Int J Public Health (2011) 56:15–24
DOI 10.1007/s00038-010-0143-6

ORIGINAL ARTICLE

Toxocara infection in the United States: the relevance of poverty, geography and demography as risk factors, and implications for estimating county prevalence

Peter Congdon · Patsy Lloyd

Received: 20 November 2009 / Revised: 3 March 2010 / Accepted: 26 March 2010 / Published online: 27 April 2010

Abstract
Objective To estimate Toxocara infection rates by age, gender and ethnicity for US counties using data from the National Health and Nutrition Examination Survey (NHANES).
Methods After initial analysis to account for missing data, a binary regression model is applied to obtain relative risks of Toxocara infection for 20,396 survey subjects. The regression incorporates interplay between demographic attributes (age, ethnicity and gender), family poverty and geographic context (region, metropolitan status). Prevalence estimates for counties are then made, distinguishing between subpopulations in poverty and not in poverty.
Results Even after allowing for elevated infection risk associated with poverty, seropositivity is elevated among Black non-Hispanics and other ethnic groups. There are also distinct effects of region. When regression results are translated into county prevalence estimates, the main influences on variation in county rates are percentages of non-Hispanic Blacks and county poverty.
Conclusions For targeting prevention it is important to assess implications of national survey data for small area prevalence. Using data from NHANES, the study confirms that both individual level risk factors and geographic contextual factors affect chances of Toxocara infection.

Keywords Toxocara infection · Poverty · Ethnicity · Geographic prevalence · Bayesian

P. Congdon (✉)
Department of Geography, Center for Statistics,
Queen Mary University of London, London E1 4NS, UK
e-mail: p.congdon@qmul.ac.uk

P. Lloyd
National Minority Quality Forum,
1200 New Hampshire Avenue Ste 575,
Washington, DC, USA
e-mail: plloyd@nmqf.org

P. Lloyd
Department of Epidemiology and Biostatistics,
School of Public Health and Health Services,
George Washington University, Washington, DC, USA

Introduction

Survey data and other prevalence estimates of Toxocara infection in the United States suggest that this public health concern is differentiated with respect to major regional subdivisions, socio-economic status and demographic variables (Won et al. 2008; Plaut et al. 1996; Tolan and Laufer 2009). It is important to understand how these differences are manifested geographically in different small areas since recent studies confirm Toxocara infection is closely related to poverty status in the United States, which is itself spatially concentrated (Hotez 2007, 2008). Poverty status is also differentiated according to race and region, being higher in US non-metropolitan areas, especially in the South and among non-Hispanic Blacks and Native Americans (USDA 2004).

This paper carries out an analysis of Toxocara infection levels in the US with regard to category and region of residence, family poverty status and demographic stratifiers. The analysis is oriented to providing county-level estimates of Toxocara infection; there are just over 3,000 counties in the US with an average population of around 100,000. Like other recent studies (Won et al. 2008), the analysis is based on the National Health and Nutrition Examination Survey (NHANES), one of few sources of

information on US population prevalence of Toxocara infection.

Exposure to Toxocara is common due to soil contamination, commonly in the peri-domestic environment (Habluetze et al. 2003; Mizgajskaa 2001), and is increased by poverty, poor hygiene, the infection rate in dogs, and risk of potential contact. For example, toddlers playing in contaminated park areas are at heightened risk of infection, and certain occupations such as farmers may be at greater risk (Won et al. 2008; Deutz et al. 2005). In the US there is evidence of higher prevalence of infection among those living in non-metropolitan areas (Won et al. 2008). Non-US studies have also found lower infection rates in both humans and dogs in urban settings (Deutz et al. 2005). However, as discussed below, reported associations between geographic residence and infection risk may be confounded by interactions between risk factors.

In this paper, we analyze Toxocara infection variations using regression models which incorporate both person and family level risk factors and geographic modifiers. The latter include differences in infection risk between four US regions, and between urban and rural subdivisions of those regions. Since the objective is to produce county-level estimates of infection in the US, the survey regression is oriented to provide infection rate estimates for subpopulations for which matching information is available at county level in the US; these are population totals by ethnicity, age and gender, and household totals in each county according to poverty status.

Toxocariasis and population health

Human toxocariasis is a zoonotic helminth infection caused by the parasitic roundworms *Toxocara canis* and *Toxocara cati*, especially the former. Eggs are passed in the stool and need to embryonate, a process usually lasting several weeks (depending on the temperature). Therefore, Toxocara infections are most commonly the result of contact with the soil (environmental contamination), though they may also result from direct contact with pets, or indirectly by eating unwashed egg-contaminated raw vegetables or meat. Recent studies also show Toxocara eggs present in dog fur, but the epidemiological consequences of these findings are not yet clear (Roddie et al. 2008). Once infected, larvae can migrate throughout the body and have been found in every tissue and organ system.

Toxocara infection is under-diagnosed since infection is generally asymptomatic, yet it may lead to severe conditions, more frequently observed in children than in adults, including visceral larva migrans and ocular larva migrans (Despommier 2003). Other clinical manifestations, often with long pre-latency periods, include so-called covert toxocariasis, with outcomes including neurological lesions and disorders, asthma and abdominal pain (Nelson et al. 1996; Nathwani et al. 1992; Magnaval et al. 1997; Cooper 2008; Macpherson 2005).

Given the diversity of possible clinical outcomes, obtaining geographic area data on disease cases linked to Toxocara infection are relatively complex. However, to indicate those areas and communities at high risk of infection and in need of prevention measures in the United States, an estimation of Toxocara infection rates at a disaggregated area scale is warranted.

