Supplemental Material

Conserved patterns of incomplete reporting in pre-vaccine era childhood diseases

C. E. Gunning, E. Erhardt, and H. J. Wearing. 2014.

Results Overview Tables

Disease	Area	Era	N	Start	End
Measles	U.S. Cities	Early	574	1924-01-05	1934-12-29
Measles	E&W Cities	Early	286	1944-01-09	1954-12-26
Measles	U.S. Cities	Late	574	1935-01-05	1945-12-29
Measles	E&W Cities	Late	312	1955-01-09	1966-12-25
WC	U.S. Cities	Early	574	1924-01-05	1934-12-29
WC	U.S. Cities	Late	574	1935-01-05	1945-12-29

Table S1. Time range and sample number of subdivided city case reports. WC: whooping cough; E&W: England & Wales; N: number of sampled case reports.

Model	Subset	Slope	CI	Correlation	CI
City vs. State	Measles	0.93	(0.67, 1.23)	0.66	(0.50, 0.77)
City vs. State	Whooping cough	0.82	(0.53, 1.19)	0.55	(0.34, 0.73)
Measles vs. Whooping cough	City	1.00	(0.81, 1.15)	0.90	(0.78, 0.96)
Measles vs. Whooping cough	State	0.95	(0.82, 1.07)	0.89	(0.82, 0.94)

Table S2. Linear models of reporting probability variation (microdata method) between area and between disease, constructed from 1E+04 bootstrap draws (see Methods). Median and 95% CI of model results and correlations are shown. Slope estimates are in logit_2 units. For the measles subset, doubling the state reporting probability odds, i.e. from 50% (1/1 odds, $\text{logit}_2(\text{odds})=0$) to 66% (2/1 odds, $\text{logit}_2(\text{odds})=1$) approximately doubles the city reporting probability odds.

Area	Response	DF	R^2	Parameter	Estimate	Std. Error	t value	Pr(> |t|)
State	Measles	41	0.515	(Intercept)	-2.25	0.105	-21.5	6.01e-24
				prop.white	5.84	1.06	5.5	2.23e-06
				prop.labforce	16.2	4.32	3.76	0.000538
				sd.housesize	2.33	0.849	2.75	0.0089
State	Whooping cough	41	0.404	(Intercept)	-4.32	0.12	-35.9	1.39e-32
				prop.white	6.59	1.33	4.96	1.28e-05
				sd.housesize	4.28	0.987	4.33	9.23e-05
				prop.labforce	14	5.01	2.79	0.00791
				prop.male	-18.9	9.75	-1.93	0.06
City	Measles	45	0.317	(Intercept)	-1.25	0.17	-7.34	3.18e-09
				prop.school	36.8	8.04	4.58	3.68e-05
				mean.housesize	-1.66	0.646	-2.57	0.0135
City	Whooping cough	46	0.128	(Intercept)	-3.31	0.203	-16.3	9.31e-21
				prop.school	26.4	9.38	2.81	0.00721

Table S3. Linear models of reporting probability (microdata method) in response to demographic covariates. Models were constructed via forward selection using the BIC. For each model, parameters are shown in decreasing order of BIC reduction. Separate models are denoted by horizontal lines. Regardless of disease, proportion white is the best predictor of state reporting probabilities, while proportion attending school is the best predictor of city reporting probabilities. Reporting probability was logit_2 transformed and all predictors were zero-centered prior to model construction.

Demographic Data Figures

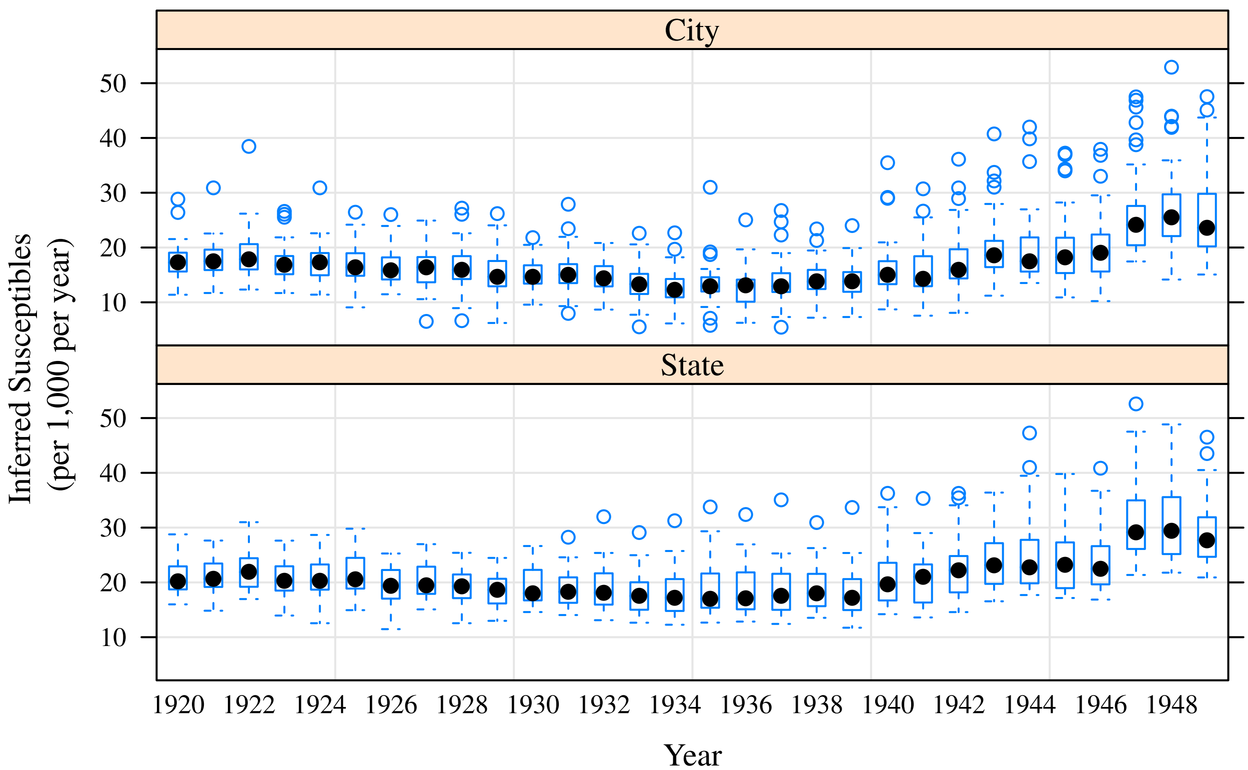

Figure S1. Distribution of yearly new susceptibles. The per capita rate of births + migration was inferred from IPUMS census records and the 1930 population size. Lower depression-era birth rates in the early 1930s and high post-war birth rates in the mid to late 1940s are apparent. Median yearly state birth rates are generally higher than associated city medians, likely due to the higher birth rates of rural populations.

