Published in final edited form as:

Sex Transm Infect. 2015 December ; 91(8): 610–614. doi:10.1136/sextrans-2014-051932.

Per-capita Incidence of Sexually Transmitted Infections Increases Systematically with Urban Population Size: a cross-sectional study

Oscar Patterson-Lomba^{*1}, Edward Goldstein², Andrés Gómez-Liévano³, Carlos Castillo-Chavez⁴, and Sherry Towers⁴

¹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA

²Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Centre for International Development, Harvard Kennedy School, Harvard University, Cambridge, MA, USA

⁴Simon A. Levin Mathematical, Computational, and Modelling Sciences Centre, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

Abstract

Objectives—Rampant urbanization rates across the globe demand that we improve our understanding of how infectious diseases spread in modern urban landscapes, where larger and more connected host populations enhance the thriving capacity of certain pathogens.

Methods—A data-driven approach is employed to study the ability of sexually transmitted diseases (STDs) to thrive in urban areas. The conduciveness of population size of urban areas and their socio-economic characteristics are used as predictors of disease incidence, using confirmed-case data on STDs in the United States as a case study.

Results—A *superlinear* relation between STD incidence and urban population size is found, even after controlling for various socio-economic aspects, suggesting that doubling the population size of a city results in an expected increase in STD incidence larger than twofold, provided that all other socio-economic aspects remain fixed. Additionally, the percentage of African Americans, income inequalities, education, and per capita income are found to have a significant impact on the incidence of each of the three STDs studied.

^{*}opatters@hsph.harvard.edu.

Competing Interest: None declared.

Contributorship statement: OPL, EG, AG and ST conceived and designed the study. OPL carried the statistical analyses and wrote a first draft. All authors contributed to the writing and revision of the final manuscript.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ and co-owners or contracting owning societies (where published by the BMJ on their behalf), and its Licensees to permit this article (if accepted) to be published in Sexually Transmitted Infections and any other BMJ products and to exploit all subsidiary rights, as set out in our licence.

Conclusions—STDs disproportionally concentrate in larger cities. Hence, larger urban areas merit extra prevention and treatment efforts, especially in developing countries where urbanization rates are higher.

Keywords

Sexually Transmitted Diseases; Incidence; Urban Population; Urbanization; Superscaling; Socio-economic Drivers

1 INTRODUCTION

Rapid urbanization is taking place across the world, especially in developing countries. By 2050, over 60% of the world's population is expected to live in urban areas [1]. Rampant urban population growth and densification, in combination with socio-economic disparities and precarious public health systems, may create favourable conditions for the wide spread of certain infectious diseases (IDs) [2–7].

Population growth, international travel and migration are common features of urbanization that can turn cities into gateways and enhancers of IDs spread [3, 8]. Urbanization may have played a critical role in the worldwide dissemination of the human immunodeficiency virus (HIV) [5, 9], the worsening of epidemics of major respiratory viruses (e.g., Influenza, RSV) [10] and tuberculosis [2, 3, 11]. Also critical is that larger populations and densities may be associated with environments that promote increased human interactions, and changes in social norms and mixing patterns [12], possibly affecting ID transmission.

Even though urban dwellers enjoy, on average, better health, education and income, socioeconomic disparities being often exacerbated in urban environments [8, 11, 13]. These disparities can be determinants for an individual's overall health [3, 4, 6], and could also play a key role in infection patterns.

Some aspects limiting further understanding the dynamics of IDs in urban areas are i) lack of data within and across cities; ii) when these data are available, most studies have focused on country-, state- or county-level associations; and iii) mechanistic models are often context (city/region) specific and complex, reducing our ability to factor out generalizable insights.