Methods

Data sources and variables

The analysis here to develop such small area estimates uses seroprevalence data from the NHANES III survey (1988–1994). The diagnosis of toxocariasis depends upon serological tests, as neither eggs nor larvae occur in the feces of human hosts (Lynch et al. 1988). Surplus sera were collected from a sample of the NHANES III participants aged 6 or over, and tested for antibodies to Toxocara in an enzyme immunoassay (EIA) using *T. canis* excretory-secretory (TES) antigens from infective-stage larvae (CDC 2007; Hotez and Wilkins 2009). Results from the EIA measure total immunoglobulin antibodies and are reported as a titer; the assay detects infections caused by both *T. canis* and *T. cati*. All results are reported in binary form, namely positive or negative. A positive result indicates an infection with *Toxocara* sp. at some point in time, i.e., either a current or past infection (Hotez 2008, p. 7; Cooper 2008, p. 552).

The analysis of the survey data (i.e., of the binary seropositivity responses and relevant risk factors) is specifically oriented to producing county-level estimates of prevalence for subpopulations for which information is routinely available for US counties. For the purpose of estimating infection risk, age is accordingly categorized into nine age-bands (6–9, 10–19, …, 70–79, 80 and older), while ethnicity has four categories: White non-Hispanic, Black non-Hispanic, Hispanic, other non-Hispanic.

As Hotez (2008) demonstrates, there is a strong poverty effect for the risk of Toxocara infection. Here, the poverty index ratio (PIR), namely the ratio of family income to a poverty threshold produced by the Census Bureau, is used to determine binary poverty status, distinguishing between those in poverty with ratios less than 1 from remaining families.

Geographic context is available in terms of region of residence, namely 1 = Northeast, 2 = Midwest, 3 = South and 4 = West (see Table 1, and the metropolitan status of the county of residence: metropolitan counties are

Table 1 Census region definitions

Northeast region	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
Midwest region	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
South region	Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia
West region	Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming

defined as central counties of metropolitan areas of over 1 million population, and all others classified as non-metropolitan areas.

Data were included from 20,396 participants with complete or partly missing data. Missing data are confined to the PIR, with around 15.8% of males, and 6.6% of females having missing values. The main regression to estimate prevalence risk is therefore preceded by a model to estimate missing PIR values. Appendix 1 discusses estimation of missing PIR values, making use of the positive correlation between PIR and two other measures of socioeconomic status (occupational prestige and education).

Form of regression model

While recent research highlights poverty as an important risk factor, tabulations from the NHANES III data suggest that geographic context (i.e., region and metropolitan status) may also be relevant. Thus, Table 2 (which includes adjustment for missing poverty status) shows that males living in non-metropolitan areas in the Northeast and South, but not in poverty, have higher infection rates than males living in poverty in the West and Midwest. Similarly, ethnicity has an apparent effect beyond the impact of poverty. For example, Table 3 shows that Black non-Hispanic males not in poverty have a higher infection rate than Hispanic males in poverty.

However, a regression analysis is required to establish more conclusively the effects of geographic context, ethnicity and poverty by controlling for the inter-correlation between risks of different kinds. Of particular interest is to establish whether contextual effects remain after controlling for individual variables (or "compositional" effects) (Merlo et al. 2005; Adams et al. 2009). It is also important to assess whether contextual effects vary by the demographic or social group of individuals, namely whether there are interactions between context and individual variables. For example, Morenoff and Lynch (2004, p. 438) argue that research on how contextual effects vary across demographic subgroups may yield new insights into the connection between environment and health.

The survey regression model used to estimate the presence of Toxocara antibody includes age group, ethnicity, poverty status, census region, and type of area (metropolitan vs. non-metropolitan). Separate regression models were used for males and females to allow for gender-specific effect modification. The regression models incorporate differential survey weights for subjects using a weighted likelihood (Graubard et al. 1997).

To represent the form of the model, let A and E denote age and ethnic categories, respectively, M and R denote metropolitan status and region, respectively, and S denote poverty status ($S = 1$ for family in poverty, $S = 0$ otherwise). The survey data (see Table 3) indicate a stronger poverty effect for non-Hispanic Whites, and so the coefficient on S is taken as ethnic-specific.

As well as main effects in these variables, the model includes interactions $u_1(E, R, M)$ between ethnicity, region and metro status, and interactions $u_2(A, R, M)$ between age, region, and metro status. These account for potential geographic or contextual variation in the impacts of person level demographic variables.

The age effects and interaction terms are modeled as collections of random effects; other parameters are estimated as fixed effects. Models are estimated using a Bayesian approach via the WINBUGS program (Lunn et al. 2000).

To predict the probability p_i of infection, a log link is adopted, namely

$$\log(p_i) = \theta + \alpha A_i + \beta E_i + \kappa R_i + \lambda M_i + \Delta_{E_i} S_i + u_1(E_i, R_i, M_i) + u_2(A_i, R_i, M_i),$$

so that (after exponentiation) coefficients can be interpreted as relative risks on the infection probability (Blizzard and Hosmer 2006).

Obtaining county estimates

The use of a log link in the infection probability regression facilitates the subsequent small area estimation stage. Specifically the survey model coefficients relevant to a particular subpopulation in a particular category of county (e.g., females aged 20–29, of White non-Hispanic ethnicity, living in a Northeast metropolitan county) are adjusted for the observed rate of family poverty in each county.