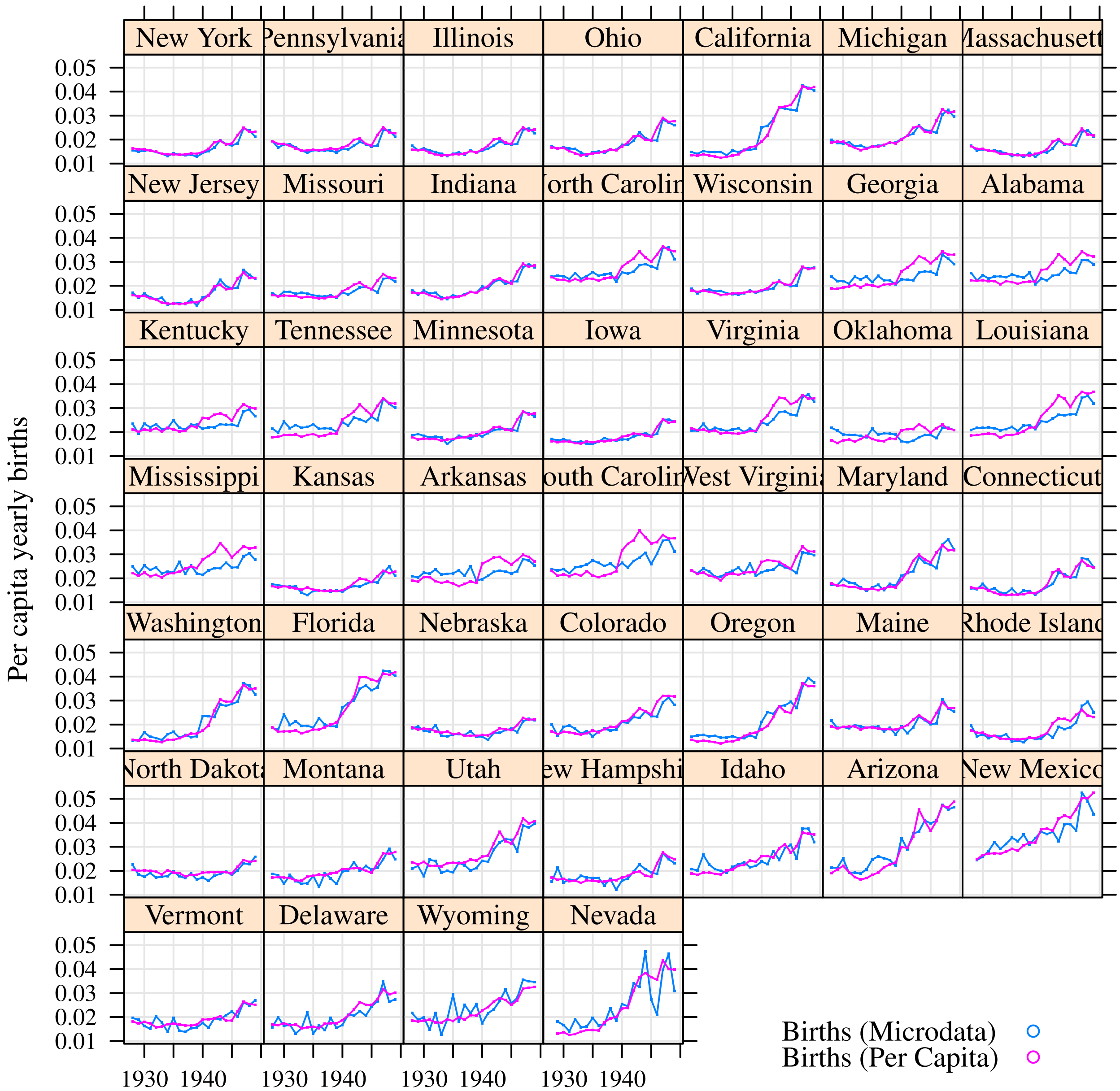

Figure S2. Comparison of per capita and microdata method by year. States are ordered by population size (1930, microdata method). The largest states show a close concordance between methods. The microdata method shows significant stochastic variation in smaller states.

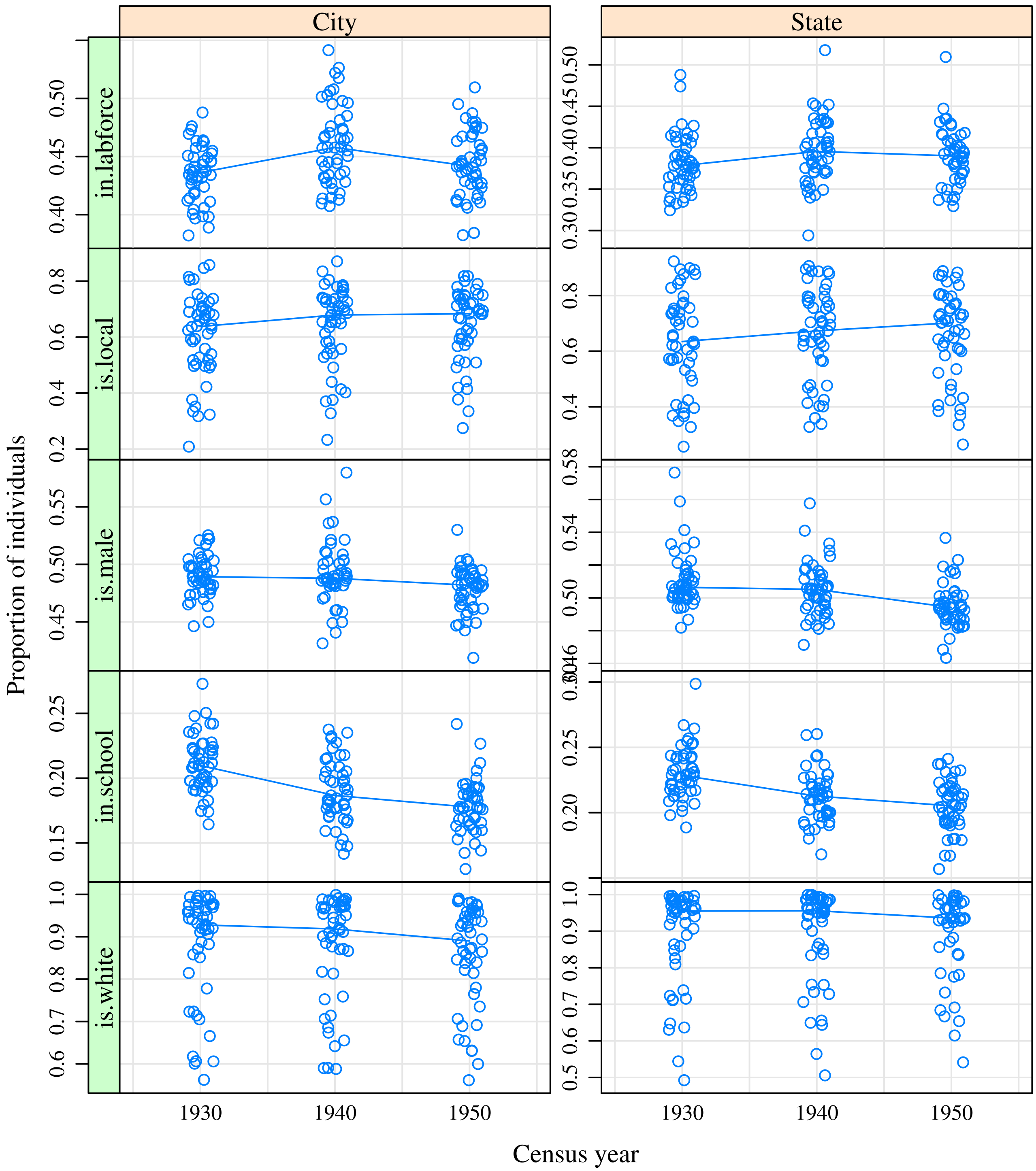

Figure S3. Distribution of categorical demographic covariates for each census (solid line follows decadal median). For some covariates, the sample *universe* (i.e., to whom the census question was asked) and precise question text is not comparable between censuses (for details see `https://usa.ipums.org/usa-action/variables/search`). Thus, the marked decrease in schooling in 1950 is possibly caused by the change in universe from all persons (pre-1950) to persons under age 30. Proportion white (prop.white) and proportion male (prop.male) are fully comparable between decades.