We use data on confirmed cases of STDs as a case study to begin addressing some of these deficiencies. The data include annual cases of disease incidence across U.S. cities from 2007 to 2011 for syphilis, gonorrhoea and chlamydia, three IDs of public health interest given their long-term complications, particularly if untreated. Using metropolitan statistical areas (MSAs) as unit of analysis, we go beyond the rural-versus-urban comparisons, or the study of a particular disease in one city at a time [3, 14, 15]. Instead we take a cross-sectional look at all MSAs for three STDs, an approach that allows the identification of systemic patterns that can generate insights on the interplay between urban population size and the spread of STDs.

On average, the rates of chlamydia, gonorrhoea and syphilis are higher in urban areas, as compared to the national rates in the U.S. (see Figure 1). Previous studies have found that

the percentage of African Americans in a population is an important factor to consider [17–19], with STD rates in African Americans largely exceeding those in whites [14]. These STD burden disparities can be traced back to a greater proportion of African-Americans living in geographic clusters characterized by low educational attainment, high unemployment rates and lower socioeconomic status [18, 20]. Hence, accessibility to healthcare services, education, as well as income and social inequalities are key factors in the spread of STDs.

Noteworthy, the role of population size– and more specifically, the population size of Metropolitan Statistical Areas (MSAs)– where individuals reside has received much less attention. In this work we study the associations between STD incidence and population size of MSAs, while also controlling for the confounders mentioned above.

So far, most of the studies that have analysed these issues often considered states, counties or single cities/regions, as the unit of analysis (e.g., [17, 19]). Naturally, our approach will build upon the existing knowledge by focusing instead on MSAs; to our knowledge, the first study to do so.

1.2 Why Metropolitan Statistical Areas?

MSAs' definitions are not political or administrative, but based on levels of economic integration and commuting [21]. MSAs are commonly used in urban economics, regional science and economic geography, and so employing MSAs as the unit of analysis may be more appropriate to study human infection processes. An MSA is defined as a core area (i.e., a city) containing a large population nucleus (> 50, 000 people), along with adjacent suburban counties that share a high degree of socio-economic integration [22]. By definition, a significant fraction of people (>25%) within the boundaries of an MSA commute on a regular basis to and from the core population, potentially having relevant implications for diseases dynamics. Thus, MSAs can be seen as highly interconnected metapopulations containing an urban hub where IDs can spread relatively faster. For these reasons, MSAs are a more suitable unit of analysis for studying disease infection patterns in urban landscapes than states or counties.

2 MATERIALS AND METHODS

2.1 Data Sources

Our response variables are the annual counts of chlamydia, gonorrhoea and syphilis (primary, secondary and congenital) from 2007 to 2011, in the counties of the 48 contiguous states, as reported by the Centers for Disease Control and Prevention (CDC) [16]. This dataset does not include any individual level data.

The explanatory covariates are:

- Population of counties (2010 estimates from Census Bureau).
- Percentage of African-American residents (Census Bureau).
- Income inequality (Gini index), describing the income distribution in a population (*Measure of America* [23]).

- Education index, based on two sub-indices: Educational Attainment Index and Enrolment Index (details in [23]).
- Percent insured under 65 years of age (Census Bureau).
- Income per-capita in units of 1000 U.S. dollars (Census Bureau).

Since the STD and the covariate data were at the county level, we constructed MSA-level metrics using county-level data (see Supplementary Information (SI) for details and for summary statistics of these variables). Of the 375 MSAs within the 48 contiguous states, our dataset had information on 364. Statistical analyses were performed in **R**.