Table 2 Toxocara infection, geographic context and poverty, totals and rates United States, 1988–1994

Gender	Urbanity	Region	Not in poverty		In poverty		Infection rate by poverty status			Relative risk according to poverty status
			Non-infected	Infected	Non-infected	Infected	All cases	Not in poverty	In poverty	
Males	Non-metro	Northeast	639	172	64	28	0.222	0.212	0.307	1.45
		Midwest	985	168	99	22	0.149	0.146	0.181	1.24
		South	1,493	350	208	123	0.217	0.190	0.372	1.96
		West	593	33	73	16	0.068	0.052	0.176	3.38
		All	3,711	723	444	189	0.180	0.163	0.299	1.83
	Metro	Northeast	803	131	126	34	0.151	0.140	0.211	1.50
		Midwest	886	110	99	17	0.114	0.110	0.144	1.31
		South	940	158	114	33	0.153	0.144	0.222	1.54
		West	1,086	132	118	26	0.116	0.108	0.184	1.70
		All	3,715	530	456	109	0.133	0.125	0.193	1.55
	All male subjects		7,425	1,253	900	298	0.157	0.144	0.249	1.72
Females	Non-metro	Northeast	740	131	89	18	0.152	0.150	0.170	1.13
		Midwest	1,074	96	151	43	0.102	0.082	0.222	2.70
		South	1,666	258	350	122	0.158	0.134	0.259	1.93
		West	583	38	143	15	0.068	0.061	0.094	1.53
		All	4,063	523	734	198	0.131	0.114	0.213	1.87
	Metro	Northeast	825	97	178	36	0.117	0.105	0.170	1.62
		Midwest	880	72	119	23	0.087	0.076	0.164	2.17
		South	892	136	156	64	0.160	0.132	0.292	2.21
		West	1,244	138	117	23	0.106	0.100	0.168	1.68
		All	3,842	443	570	148	0.118	0.103	0.206	1.99
	All female subjects		7,905	965	1,304	346	0.125	0.109	0.210	1.93

Totals using subject survey weights

Table 3 Ethnicity, poverty and Toxocara infection rates United States, 1988–1994

Gender	Ethnicity	Not in poverty		In poverty		Infection rate by poverty status			Relative risk according to poverty status
		Non-infected	Infected	Non-infected	Infected	All cases	Not in poverty	In poverty	
Males	White non-Hispanic	5,939	903	404	141	0.141	0.132	0.259	1.97
	Black non-Hispanic	606	176	208	86	0.243	0.225	0.292	1.30
	Hispanic	621	101	253	55	0.151	0.139	0.178	1.27
	Other	259	74	35	16	0.235	0.222	0.319	1.44
	All males	7,425	1,253	900	298	0.157	0.144	0.249	1.72
Females	White non-Hispanic	6,382	640	637	154	0.102	0.091	0.195	2.14
	Black non-Hispanic	717	146	324	98	0.190	0.169	0.233	1.38
	Hispanic	615	85	288	69	0.146	0.122	0.194	1.59
	Other	191	94	54	25	0.327	0.331	0.311	0.94
	All females	7,905	965	1,304	346	0.125	0.109	0.210	1.93

Thus let P_j denote the poverty rate in county j, and let R_j and M_j denote the region and metropolitan status of that county. Then for a given gender, age and ethnic-specific infection rates $r(A, E)$ are derived using two sets of rates. One set is for people in each county in poverty, so that their estimated infection rate includes the poverty effect (specific to ethnicity E) Δ_E, namely

$$r_{pj}(A,E) = \exp\big(\theta + \alpha A + \beta E + \kappa R_j + \lambda M_j + \Delta_E + u_1\big(E, R_j, M_j\big) + u_2\big(A, R_j, M_j\big)\big).$$

The other set of rates is for remaining portion of each county's population, namely those not in poverty, so that the poverty effect Δ is not included, namely

$$r_{\mathrm{rj}}(\mathrm{A},\mathrm{E}) = \exp\big(\theta + \alpha\mathrm{A} + \beta\mathrm{E} + \kappa\mathrm{R_j} + \lambda\mathrm{M_j} + u_1\big(\mathrm{E},\mathrm{R_j},\mathrm{M_j}\big) + u_2\big(\mathrm{A},\mathrm{R_j},\mathrm{M_j}\big)\big).$$

The overall age–ethnic infection rate for each county is then obtained as a weighted average according to the poverty composition in each county:

$$r_{\mathrm{j}}(A,E) = P_{\mathrm{j}} r_{\mathrm{pj}}(A,E) + \big(1 - P_{\mathrm{j}}\big) r_{\mathrm{rj}}(A,E).$$

Numbers of prevalent cases in each county may be estimated using county populations $N_{\mathrm{j}}(A, E)$ in conjunction with the model rates $r_{\mathrm{j}}(A, E)$.

To form overall county rates for each ethnic group, standard age weights w_{A} (e.g., from the US 2,000 population) may be used, so that

$$r_{\mathrm{j}}(E) = \sum_A w_A r_{\mathrm{j}}(A,E).$$

To derive overall rates for each county regardless of ethnic group, ethnic weights w_E from the US 2,000 population are used, so that

$$r_{\mathrm{j}} = \sum_E w_E r_{\mathrm{j}}(E).$$

Results

Regression results

Table 4 shows model results for the separate male and female regressions in terms of relative risks for age and ethnic groups, poverty status, region and metro status. These are fixed effect parameters estimated in relation to a reference group (with relative risk 1), except for age parameters which are modeled as random effects and have an average relative risk of 1.

Table 4 shows that there is a significant poverty effect such that people living in families with income below the poverty threshold have an elevated risk of Toxocara infection. The poverty effect is strongest for White families. Even after allowing for poverty, seropositivity is elevated among Black non-Hispanics and other ethnic groups compared to White non-Hispanics and Hispanics.