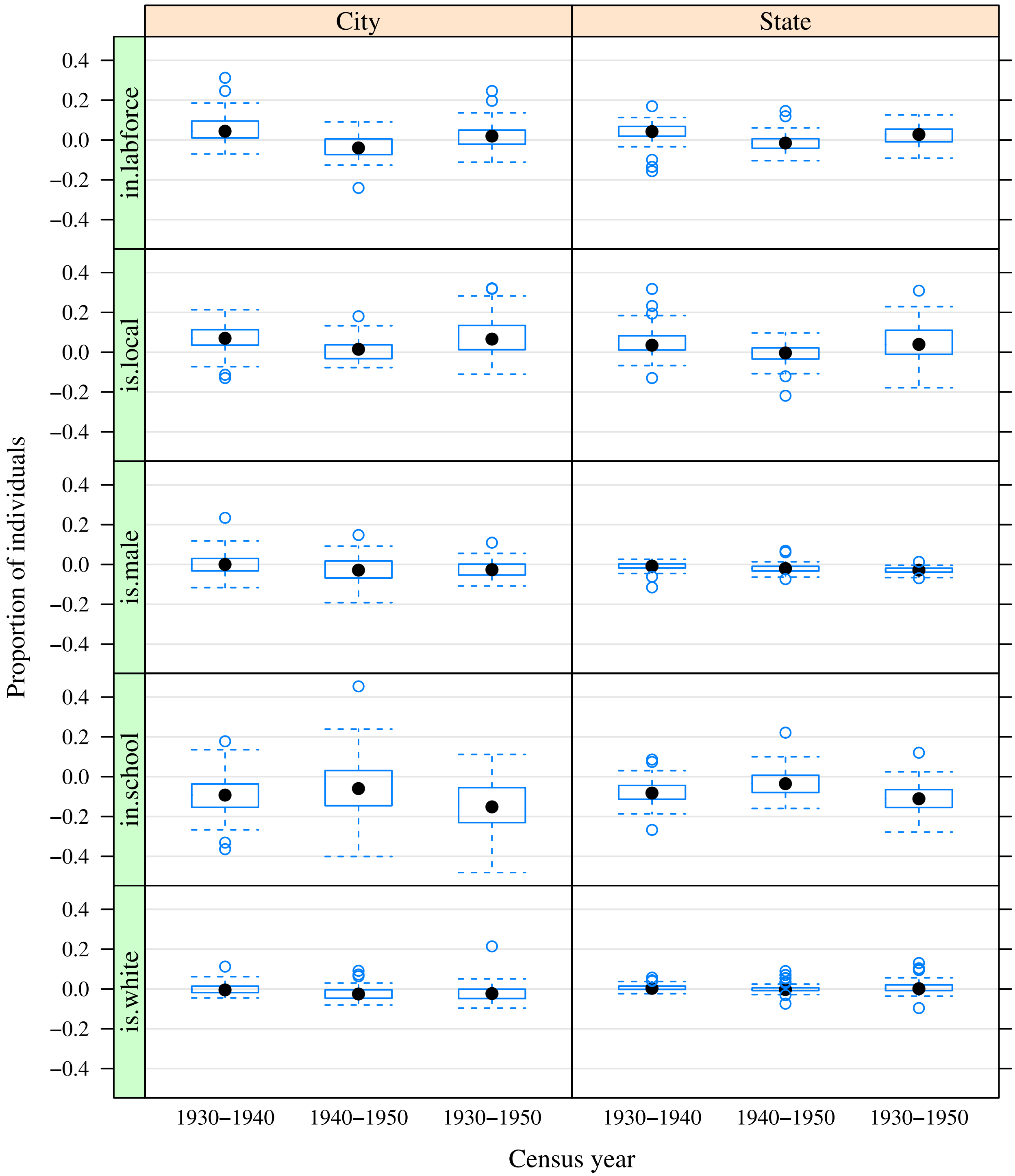

Figure S4. Distribution of inter-decadal change in categorical demographic covariates. Data as in Figure S3, showing each variable's change in proportion from one decade to the next relative to its initial proportion.

Reporting Probability Figures

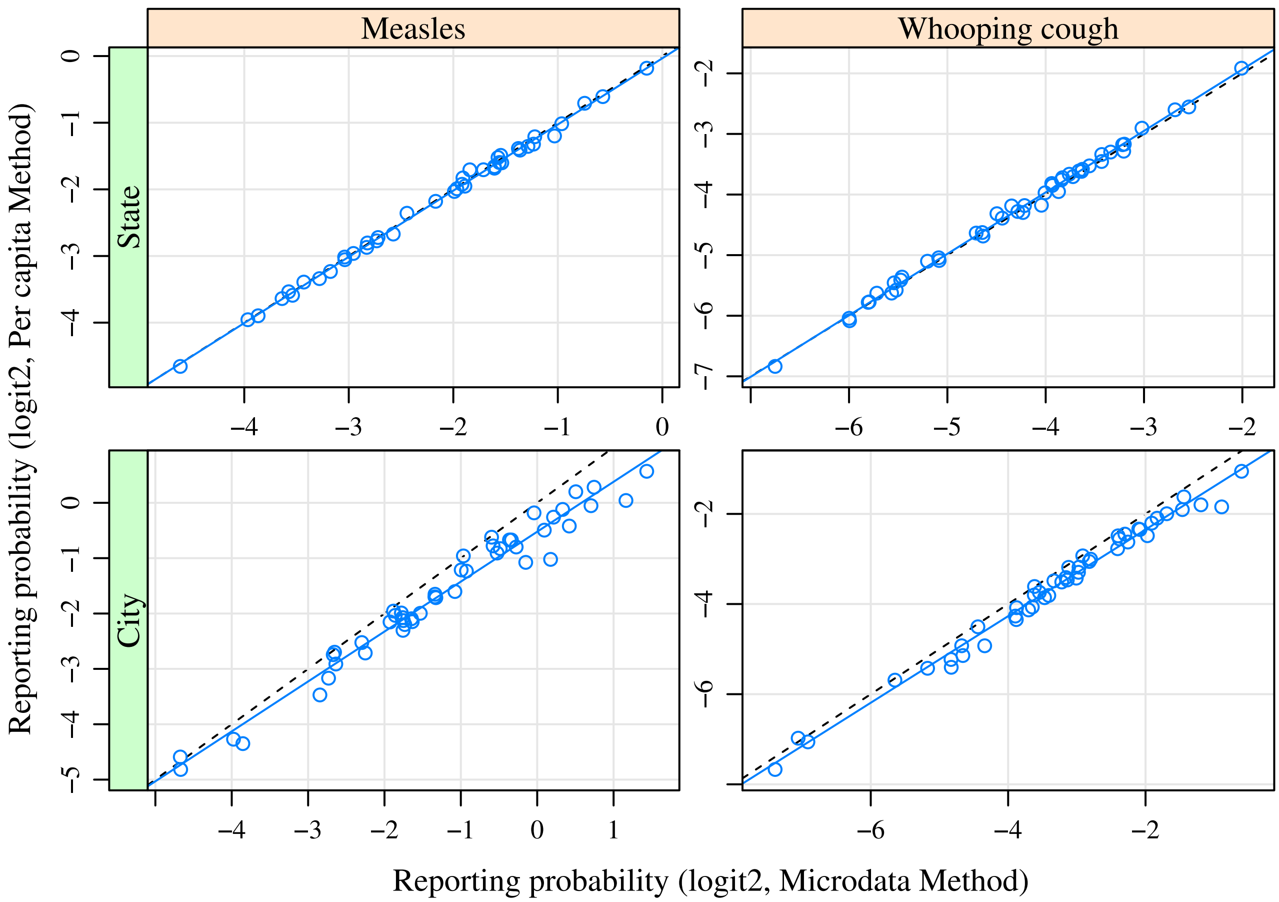

Figure S5. Comparison of estimated reporting probabilities between the per capita and the microdata method. A close concordance between methods is evident for states. For cities, the per capita method estimates more susceptibles and lower reporting probabilities than the microdata method. This is likely caused by the per capita method's use of state birth rates as proxies for city birth rates. State birth rates include rural areas, which generally have higher birth rates than urban areas (Figure S1). All values were logit_2 transformed to correct for heteroskedasticity. Black dashed line: 1-1 line; blue line: linear model.