2.2 The Regression Model

We assess how well the variation of urban STD incidence can be explained by the population size of urban areas and their socio-economic characteristics. Ignoring the impact of the socioeconomic covariates for the moment, we approach this issue using a framework from urban scaling theory [24], which supposes that Y , an urban metric, and N , the population size of an urban area, satisfy the scaling relationship

$$E[Y|N] = Y_0 N^{\alpha_N}. \quad (1)$$

Here Y is considered a random variable with Y_0 as the baseline value. The scaling exponent, α_N , measures the average relative change in Y with respect to N . Assuming that for each city indexed $i \in \{1, 2, \dots, 364\}$,

$$Y_i = Y_0 N_i^{\alpha_N} 10^{\varepsilon_i}, \quad (2)$$

where ε_i is a random noise, we linearize Eq.(2) by taking base 10 logarithms, yielding

$$\log(Y_i) = \log(Y_0) + \alpha_N \log(N_i) + \varepsilon_i. \quad (3)$$

Empirical regularities can then be investigated using regression techniques, allowing the assessment of population size as a predictor of Y , and also the estimation of the scaling exponent α_N . It has been found that population size is a strong predictor of several urban metrics. Quantities reflecting interaction processes such as new HIV cases, innovation and violent crime have a scaling exponent $\alpha_N > 1$ (*superlinear* scaling), whereas those accounting for a city's physical infrastructure display $\alpha_N < 1$ [24–26].

In the context of IDs, an informative interpretation of α_N is: Let Y_i and N_i be, respectively, the number of infected cases and the population of MSA i . Dividing both sides of Eq.(1) by N_i yields $Y_i/N_i = Y_0 N_i^{\alpha_N - 1} = f_i$, the fraction of the population that is infected, or the per-capita incidence. When $\alpha_N > 1$, $f(N) = Y_0 N^{\alpha_N - 1}$ is an increasing function of N . Therefore, when the scaling relation is *superlinear*, the expected per-capita incidence will increase with population size. In fact, incidence increases systematically on a per capita basis by $(\alpha_N - 1) \times 100\%$ with every doubling of the population.

We are also interested in discerning which socio-economic factors are associated with the patterns of STD incidence in cities. To that end, we extend model (3) by adding the covariates of interest, yielding

$$\log(Y_i) = \log(Y_0) + \alpha_N \log(N_i) + \alpha_A A_i + \alpha_G G_i + \alpha_E E_i + \alpha_S S_i + \alpha_I I_i + \varepsilon_i, \quad (4)$$

where A, G, E, S and I represent, respectively, % African American, Gini, Education Index, % Insured and per-capita Income.

We define ‘incidence’ as the sum of reported cases between 2007 and 2011, a five-year cumulative incidence. A parsimonious way to model incidence data is to assume that the counts are Poisson distributed. However, since these data are over-dispersed compared to the Poisson model, we employ Negative Binomial regression (see SI for details).

To help account for the issue of multicollinearity among covariates we use a variable-selection algorithm (*stepAIC*) to determine the most parsimonious model that better explains the incidence variability, excluding the less significant covariates via the Akaike Information Criterion (AIC). Employing LASSO regression yielded similar results (see SI).

Using this scaling framework, we address a series of questions. First, is the relationship between city size and incidence *superlinear*? Second, are these scaling relationships different for each disease? Third, does this relationship persist after controlling for important socio-economic confounders? And fourth, which socio-economic covariates are significantly associated with STD incidence?

3 RESULTS

We first estimate the scaling exponents α_N using the simpler model in (3). The estimates are $\alpha_N = 1.04$ (0.02), 1.10 (0.04) and 1.29 (0.05), respectively for chlamydia, gonorrhoea and syphilis (standard errors in parenthesis). To assess whether the scaling parameter estimates α_N are significantly larger than one for each disease, we test the null $H_0: \alpha_N = 1$. We can reject the null with a significance level of 0.01 to guard against type 1 error (except for the case of chlamydia with $p = 0.02$), evidencing that all three diseases feature *superlinear* scaling and that larger urban areas seem to be conducive of increased per-capita STD incidence.

We can also conclude that the scaling patterns for these three STDs in MSAs are different from each other ($p \ll 0.01$) (see SI for details).

To evaluate whether these superscaling patterns persist even after controlling with important socio-economic factors, we employ the full model (4). The results are given in Table 1. Figure 2 depicts eloquently the *superlinear* pattern between STD incidence and population size when using model (4). It can be shown that, indeed, after accounting for important confounders, the scaling exponents are still larger than one (with significance level of 0.02).