As to demographic variation, seroprevalence rates are significantly higher at ages 60–79 among males, though there is virtually no female age gradient. While prevalence is higher among Black and other ethnic groups for both genders, the ranking of these two groups differs between males and females.

Table 4 Estimated relative risks of infection (posterior means, with 95% intervals), by gender United States, 1988–1994

Males			Females		
Mean	2.5%	97.5%	Mean	2.5%	97.5%
Ethnicity					
White (reference)					
1.00			1.00		
Black					
1.64	0.98	2.35	1.48	0.91	2.22
Hispanic					
1.09	0.70	1.57	1.17	0.70	1.71
Other					
1.65	1.12	2.21	2.24	1.44	2.70
Age					
6–9					
0.82	0.60	0.99	0.98	0.86	1.08
10–19					
0.88	0.77	1.00	1.00	0.92	1.09
20–29					
0.92	0.80	1.03	1.02	0.94	1.13
30–39					
0.96	0.81	1.09	1.02	0.96	1.12
40–49					
0.98	0.87	1.09	1.01	0.95	1.10
50–59					
1.03	0.92	1.18	1.00	0.93	1.08
60–69					
1.14	1.00	1.38	1.00	0.92	1.08
70–79					
1.18	1.02	1.43	0.99	0.89	1.07
80+					
1.17	0.99	1.44	0.99	0.87	1.10
Poverty[a]					
Whites					
1.80	1.53	2.09	1.95	1.65	2.28
Blacks					
1.32	1.06	1.62	1.37	1.09	1.69
Hispanic					
1.32	1.03	1.74	1.49	1.11	1.97
Other					
1.55	1.06	2.26	1.31	1.02	1.80
Region					
Northeast (reference)					
1.00			1.00		
Midwest					
0.95	0.65	1.33	0.70	0.36	1.23
South					
1.38	1.02	1.84	1.45	0.87	2.31
West					
0.70	0.45	0.93	0.79	0.47	1.34

Table 4 continued

Males			Females		
Mean	2.5%	97.5%	Mean	2.5%	97.5%
Metro status					
Non-metropolitan (reference)					
1.00					
			1.00		
Metropolitan					
0.99	0.73	1.28			1.81
			1.19	0.76	

[a] Poverty effect specific for ethnicity

Contextual effects

As to contextual effects, Table 4 shows that infection rates are significantly higher for both genders living in the South, and lower for both genders living in the West. These effects remain after controlling for age, ethnicity and poverty, and so appear as genuine contextual effects. On the other hand, the overall non-metropolitan effect seems not to be significant.

More detailed contextual effects are evident from the estimates of the interactions between ethnicity, region and urbanity (Table 5; expressed in terms of relative risks, and limited to significant effects).

These indicate first that significant contextual effects are more evident for males. Second, for both males and females, the low relative risk associated with the West region is concentrated among White non-Hispanics. The third main feature is that the high relative risk associated with the South is concentrated among those of Black non-Hispanic and other ethnic groups, though White non-Hispanic males in the non-metropolitan South also have excess risk. The fourth feature is that for White non-Hispanic males, higher than average risks are confined to non-metropolitan settings, namely in the Northeast and South. So the overall non-metropolitan excess infection risk apparent before regression analysis, and reported by Won et al. (2008, p. 554), may be particularly due to an excess non-metropolitan risk among Whites. The latter form of excess risk appears to be a genuine contextual effect for White males—one remaining after a regression analysis also controlling for impacts of individual level ethnicity and poverty. By contrast, for those of Black and other ethnicity, high relative risks occur in both metropolitan and non-metropolitan settings.

Prevalence maps

Maps displaying the estimated county Toxocara infection rates r_j according to gender are presented in Figs. 1 and 2.

Table 5 Interactions between race and geographic context, significant relative risks (in descending order), United States, 1988–1994

Race	Metro status	Region	Mean relative risk[a]	2.5%	97.5%
Males					
Other	Metro	S	2.75	1.51	4.38
Other	Non-metro	S	2.29	1.08	4.48
Bl-NH	Non-metro	S	2.18	1.52	3.16
Bl-NH	Metro	S	1.78	1.15	2.60
Bl-NH	Non-metro	MW	1.69	1.09	2.54
Other	Metro	W	1.64	1.07	2.47
Bl-NH	Metro	NE	1.61	1.06	2.37
Bl-NH	Metro	MW	1.59	1.09	2.31
Wh-NH	Non-metro	NE	1.45	1.08	1.91
Wh-NH	Non-metro	S	1.41	1.02	2.02
Wh-NH	Metro	MW	0.66	0.47	0.93
Wh-NH	Metro	W	0.58	0.43	0.81
Wh-NH	Non-metro	W	0.37	0.23	0.59
Females					
Other	Metro	S	4.49	2.78	7.00
Other	Metro	W	3.70	2.39	5.60
Other	Non-metro	S	3.60	1.40	9.03
Other	Metro	NE	2.78	1.65	4.50
Bl-NH	Non-metro	S	1.95	1.29	2.92
Bl-NH	Metro	S	1.67	1.04	2.58
Hispanic	Metro	S	1.65	1.00	2.67
Wh-NH	Metro	W	0.50	0.32	0.76
Wh-NH	Non-metro	W	0.38	0.22	0.62

[a] Exponential of combined effect of region, ethnicity, metro status and interaction u_1

The overall population weighted prevalence rates in all 3,141 counties are 14.6 and 12.6% for males and females, respectively. The maps confirm that the level of infection by Toxocara is higher in the South and the Northeast of the US, with lower burden in the West and Midwest (cf. Herrmann et al. 1985, who report lower rates in the West of the US). In terms of the geographic patterning in estimated county rates, there is elevated prevalence especially in higher poverty counties in the South and Northeast, as well as in non-metropolitan counties in the Northeast.