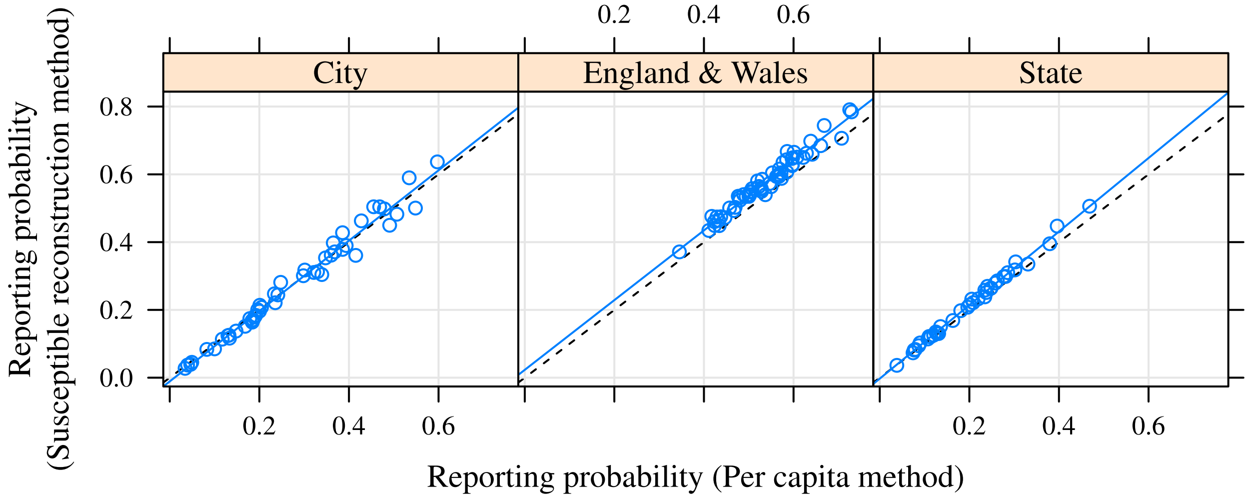

Figure S6. Comparison of reporting probability estimation methods. A close correspondence between the per capita method and the susceptible reconstruction method is evident (dashed black line: 1-1 line; blue line: linear model between the two methods).

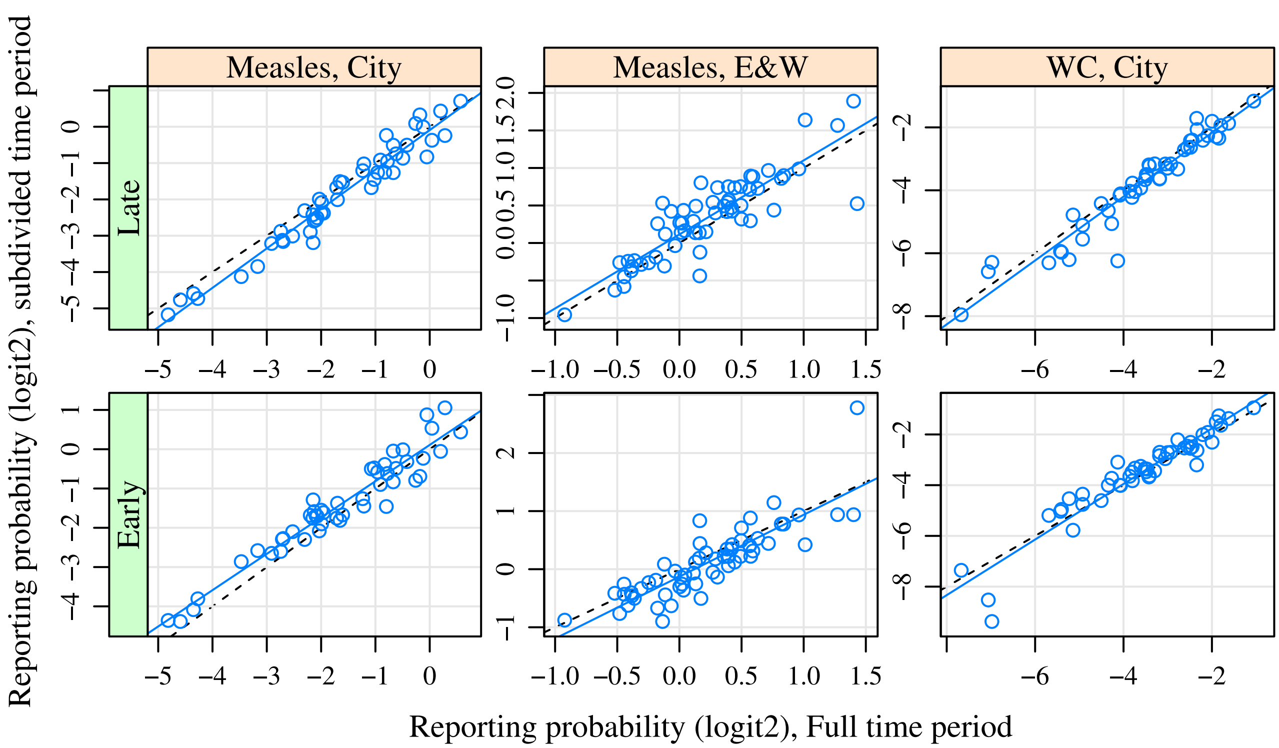

Figure S7. Time period comparison (per capita method, E&W = England & Wales, WC = whooping cough). For cities, case report timeseries were divided into early and late portions (see Table S1). Reporting probabilities estimated from these sub-periods generally match those from the full period. U.S. city measles is the exception, with more complete reporting in the early period. All values were logit_2 transformed to correct for heteroskedasticity. Black dashed line: 1-1 line; solid blue line: linear model.

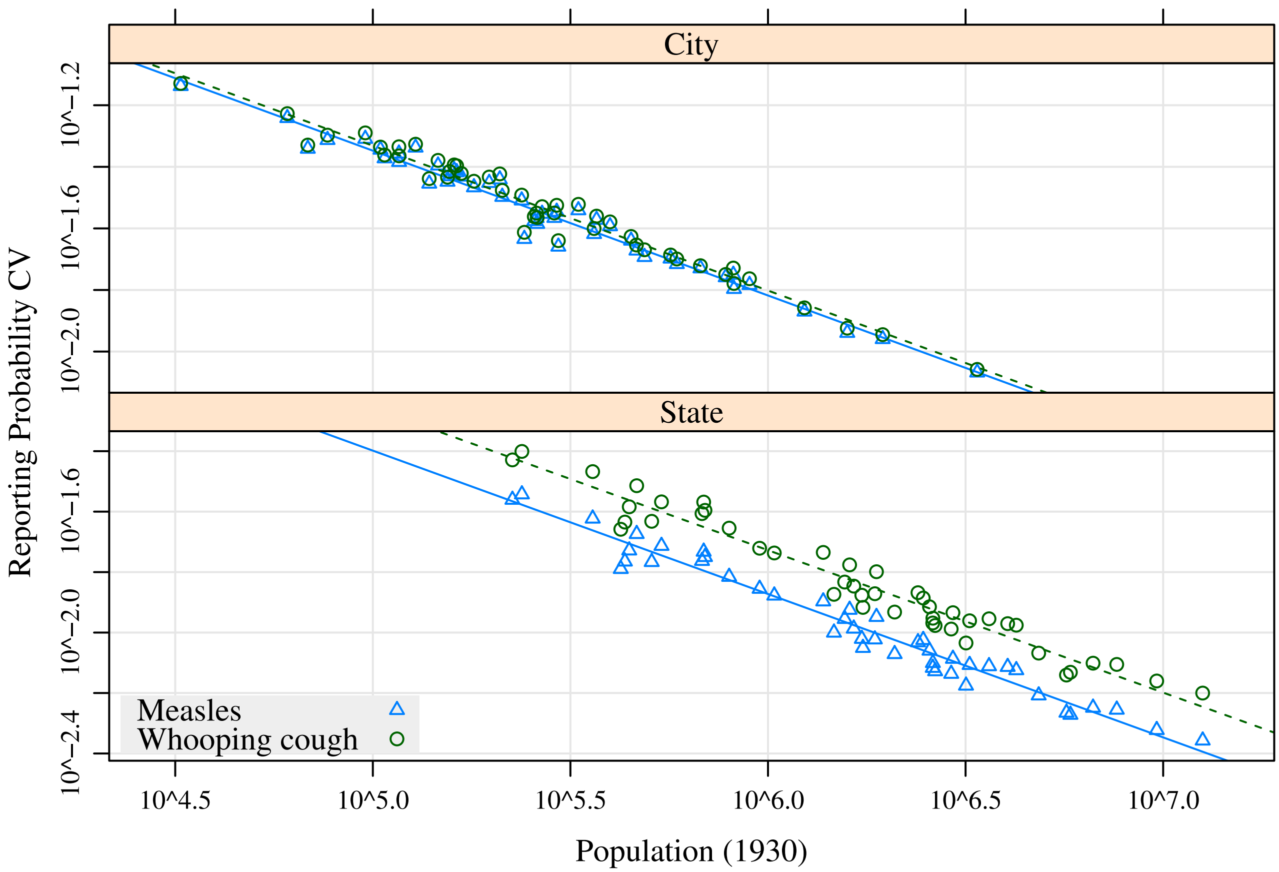

Figure S8. Coefficient of Variation (CV) of reporting probability bootstrap draws (microdata method) for cities (top) and states (bottom) with linear regression fits, showing power law scaling of normalized variation with population size (1930).