Inferences from socio-economic covariates

In all three cases, estimates of α_A are significantly larger than zero, indicating that, everything else equal, an increase in the % of African Americans in cities is associated with an increase in the incidence of STDs. In fact, the largest increases are for gonorrhoea and syphilis.

Estimates of α_G remain significant and positive in all three STDs, suggesting that larger income inequality is associated with higher incidence. Education, not surprisingly, is consistently associated with a reduction of the expected incidence. Percent Insured is only selected for the case of gonorrhoea, with a positive association with incidence. Lastly, per-capita Income is also significant for all three STDs, with higher income being associated with a reduction in expected incidence, except for the case of syphilis (see SI for a more detailed interpretation of these associations).

4 DISCUSSION

Characterizing the relationship between the spread of STDs and urban socio-economic aspects may offer suitable venues to determine effective ways of deploying control efforts in urban populations. This issue gathers particular relevance given the upward trends in urbanization worldwide, especially in developing countries where large socio-economic disparities can create favourable conditions for the large-scale spread of certain IDs.

Here we show that, even after controlling for important socio-economic confounders, annual incidence of chlamydia, gonorrhoea and syphilis across U.S. metropolitan areas follow a *superlinear* scaling pattern, suggesting that the per-capita incidence of STDs increases, on average, with the population size of MSAs. In other words, when living in urban areas with similar socio-economic indicators, a person is at higher risk of contracting an STD if he/she resides in a larger area. This empirical regularity strongly suggests that larger MSAs merit extra prevention and treatment efforts.

In addition, the scaling patterns for each STD are significantly different. To begin to understand why, note that chlamydia is the least symptomatic of the three [16], making it the most difficult to detect and prevent, and is therefore the most prevalent of the three.

Gonorrhoea is more symptomatic and less prevalent, and syphilis is the most symptomatic and the least prevalent. Moreover, syphilis has the lowest infection risk per sexual act, with chlamydia and gonorrhoea having comparable infectiousness (ref. *Australasian contact tracing manual*). Certainly, the symptomatology and inherent infectivity of a disease affect its overall transmissibility, with more asymptomatic and infectious STDs (e.g., chlamydia), affecting a larger proportion of the population, compared to less transmissible ones, such as syphilis, which presumably propagate largely through high-risk (core) sexual networks [27] where individuals engage in riskier sexual practices.

At the same time, note that the associations of incidence with population size (quantified by α_N) decrease as STDs increase their overall transmissibility and prevalence. Thus, this scaling analysis suggests that the easier an STD can spread, the lower the association of population size with its transmission potential. According to this view, population size has

an enabling effect on STD transmission, but more so for STDs that transmit with less ease. Hence, the spreading capacity of syphilis is considerably tied in to the population size, which could be, in turn, due to the disproportionate presence of larger high-risk sexual networks in larger cities. Arguably, the structure and size of high-risk sexual networks are affected by the size of the cities in which they exist.

In connection with this, chlamydia is the most evenly distributed geographically in the U.S., with gonorrhoea and syphilis being more geographically clustered (in that order) [28]. Our estimates of the scaling exponent for each STD also increase in that order, indicating that spatial clustering and *superlinearity* are positively correlated [26]. Moreover, the level of clustering increases as the level of aggregation decreases (from states, to MSAs to counties; see [28] and SI). Extrapolating this trend we expect that at the community level spatial heterogeneities are even higher, implying that few communities feature very large rates of STDs compared to the rest of the communities. These communities with extreme levels of infection likely harbour high-risk sexual networks [27].

Our study, as others e.g., [18, 19], also suggests that the percentage of African-Americans in the population is one of the most significant covariates explaining the variability of STD incidence in cities. Moreover, given the risk-prone socio-economic environments in which African-Americans live and develop, they likely are important contributors of the maintenance and growth of the high-risk sexual networks in which the less prevalent STDs persist.