Correlations between the county estimates and rates of poverty and ethnic structure are shown in Table 6. From Table 6, it is apparent that the main determinants of high rates across the country as a whole are poverty levels and percent Black non-Hispanic. However, isolated elevated rates in the West and Midwest (see Figs. 1, 2) also reflect the location of counties with large Native American Indian communities (e.g., in Arizona, New Mexico and Montana). Estimated county prevalence rates also show a clear monotonic gradient according to poverty, as summarized by grouping counties according to poverty decile (e.g., the 10% of counties with the highest rates constitute the

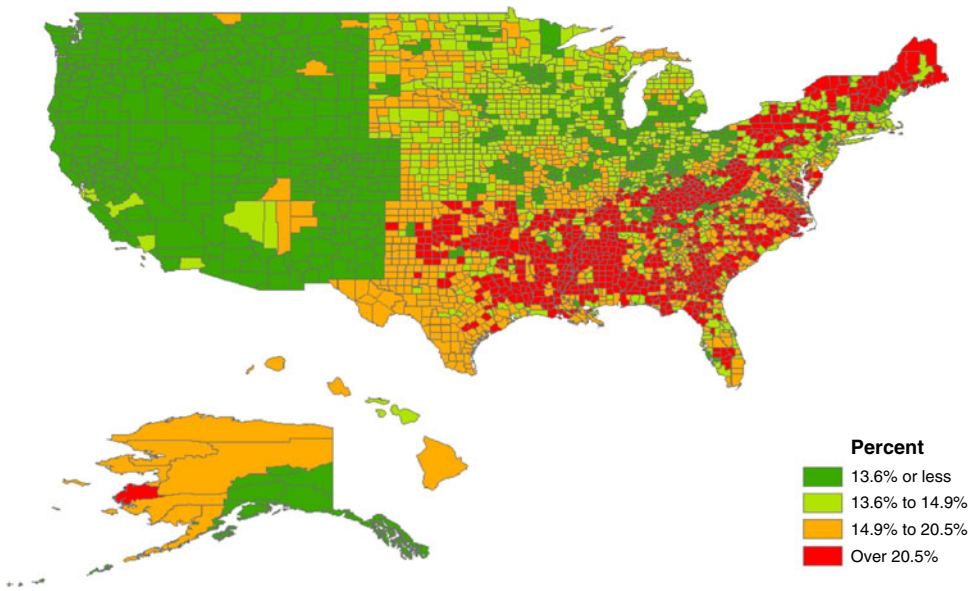

Fig. 1 Estimated county infection rates (percents for males), United States, 1988–1994

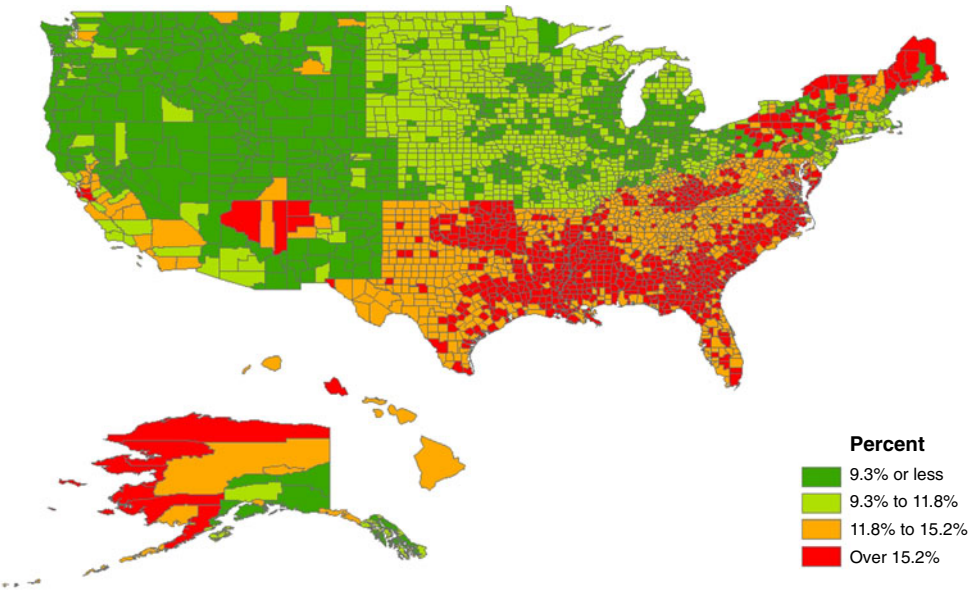

Fig. 2 Estimated county infection rates (percents for females), United States, 1988–1994

highest decile). The male infection rate varies from 12.8% in the lowest poverty decile to 19.3% in the highest poverty decile, while the female rate varies from 10.5 to 16.2%.

Discussion

By incorporating individual and socio-demographic risk factors as well as geographic context to examine infection risk, the analysis presented in this paper provides a distinct approach for using nationally collected seroprevalence data to both evaluate the effect of these factors on infection risk and apply such effects in small area prevalence estimates. The present analysis is distinct from other studies in formally taking account of missing survey data. For example, Won et al. (2008) undertake a logistic regression of NHNANES III data that include survey weights, but do not mention missing data issues, or any procedure used to take account of missing data. In particular, it is important to model missing data on poverty status as this is an important risk factor for seroprevalence (Hotez 2008).