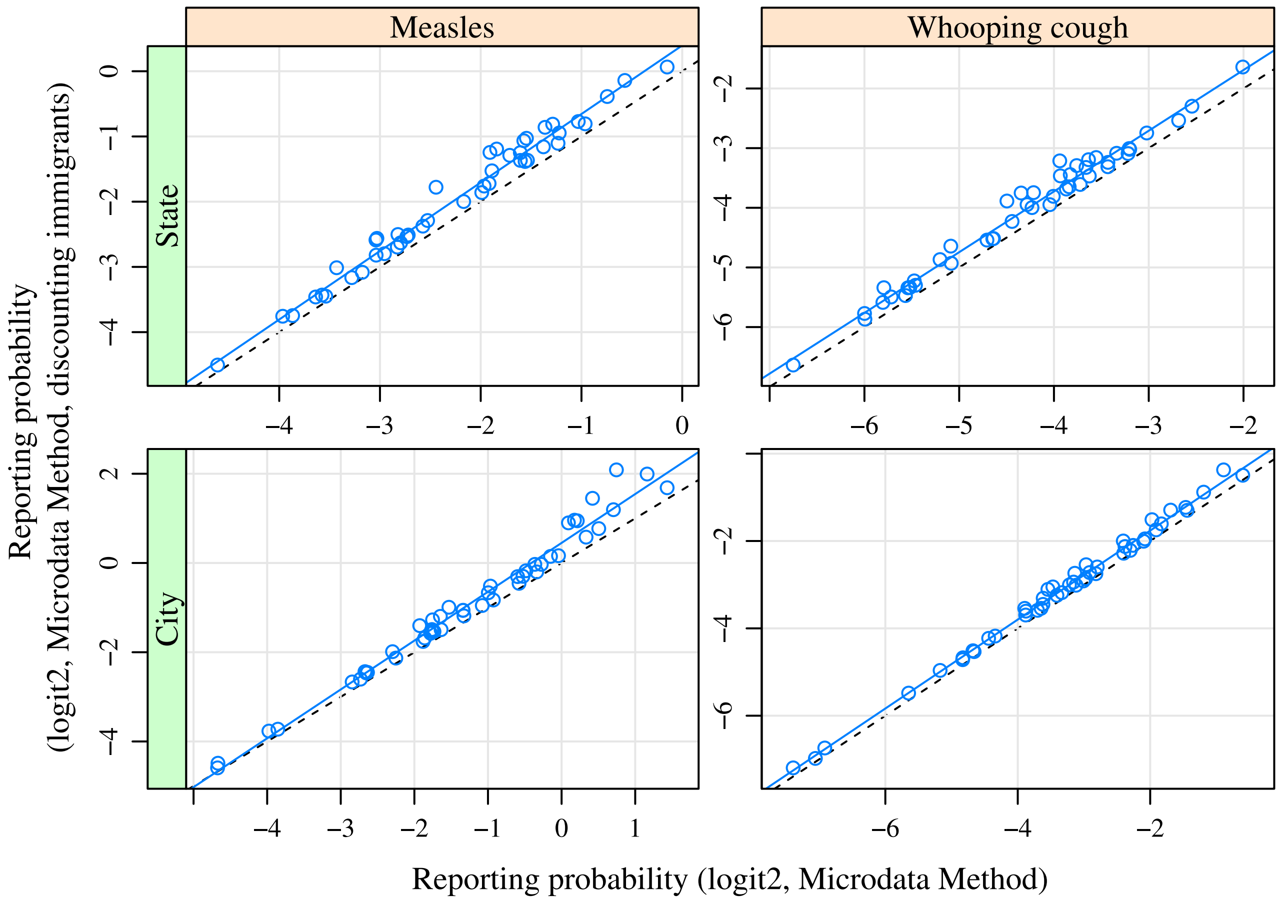

Figure S9. Effect of discounting observable immigration (i.e. out-of-state births). The y-axis shows the estimated reporting probabilities assuming that all immigrants have been infected prior to immigration, leading to fewer local susceptibles and higher reporting rates. Thus, this shows the upper bound on error due to immigration of recovered individuals. Inversely, intra-census emigration of locally-infected individuals would lower reporting rates, though no estimates of emigration are available. All values were logit_2 transformed to correct for heteroskedasticity. Black dashed line: 1-1 line; blue line: linear model.

Reference Data:

Case Reports, Demographic Covariates, and Reporting Estimates

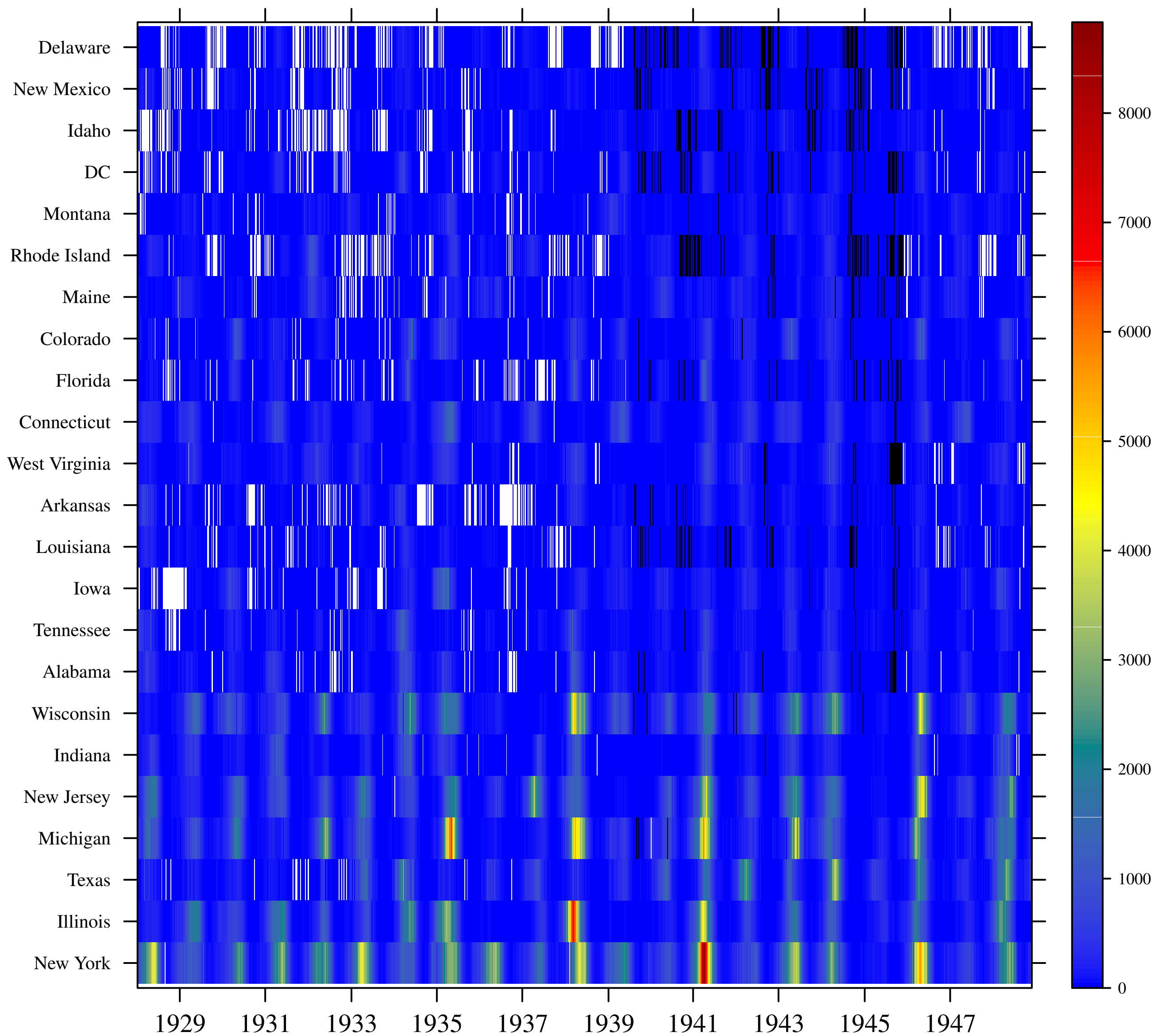

Figure S10. Measles, State. Case reports per sample period, with locations ordered by population size (black = 0; white = missing). Every other location is shown.