The indication that larger levels of income inequality—which are often mirrored in education and healthcare access disparities—are associated with an increase in the expected number of STD cases is in line with previous findings [17], and supports the view that larger income inequality correlates with worse health outcomes [29].

Paradoxically, our analysis indicates that more medical coverage is associated with larger incidence of gonorrhoea. Univariate regression of incidence with % Insured yielded a negative association. However, after controlling for population size and % African American, the association reverses its direction. Thus this finding should be interpreted cautiously (see SI for details).

The association of income and incidence differs between STDs, with incidence increasing with income only in the case of syphilis. This association may be confounded by the contribution of men that have sex with men (MSM) in the transmission of syphilis (CDC recently reported that 75% of cases of syphilis are currently in MSM), particularly given that larger and more affluent cities typically harbour larger populations of MSM [30]. Hence, a key covariate for future analyses is the percent of residents who are MSM (see SI for further discussion).

A limitation in further improving the interpretation of these associations is that the data did not contain any information regarding the gender, age, race or sexual orientation of those infected. As a result we only inferred associations at the MSA level. To avoid committing “ecological fallacies”, associations inferred at a certain level of aggregation (e.g., MSA level) cannot be extrapolated to other levels of aggregation (e.g., individual level) without

additional evidence. Moreover, *causality* will be hard to infer without additional information on those infected. Noteworthy, our study does not aim to infer a *direct* causal link between population size and the chances of contracting an STD. Our claim is, instead, that population size *alone* can be a good aggregate proxy for other myriad factors that *directly* affect the spreading potential of an STD.

The systemic associations found herein might not apply to a particular city. There may be large cities with relatively low levels of STD incidence. What we find is that these types of cities are not the norm. In fact, it is often cities that deviate from their expected behaviour that are the most interesting for both policy and scientific analyses [25]. Another limiting aspect is that the number of reported STD cases could have been subject to biases or inaccurate reporting. Any systematic under- or over-reporting may have influenced the inferred patterns. For example, the lack of information on healthcare services, such as diagnostic/testing rates, precludes us from ascertaining how much of the *superscaling* effect is due to a better testing/diagnostic infrastructure in larger and more developed cities.

Furthermore, although our regression model choice is reasonable and interpretable, we recognize that other regression models could have been used. As in most linear regression models, ours makes no assumption about the distribution of the predictors. Hence, the large variation in the distribution of MSA populations [21] does not affect the estimations.

In summary, this study quantifies the associations between STD incidence and urban population size, while controlling for important socio-economic confounders. To our knowledge it is the first to take this approach. The superlinear scaling found, in essence, denotes a continuous and systematic enabling of STD transmission with increasing population size so that larger cities suffer from disproportionately larger burdens of STD incidence. Moreover, population size appears to have a greater enabling effect for STDs that transmit with less ease. These insights indicate that incidence hotspots are likely located in larger cities, and thus provide a criterion for effectively allocating interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the reviewers for insightful comments that helped improve the manuscript. We also thank Sarah Kidd and Lyn Finelli at the Centers for Disease Control and Prevention, for facilitating access to the STD data used herein. Finally, we thank Jay Taylor and Julia Wu for useful discussions.

This work was supported by National Institutes of Health (NIH) training grant number T32AI007358-26 (OPL); by the National Institute of General Medical Sciences (NIGMS) award number U54GM088558 and by the NIH K01 award 1K01AI101010-01 (EG); and by NIGMS grant number 1R01GM100471-01 (CCC and ST).