However, even if missingness is allowed for, there are still possible drawbacks in making detailed inferences (i.e., relating to particular subgroups) from survey data where numbers, in particular, demographic and socioeconomic sub-groups may be small. For example, numbers in other ethnic groups in NHNANES III are small (e.g., an unweighted total of 150 among 9,876 male subjects, raised

Table 6 Correlations between prevalence, county poverty and ethnic structure United States, 1988–1994

County variable	Prevalence	
	Male	Female
White (%)	−0.35	−0.52
Black (%)	0.54	0.63
Hispanic (%)	−0.06	0.01
Other (%)	−0.01	0.06
Poverty rate 2000	0.50	0.51

to 385/9,876 after weighting). A possible expedient (adopted by Won et al. 2008) is to exclude other ethnic groups from any regression analysis, but here they are included, while admitting that inferences may be less precise than for other groups.

While poverty is confirmed as an important risk factor in the full regression analysis, seropositivity is elevated among Black non-Hispanics and other ethnic groups even after allowing for poverty. Such differentials are not fully explained (and are possibly under-investigated) in the literature, and inferences for the other group are subject to the caveats mentioned above. However, they may reflect differential contextual exposures, linked inter alia to ethnic residential clustering and segregation (e.g., in disadvantaged urban enclaves), as distinct from the impact of family poverty per se (Hotez 2008 p. 5). Potential impacts of segregation on the residential environment, and risk of infectious disease, in general, are considered by Acevedo-Garcia (2000). Earlier studies by Glickman and Schantz (1981) and Herrmann et al. (1985) mention the higher prevalence of pica (ingestion of non-food substances) and geophagia among black children, though contemporary data are limited. This ethnic differential in risk affects prevalence estimates for counties with large Black non-Hispanic populations, most notably in the South and the Northeast.

The regression also shows higher seroprevalence rates among adult males than females, a differential also reported by Glickman and Schantz (1981), and possibly reflecting differentials in occupational exposures, including occupations involving soil exposures (Won et al. 2008). Significant age effects are absent for females, but are significantly elevated for males aged 60–79. The reasons for this are unclear, but higher proportions of males in this age group (as compared to other ages) are located in non-metropolitan residential settings, and males aged 60–79 in non-metropolitan areas show relatively high seroprevalence rates (over 20%) in the original survey data, possibly reflecting past occupational exposures. An increase in seroprevalence among older adults is also reported by Genchi et al. (1990).

Results also confirm that infection rates are influenced by geographic context as well as socio-demographic risk factors. Rates are highest in the South and lowest in the West, and rates may also differ between type of area within region, with non-metropolitan areas in the Northeast generally having a larger burden of infection. However, some of the non-metropolitan effect detected in other studies (Won et al. 2008) seems not to be retained in a regression model allowing for the interplay between ethnicity, poverty and geographic setting. The apparent excess infection risk of non-metropolitan residence among Black non-Hispanics may be reduced by allowing for family poverty, which is highest among Black groups in the rural South.

While non-metropolitan settings may not, in general, have excess risk in the regression analysis, the reverse is not true either. Thus, there is no evidence of excess risk in densely populated metropolitan settings. In this connection, it is relevant to mention increased awareness of health threats from dog waste in urban areas, and prevention strategies partly following success of legal interventions. In 1978, the Canine Waste Law was enacted in New York City requiring dog owners and walkers to pick up and dispose dog waste, and this example has been followed in other major cities (Brooke 1984; Beck 2000; Macpherson 2005).

The analysis of this paper analysis confirms that seroprevalence in the US is relatively high compared to that in some other developed countries, although the latter estimates are often based on very small samples. The overall US rates of 14.6% (males) and 12.6% (females) are close to those of Won et al. (2008), but higher than seroprevalence rates of 2.4% for Denmark reported by Stensvold et al. (2009), 6.6% among 201 healthy Italians reported by Nicoletti et al. (2008) and 13.7% among healthy Slovakian blood donors reported by Havasiova et al. (1993). Considerably higher prevalences among rural/farm workers in Italy and Austria are reported by Genchi et al. (1990) and Deutz et al. (2005). Relatively high rates in the US may reflect its high poverty rate as compared to many other developed countries (combined with wide socioeconomic disparities), and its ethnic composition given, for example, evidence of higher seroprevalence among Black non-Hispanics. Impacts on geographic variations in Toxocara infection of residential segregation and socio-spatial inequality are, however, an under-researched subject.

By contrast, US seroprevalence rates are lower than reported for some middle income or developing countries, with rates of toxocariasis reaching 40% or higher in Indonesia and Brazil (Noordin et al. 2005; De Andrade Lima Coelho et al. 2005). Application of prevalence modeling to such settings may need to be sensitive to climatic and seasonal effects (e.g., Paquet-Durand et al. 2007), as well as to major demographic variations (e.g., among aboriginal groups) (e.g., Fan et al. 2004). The prevalence estimation method outlined in the current paper

relies upon comprehensive population health survey, and either small area population census data, or intercensal population and poverty estimates for small areas. The NHANES survey in the US is unusual, compared to health surveys even in other developed countries, including serologic tests for Toxocara infection and Hepatitis-C. Given the typically less comprehensive nature of census data for small areas in less developed countries, and the complexity and cost of performing population health surveys in many such societies, simplified prevalence estimation methods will need to be applied in such settings.