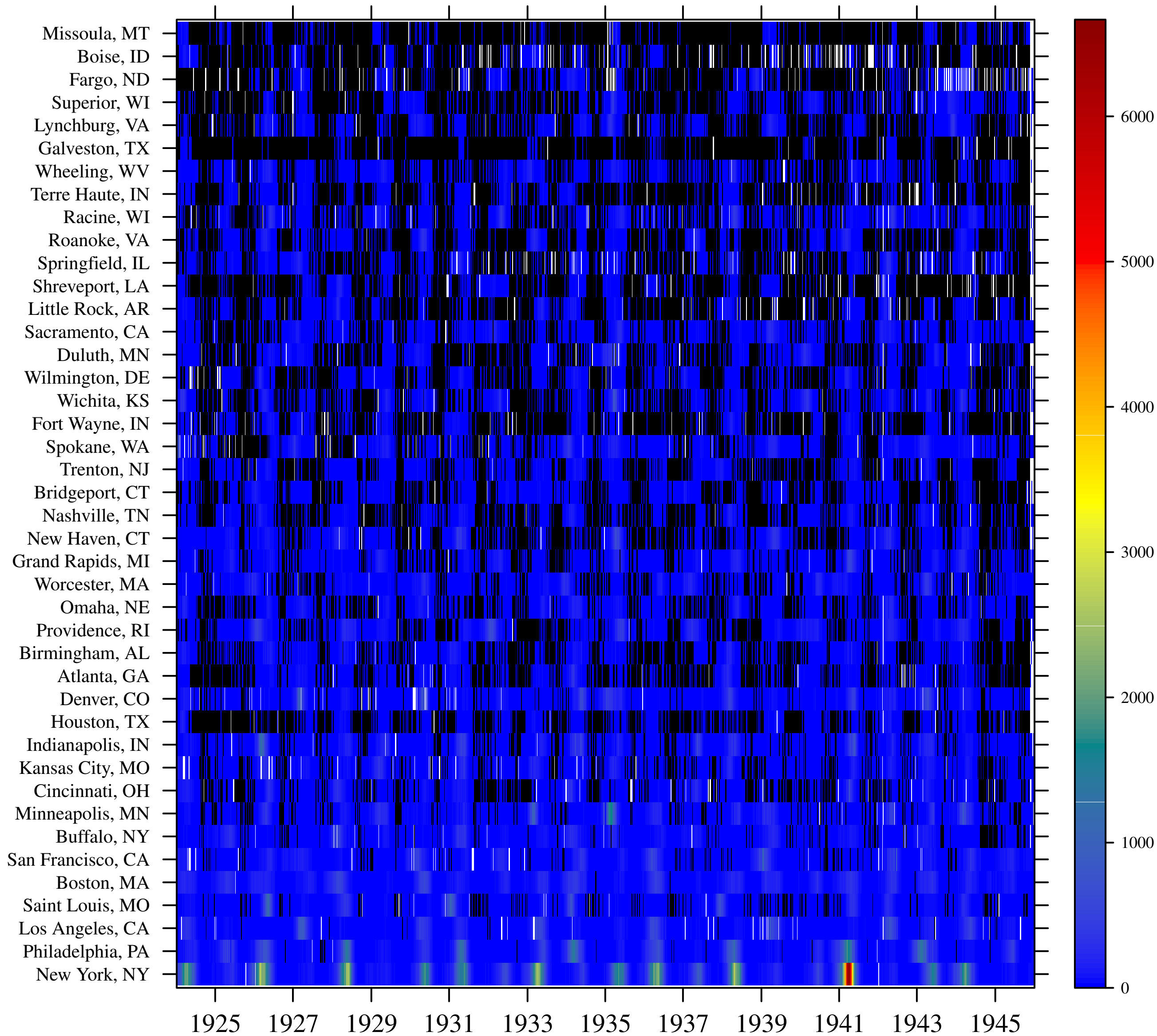

Figure S11. Measles, City. Case reports per sample period, with locations ordered by population size (black = 0; white = missing). Every other location is shown.

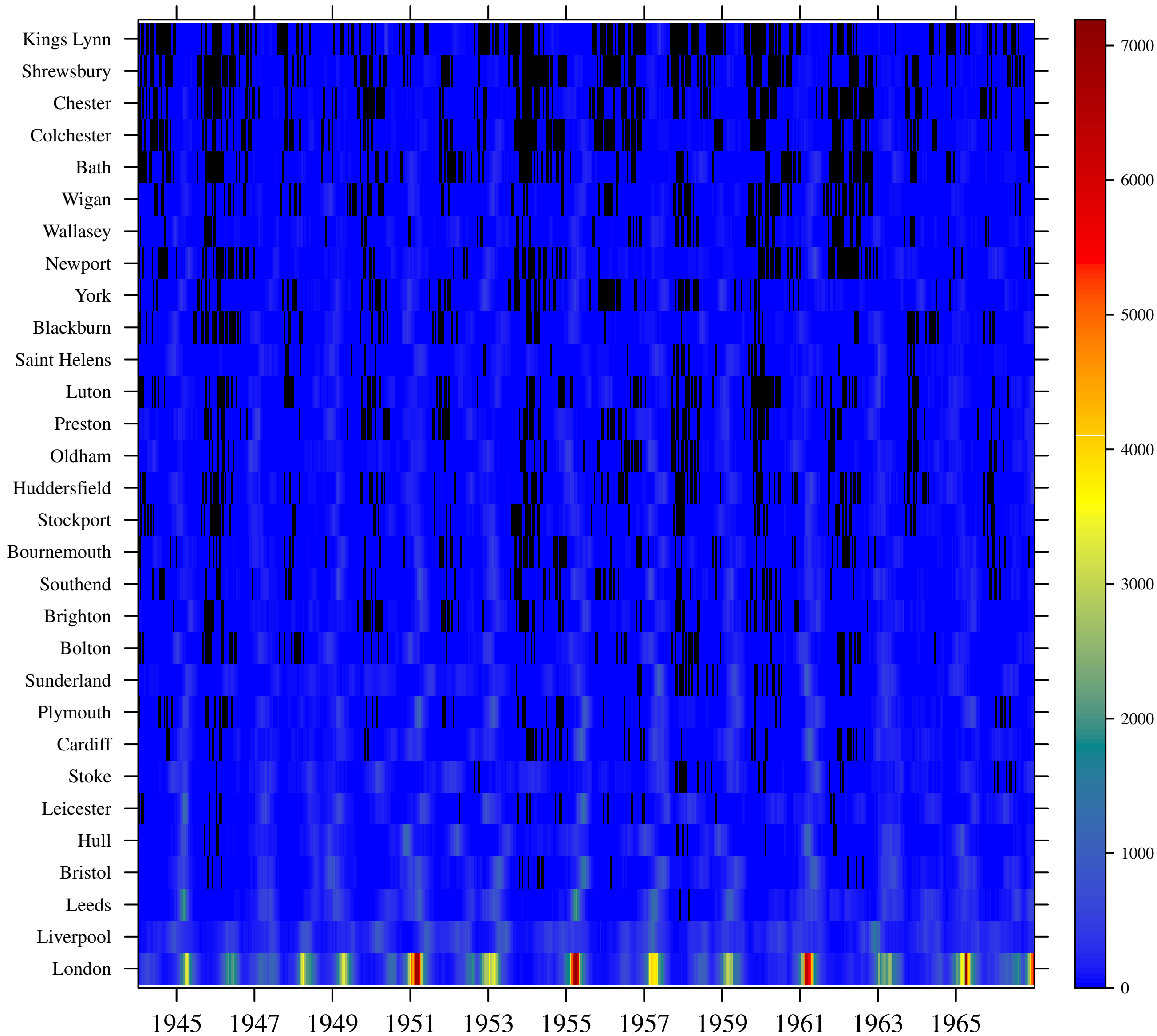

Figure S12. Measles, England & Wales. Case reports per sample period, with locations ordered by population size (black = 0; white = missing). Every other location is shown.

