protection* 2000; **63**: 1538–1543.
50. **Roseman M, Kurzynske J.** Food safety perceptions and behaviors of Kentucky Consumers. *Journal of Food Protection* 2006; **68**: 1412–1421.
51. **Patil SR, Cates S, Morales R.** Consumer food safety knowledge, practices, and demographic differences: findings from a meta-analysis. *Journal of Food Protection* 2005; **68**: 1884–1894.
52. **Scallan E, *et al.*** Factors associated with seeking medical care and submitting a stool sample in estimating the burden of foodborne illness. *Foodborne Pathogens and Disease* 2006; **3**: 432–438.
53. **Marcus R, *et al.*** Re-assessment of risk factors for sporadic Salmonella serotype Enteritidis infections: a case-control study in five FoodNet Sites, 2002–2003. *Epidemiology and Infection* 2007; **135**: 84–92.
54. **Jones TF, *et al.*** Salmonellosis outcomes differ substantially by serotype. *Journal of Infectious Diseases* 2008; **198**: 109–114.
55. **Jones TF, Scallan E, Angulo FJ.** FoodNet: overview of a decade of achievement. *Foodborne Pathogens and Disease* 2007; **4**: 60–66.
56. **Thomas MK, *et al.*** Estimated numbers of community cases of illness due to *Salmonella*, *Campylobacter* and *Verotoxigenic Escherichia coli*: pathogen-specific community rates. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2006; **17**: 229–234.
57. **Guthrie KA, Sheppard L.** Overcoming biases and misconceptions in ecological studies. *Journal of the Royal Statistical Society* 2001; **164**: 141–154.
58. **Greenland S, Morgenstern H.** Ecological bias, confounding, and effect modification. *International Journal of Epidemiology* 1989; **18**: 269–274.
59. **Brenner H, *et al.*** Effects of nondifferential exposure misclassification in ecologic studies. *American Journal of Epidemiology* 1992; **135**: 85–95.
60. **Greenland S.** Divergent biases in ecologic and individual-level studies. *Statistics in Medicine* 1992; **11**: 1209–1223.