The relevance of the current analysis to disease control and public health strategy is apparent from Hotez (2008, p. 7) when he argues that "an important obstacle to the control or elimination of the neglected infections of poverty in the US is the absence of reliable population-based estimates of prevalence and disease burden data about these conditions". Indications of geographic variation as developed in this paper assist in targeted interventions or need assessments, including measuring Toxocara infection levels among marginalized social and demographic groups, or in particular contexts (e.g., in high poverty segregated urban areas), assessing comorbidity between toxocariasis and the growing burden of conditions such as asthma and rheumatoid arthritis (Cooper 2008; Kaplan et al. 2005), preventing indiscriminate deposition of dog and cat feces in parks and play areas (Despommier 2003), routine treatment of non-feral dogs and cats, and control of stray and wild dog populations.

There are some limitations with this analysis. First, some of the risk factors collected in NHANES III are not used, since the goal of the regression models is to provide rates for population subgroups in counties (and possibly lower area scales) where up to date subpopulation estimates are routinely provided and that the estimated rates can be applied to. Though additional risk factors (e.g., dog ownership) may explain variation in infection status, poverty status, race/ethnicity, age, region, and area type provide sufficient information for the purpose of identifying communities of high risk of infection. Second, the publicly available geographic detail from NHANES III is relatively aggregated and there are likely to be intra-regional contextual differences, beyond those relating to metropolitan status. Third, the survey data used are from the mid-1990s and current Toxocara prevalence rates are likely to be different.

Appendix 1: Modeling missing data with particular regard to PIR

Missingness in the predictors is confined to the income to poverty ratio, conventionally described as the PIR. This variable is in fact a negative measure of poverty and a positive measure of socioeconomic status (SES). To account for missing PIR values, two extra variables are introduced. These are not used in the binary Toxocara risk model, but are additional measures of SES relevant to effectively modeling missing PIR values.

The extra variables are education (EDUC) of the household reference person, and Duncan's socioeconomic index value (SEI) for the occupation of the household reference person. These variables are also subject to missingness, though for education under 1 in 200 values are missing (0.29% for males, 0.44% for females).

A marginal/conditional regression sequence is used (Ibrahim et al. 1999) to model the joint distribution of EDUC, PIR and SEI conditional on the fully observed attributes (ethnicity, region, urbanity and age), denoted collectively as X_{obs}. In this analysis, PIR is continuous, though in the subsequent prevalence model it is converted to a binary variable. The marginal regression involves a multinomial logit regression p(EDUC|X_{obs}) of education (four categories) on ethnicity, region and urbanity. The first conditional regression p(SEI|EDUC, X_{obs}) models SEI conditional on education, and known attributes. The second conditional regression p(PIR|SEI, EDUC, X_{obs}) then models PIR conditional on both education and SEI, and on known attributes.

For subjects with missing PIR values, the posterior means of the PIR from this conditional regression are substituted for missing values (the means are from the second half of a two-chain run of 10,000 iterations in WINBUGS). The results from this conditional regression are interesting in showing that income to poverty ratios increase (as might be anticipated) with SEI and education; in addition, PIR is lower for Black non-Hispanic, Hispanic and other ethnic groups. PIR is also higher in urban counties and for people living in the West region.

To allow for the possibility that missingness in PIR values is informative, there is in addition to the marginal/conditional regressions, a logit regression of the binary missingness indicator for PIR ($R = 1$ for PIR missing, $R = 0$ otherwise) on all three SES indicators and on observed demographic and geographic attributes. This regression produced a significant positive coefficient on the PIR value, so that the chance of missing PIR response is greater at higher PIR values.

References

Acevedo-Garcia D (2000) Residential segregation and the epidemiology of infectious diseases. Soc Sci Med 51:1143–1161

Adams R, Howard N, Tucker G, Appleton S, Taylor A, Chittleborough C, Gill T, Ruffin R, Wilson D (2009) Effects of area

deprivation on health risks and outcomes: a multilevel, cross-sectional, Australian population study. Int J Public Health 54:183–192

Beck A (2000) The human–dog relationship: a tale of two species, Chap 1. In: Macpherson C, Meslin F, Wandeler A (eds) Dogs, zoonoses and public health. CABI Publishing, Wallingford, pp 1–16

Blizzard L, Hosmer D (2006) Parameter estimation and goodness-of-fit in log binomial regression. Biometrical J 48:5–22

Brooke J (1984) 6 Years of canine-waste law: all in all, a cleaner New York, Sect. I. New York Times 25 (16 June)

Centre for Disease Control (2007) Documentation, codebook, and frequencies; surplus sera laboratory component: antibody to Toxocara larva migrans. NHANES III, Series 11 Data Files 26A. Available via http://www.cdc.gov/nchs/about/major/nhanes/archive_whatsnew.htm. Accessed 30 March 2010

Cooper P (2008) *Toxocara canis* infection: an important and neglected environmental risk factor for asthma? Clin Exp Allergy 38:551–553

De Andrade Lima Coelho R, De Carvalho L, Perez E, Araki K, Takeuchi T (2005) Prevalence of toxocariasis in northeastern Brazil based on serology using recombinant *Toxocara canis* antigen. Am J Trop Med Hyg 72:103–107

Despommier D (2003) Toxocariasis: clinical aspects, epidemiology, medical ecology, and molecular aspects. Microbiol Rev 16:265–272

Deutz A, Fuchs K, Auer H, Kerbl U, Aspock H, Kofer J (2005) Toxocara infestations in Austria: a study on the risk of infection of farmers, slaughterhouse staff, hunters and veterinarians. Parasitol Res 97:390–394