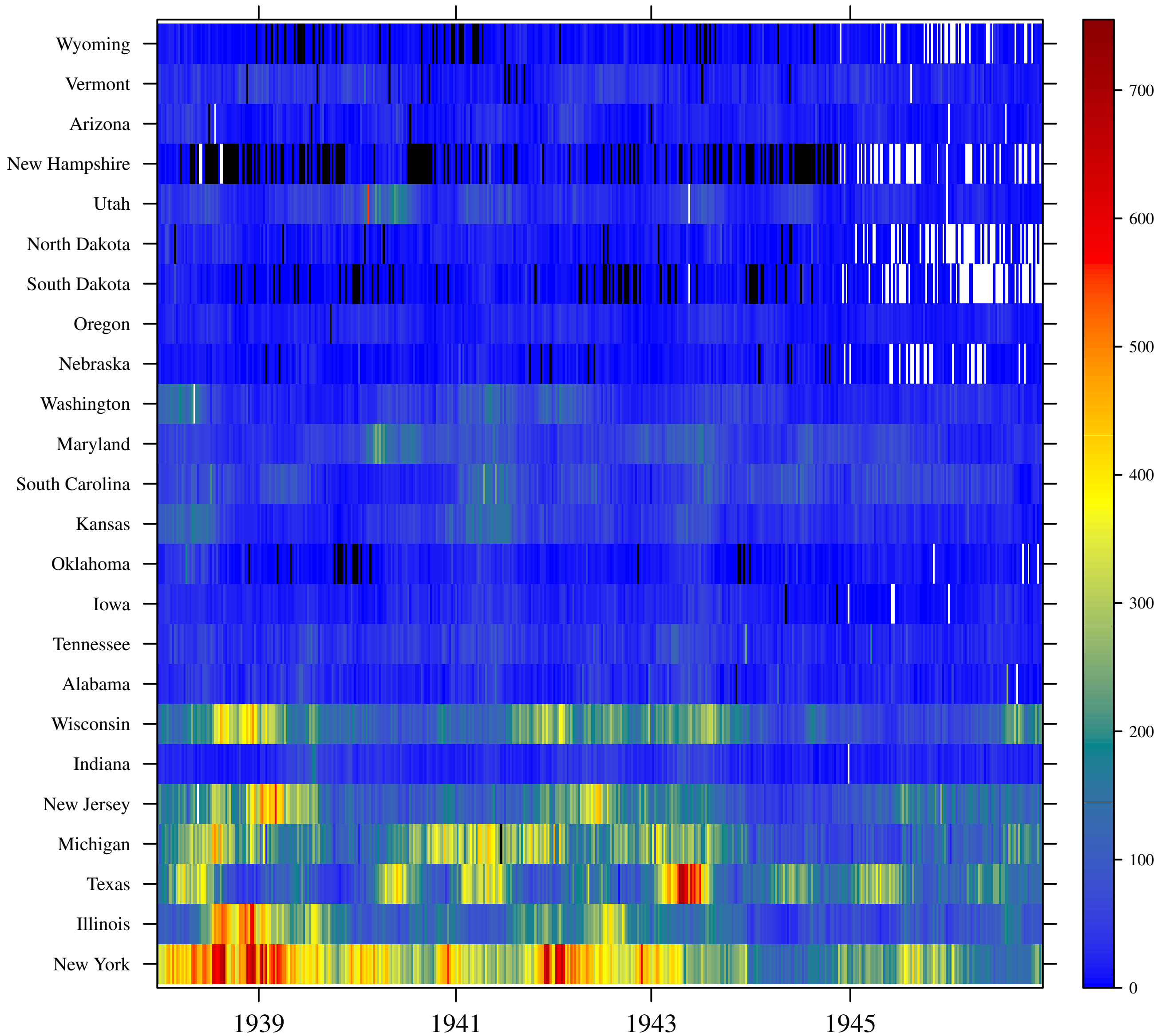

Figure S13. Whooping cough, State. Case reports per sample period, with locations ordered by population size (black = 0; white = missing). Every other location is shown.

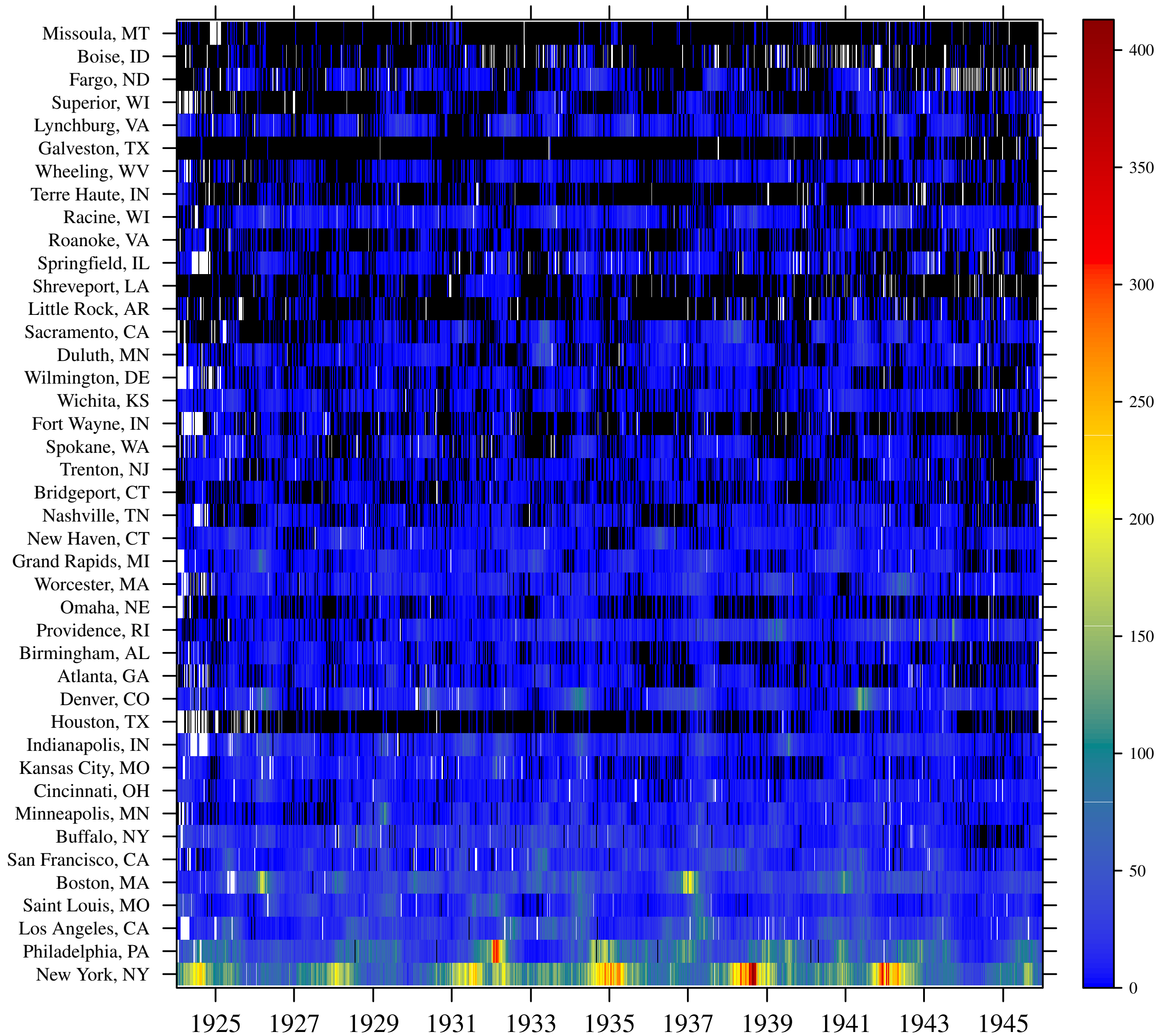

Figure S14. Whooping cough, City. Case reports per sample period, with locations ordered by population size (black = 0; white = missing). Every other location is shown.