References

1. United nations, world urbanization prospects: The 2011 revision.
2. Aparicio JP, Capurro AF, Castillo-Chavez C. Markers of disease evolution: the case of tuberculosis. *J Theor Biol.* 2002; 215:227–237. [PubMed: 12051976]

3. Alirol E, Getaz L, Stoll B, et al. Urbanisation and infectious diseases in a globalised world. *Lancet Infect Dis.* 2011; 11:131–141. [PubMed: 21272793]
4. Galea S. Urbanization, urbanicity, and health. *J Urban Health.* 2002; 79:S1–S12. [PubMed: 12473694]
5. Cohen C, Horlacher D, MacKellar FL. Is urbanization good for a nation's health?. 2010
6. Leon DA. Cities, urbanization and health. *Int J Epidemiol.* 2008; 37:4–8. [PubMed: 18187525]
7. Bamaïyi PH. The role of demographics and human activities in the spread of diseases. *Current Trends in Technology and Sciences.* 2013; 2:253–257.
8. Morens DM, Fauci AS. Emerging infectious diseases: Threats to human health and global stability. *PLoS Pathog.* 2013; 9:e1003467. [PubMed: 23853589]
9. Quinn TC. Population migration and the spread of types 1 and 2 human immunodeficiency viruses. *Proc Natl Acad Sci.* 1994; 91:2407–2414. [PubMed: 8146131]
10. Glezen WP. The changing epidemiology of respiratory syncytial virus and influenza: impetus for new control measures. *Pediatr Infect Dis J.* 2004; 23:S202–S206.
11. Stephens C. Healthy cities or unhealthy islands? The health and social implications of urban inequality. *Environ Urban.* 1996; 8:9–30.
12. Schlapfer M, Bettencourt L, Grauwin S, et al. The scaling of human interactions with city size. *J R Soc Interface.* 2014;11.
13. Dye C. Health and urban living. *Science.* 2008; 319:766–769. [PubMed: 18258905]
14. Mao L, Bian L. Spatial–temporal transmission of influenza and its health risks in an urbanized area. *Comput Environ Urban Syst.* 2010; 34:204–215.
15. Gesink DC, Sullivan AB, Miller WC, et al. Sexually transmitted disease core theory: roles of person, place, and time. *Am J Epidemiol.* 2011; 174:81–89. [PubMed: 21540320]
16. Center for Disease Control and Prevention. Sexually transmitted disease surveillance 2012. Atlanta: US Department of Health and Human Services; 2013.
17. Holtgrave DR, Crosby RA. Social capital, poverty, and income inequality as predictors of gonorrhoea, syphilis, chlamydia and aids case rates in the United States. *Sex Transm Infect.* 2003; 79:62–64. [PubMed: 12576618]
18. Farley TA. Sexually transmitted diseases in the Southeastern United States: location, race, and social context. *Sex Transm Dis.* 2006; 33:S58–S64. [PubMed: 16432486]
19. Semaan S, Sternberg M, Zaidi A, et al. Social capital and rates of gonorrhea and syphilis in the united states: spatial regression analyses of state-level associations. *Soc Sci Med.* 2007; 64:2324–2341. [PubMed: 17400352]
20. Marmot M. Social determinants of health inequalities. *Lancet.* 2005; 365:1099–1104. [PubMed: 15781105]
21. Ioannides Y, Skouras S. US city size distribution: Robustly Pareto, but only in the tail. *J Urban Econ.* 2013; 73:18–29.
22. Rozenfeld HD, Rybski D, Andrade JS, et al. Laws of population growth. *Proc Natl Acad Sci.* 2008; 105:18702–18707. [PubMed: 19033186]
23. Burd-Sharps, S.; Lewis, K. The measure of America: American human development report, 2008–2009. Columbia University Press; 2013.
24. Bettencourt LM, Lobo J, Helbing D, et al. Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci.* 2007; 104:7301–7306. [PubMed: 17438298]
25. Bettencourt LM, Lobo J, Strumsky D, et al. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PloS One.* 2010; 5:e13541. [PubMed: 21085659]
26. Gomez-Lievano A, Youn H, Bettencourt LM. The statistics of urban scaling and their connection to Zipf's law. *PloS One.* 2012; 7:e40393. [PubMed: 22815745]
27. Hethcote, HW.; Yorke, JA. Gonorrhea transmission dynamics and control. Vol. 56. Springer; Berlin: 1984.
28. Chesson HW, Sternberg M, Leichter JS, et al. The distribution of chlamydia, gon- orrhoea and syphilis cases across states and counties in the USA, 2007. *Sex Transm Infect.* 2010; 86:iii52–iii57. [PubMed: 21098057]