Fan C, Lan H, Hung C, Chung W, Liao C, Du W, Su K (2004) Seroepidemiology of *Toxocara canis* infection among mountain aboriginal adults in Taiwan. Am J Trop Med Hyg 71:216–221

Genchi C, Di Sacco B, Gatti S, Sangalli G, Scaglia M (1990) Epidemiology of human toxocariasis in northern Italy. Parassitologia 32:313–319

Glickman L, Schantz P (1981) Epidemiology and pathogenesis of zoonotic toxocariasis. Epidemiol Rev 3:230–250

Graubard B, Korn E, Midthune D (1997) Testing goodness-of-fit for logistic regression with survey data. In: Proceedings of the section on survey research methods. American Statistical Association, Alexandria, pp 170–174

Habluetze A, Traldia G, Ruggieri S, Attilia A, Scuppa P, Marchetti R, Menghini G, Esposito F (2003) An estimation of *Toxocara canis* prevalence in dogs, environmental egg contamination and risk of human infection in the Marche region of Italy. Vet Parasitol 113:243–252

Havasiova K, Dubinsky P, Stefancíkova A (1993) A seroepidemiological study of human Toxocara infection in the Slovak Republic. J Helminthol 67:291–296

Herrmann N, Glickman L, Schantz P, Weston M, Domanski L (1985) Seroprevalence of zoonotic toxocariasis in the United States: 1971–1973. Am J Epidemiol 122:890–896

Hotez P (2007) Neglected diseases and poverty in ''the other America'': the greatest health disparity in the United States? PLoS Negl Trop Dis 1(3):e149. doi:10.1371/journal.pntd.0000149

Hotez P (2008) Neglected infections of poverty in the United States of America. PLoS Negl Trop Dis 2(6):e256. doi:10.1371/journal.pntd.0000256

Hotez P, Wilkins P (2009) Toxocariasis: America's most common neglected infection of poverty and a helminthiasis of global importance? PLoS Negl Trop Dis 3(3):e400. doi:10.1371/journal.pntd.0000400

Ibrahim J, Lipsitz S, Chen M (1999) Missing covariates in generalized linear models when the missing data mechanism is nonignorable. J Roy Stat Soc B 61:173–190

Kaplan M, Kamanli A, Kalkan A, Kuk S, Gülkesen A, Ardiçoğlu O, Demırdağ K (2005) Toxocariasis seroprevalence in patients with rheumatoid arthritis. Turkiye Parazitol Derg 29:251–254

Lunn D, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337

Lynch N, Wilkes L, Hodgen A, Turner K (1988) Specificity of Toxocara ELISA in tropical populations. Parasite Immunol 10:323–337

Macpherson C (2005) Human behaviour and the epidemiology of parasitic zoonoses. Int J Parasitol 35:1319–1331

Magnaval J, Galindo V, Glickman L, Clanet M (1997) Human Toxocara infection of the central nervous system and neurological disorders: a case–control study. Parasitology 115:537–543

Merlo J, Yang M, Chaix B, Lynch J, Råstam L (2005) A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. J Epidemiol Community Health 59:729–736

Mizgajskaa H (2001) Eggs of Toxocara in the environment and their public health implications. J Helminthol 75:147–151

Morenoff J, Lynch J (2004) What makes a place healthy? Neighborhood influences on racial/ethnic disparities in health over the life course. In: Anderson N, Bulatao R, Cohen B (eds) Critical perspectives on racial and ethnic differences in health in late life. National Academy Press, Washington, DC, pp 406–449

Nathwani D, Laing R, Currie P (1992) Covert toxocariasis as a cause of recurrent abdominal pain in childhood. Br J Clin Pract 46:271

Nelson S, Greene T, Ernhart C (1996) *Toxocara canis* infection in preschool age children: risk factors and the cognitive development of preschool children. Neurotoxicol Teratol 18:167–174

Nicoletti A, Sofia V, Mantella A, Vitale G, Contrafatto D, Sorbello V, Biondi R, Preux P, Garcia H, Zappia M, Bartoloni A (2008) Epilepsy and toxocariasis: a case–control study in Italy. Epilepsia 49:594–599

Noordin R, Smith H, Mohamad S, Maizels R, Fong M (2005) Comparison of IgG-ELISA and IgG4-ELISA for toxocara serodiagnosis. Acta Trop 93:57–62

Paquet-Durand I, Hernández J, Dolz G, Zuñiga J, Schnieder T, Epe C (2007) Prevalence of *Toxocara* spp., Toxascaris leonina and ancylostomidae in public parks and beaches in different climate zones of Costa Rica. Acta Trop 104:30–37

Plaut M, Zimmerman E, Goldstein R (1996) Health hazards to humans associated with domesticated pets. Annu Rev Pub Health 17:221–215

Roddie G, Stafford P, Holland C, Wolfe A (2008) Contamination of dog hair with eggs of *Toxocara canis*. Vet Parasitol 152:85–93

Stensvold C, Skov J, Møller L, Jensen P, Kapel C, Petersen E, Nielsen H (2009) Seroprevalence of human Toxocariasis in Denmark. Clin Vaccine Immunol 16:1372–1373

Tolan M, Laufer R (2009) Toxocariasis: an overview. eMedicine (Parasitology). Available via http://emedicine.medscape.com/article/999850-overview Accessed 30 March 2010

United States Department of Agriculture (USDA) (2004) Rural poverty at a glance. Rural Development Research Report Number 100, USDA Economic Research Service, Washington, DC

Won K, Kruszon-Moran D, Schantz P, Jones J (2008) National seroprevalence and risk factors for zoonotic Toxocara infection. Am J Trop Med Hyg 79:552–557