29. Pickett, K.; Wilkinson, R. The spirit level: Why more equal societies almost always do better. London: Allen Lane; 2009.
30. Lieb S, Fallon SJ, Friedman SR, et al. State-wide estimation of racial/ethnic populations of men who have sex with men in the US. Public Health Reports. 2011; 126:60. [PubMed: 21351396]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Key Messages

- The focus of this study is on the incidence of sexually transmitted diseases (STDs) in metropolitan statistical areas.
- The analyses suggest that the per-capita incidence of STDs increases systematically with urban population size, more so for less prevalent STDs.
- Larger urban areas merit extra prevention efforts, a particularly important recommendation given the rise in urban populations worldwide, especially in developing countries.

FIGURE 1.

Logarithm of infection rates (cases per 10^5) for chlamydia, gonorrhoea and syphilis from 2007 to 2011. Comparison between the 50 most populous MSAs (solid lines) versus all U.S. population (dashed lines). Rates are consistently higher in MSAs, particularly for syphilis, the least common of the three STDs under study [16].

FIGURE 2.

Scaling of STD incidence with MSA population with Negative Binomial regression for chlamydia (left), gonorrhoea (centre) and syphilis (right) using model (4), as reported in Table 1. Comparing the blue lines (with slopes $\widehat{\alpha_N}$) with the red thin lines (with slope 1) shows the departures from the linear pattern in each case.

TABLE 1

Regression analysis results using model in (4).

<i>Covariates</i>	<i>Coefficients</i>	Chlamydia	Gonorrhoea	Syphilis
<i>Population</i>	$\widehat{\alpha}_N$	1.039***	1.119***	1.292***
	CI_N	[1.008, 1.070]	[1.054, 1.184]	[1.212, 1.372]
<i>African Am.</i>	$\widehat{\alpha}_A$	0.01***	0.024***	0.025***
	CI_A	[0.009, 0.011]	[0.022, 0.026]	[0.022, 0.028]
<i>Gini</i>	$\widehat{\alpha}_G$	0.645**	2.312***	5.993***
	CI_G	[0.069, 1.221]	[0.848, 3.776]	[4.505, 7.481]
<i>Education</i>	$\widehat{\alpha}_E$	-0.015	-0.051**	-0.106***
	CI_E	[-0.035, 0.005]	[-0.096, -0.006]	[-0.155, -0.057]
<i>Insurance</i>	$\widehat{\alpha}_S$	—]	0.011***	—
	CI_S		[0.003, 0.019]	
<i>Income</i>	$\widehat{\alpha}_I$	-0.005**	-0.011***	0.014**
	CI_I	[-0.009, -0.001]	[-0.019, -0.003]	[0.004, 0.024]
<i>Intercept</i>	$\log(\widehat{Y}_0)$	-2.125***	-4.705***	-8.371***
	CI	[-2.395, -1.855]	[-5.763, -3.647]	[-9.081, -7.661]
<i>Dispersion</i>	r	12.805	3.073***	2.181

Results include coefficients estimates, their 95% confidence intervals (CI), the p-values (** means $p < 0.05$ and *** means $p < 0.01$, with p-values testing the Null $\widehat{\alpha}_X \leq 0$) and the over-dispersion parameter r . A stepwise model selection was carried out using stepAIC algorithm, with “--” indicating that the algorithm excluded the variable from the regression.