	Population (1930)	White (%)	Measles (%)	CI	Whooping cough (%)	CI
Alabama	2.6e+06	65	7.7	(7.6, 7.8)	2.1	(2.1, 2.1)
Arizona	4.3e+05	86	15.2	(14.7, 15.7)	6.2	(5.9, 6.4)
Arkansas	1.9e+06	74	7.2	(7.0, 7.3)	2.2	(2.1, 2.2)
California	5.7e+06	95	25.0	(24.7, 25.3)	7.0	(6.9, 7.1)
Colorado	1e+06	99	27.2	(26.6, 28.0)	7.5	(7.2, 7.8)
Connecticut	1.6e+06	98	28.9	(28.2, 29.6)	9.1	(8.8, 9.4)
Delaware	2.4e+05	85	20.5	(19.4, 21.7)	5.0	(4.7, 5.4)
Florida	1.5e+06	71	8.3	(8.1, 8.4)	1.8	(1.7, 1.8)
Georgia	2.9e+06	63	6.2	(6.1, 6.3)	1.6	(1.5, 1.6)
Idaho	4.5e+05	99	10.5	(10.1, 10.8)	2.9	(2.8, 3.1)
Illinois	7.6e+06	96	19.7	(19.5, 20.0)	5.9	(5.8, 6.0)
Indiana	3.2e+06	96	12.7	(12.5, 12.9)	2.1	(2.1, 2.2)
Iowa	2.5e+06	99	11.0	(10.8, 11.2)	2.3	(2.2, 2.3)
Kansas	1.9e+06	96	24.4	(23.9, 25.0)	7.5	(7.3, 7.7)
Kentucky	2.6e+06	91	8.9	(8.8, 9.0)	4.0	(3.9, 4.1)
Louisiana	2.1e+06	64	3.8	(3.8, 3.9)	0.9	(0.9, 1.0)
Maine	8e+05	100	24.7	(23.9, 25.4)	10.0	(9.6, 10.5)
Maryland	1.6e+06	83	22.6	(22.1, 23.0)	8.0	(7.8, 8.2)
Massachusetts	4.2e+06	99	33.0	(32.5, 33.5)	9.9	(9.7, 10.1)
Michigan	4.8e+06	96	26.8	(26.5, 27.1)	8.6	(8.5, 8.8)
Minnesota	2.6e+06	99	17.5	(17.2, 17.8)	3.8	(3.7, 3.9)
Missouri	3.6e+06	94	10.4	(10.3, 10.6)	1.8	(1.8, 1.9)
Montana	5.4e+05	97	27.9	(26.9, 29.0)	6.8	(6.4, 7.2)
Nebraska	1.4e+06	98	12.6	(12.3, 12.9)	1.9	(1.8, 2.0)
New Hampshire	4.7e+05	100	11.8	(11.3, 12.3)	2.7	(2.6, 2.9)
New Jersey	4e+06	95	35.6	(35.1, 36.2)	11.1	(10.9, 11.3)
New Mexico	4.2e+05	92	11.0	(10.6, 11.3)	5.3	(5.1, 5.6)
New York	1.3e+07	97	24.7	(24.5, 24.9)	7.6	(7.6, 7.7)
North Carolina	3.2e+06	72	20.4	(20.1, 20.6)	8.7	(8.5, 8.8)
North Dakota	6.8e+05	98	13.7	(13.3, 14.2)	4.6	(4.4, 4.8)
Ohio	6.6e+06	95	20.0	(19.8, 20.2)	6.7	(6.6, 6.8)
Oklahoma	2.4e+06	89	5.9	(5.8, 6.0)	1.7	(1.6, 1.7)
Oregon	9.5e+05	98	21.6	(21.0, 22.2)	4.3	(4.2, 4.5)
Pennsylvania	9.6e+06	95	28.7	(28.4, 28.9)	7.1	(7.0, 7.2)
Rhode Island	6.9e+05	98	23.7	(22.9, 24.6)	10.0	(9.5, 10.5)
South Carolina	1.7e+06	54	10.1	(10.0, 10.3)	5.8	(5.6, 5.9)
South Dakota	6.9e+05	97	14.2	(13.7, 14.7)	2.4	(2.3, 2.5)
Tennessee	2.6e+06	81	7.5	(7.4, 7.7)	2.9	(2.9, 3.0)
Texas	5.8e+06	86	12.1	(12.0, 12.2)	6.5	(6.4, 6.6)
Utah	5.1e+05	98	31.6	(30.6, 32.7)	14.7	(14.1, 15.4)
Vermont	3.6e+05	100	39.2	(37.4, 41.1)	20.3	(19.0, 21.7)
Washington	1.6e+06	97	25.1	(24.6, 25.6)	6.3	(6.2, 6.5)
West Virginia	1.7e+06	94	12.0	(11.8, 12.2)	4.0	(3.9, 4.1)
Wisconsin	2.9e+06	99	45.8	(45.1, 46.6)	13.7	(13.4, 14.1)
Wyoming	2.3e+05	97	20.7	(19.6, 21.8)	4.9	(4.6, 5.3)

Table S4. States: Demographic covariates and estimated reporting probabilities of each disease (showing median and 95% CI).

	Population (1930)	White (%)	Measles (%)	CI	Whooping cough (%)	CI
Atlanta, GA	2.7e+05	62	10.2	(9.6, 10.7)	4.2	(3.9, 4.4)
Baltimore, MD	8.2e+05	81	34.7	(33.7, 35.8)	19.6	(18.9, 20.2)
Birmingham, AL	2.6e+05	61	11.0	(10.5, 11.7)	3.0	(2.9, 3.2)
Boston, MA	7.8e+05	98	34.4	(33.2, 35.6)	15.3	(14.7, 15.8)
Bridgeport, CT	1.5e+05	98	11.1	(10.3, 12.0)	3.2	(3.0, 3.5)
Buffalo, NY	5.7e+05	98	18.4	(17.7, 19.1)	9.3	(9.0, 9.7)
Charleston, WV	6.1e+04	92	20.7	(18.6, 23.3)	6.4	(5.7, 7.2)
Chicago, IL	3.4e+06	93	18.2	(17.9, 18.5)	8.1	(7.9, 8.2)
Cincinnati, OH	4.5e+05	91	14.2	(13.6, 14.9)	5.7	(5.4, 5.9)
Cleveland, OH	9e+05	92	34.1	(33.1, 35.3)	18.7	(18.1, 19.4)
Columbus, OH	2.9e+05	86	23.5	(22.3, 24.9)	8.7	(8.2, 9.2)
Dallas, TX	2.6e+05	85	17.8	(16.9, 18.8)	5.4	(5.1, 5.7)
Denver, CO	2.9e+05	98	46.6	(44.2, 49.2)	17.8	(16.9, 18.8)
Detroit, MI	1.6e+06	92	28.3	(27.7, 29.0)	14.5	(14.2, 14.9)
Flint, MI	1.5e+05	96	33.7	(31.5, 36.2)	10.6	(9.9, 11.4)
Fort Wayne, IN	1.2e+05	96	11.4	(10.6, 12.4)	1.7	(1.6, 1.9)
Grand Rapids, MI	1.7e+05	99	38.0	(35.4, 40.9)	15.4	(14.3, 16.6)
Hartford, CT	1.6e+05	97	23.3	(21.6, 25.1)	8.8	(8.2, 9.6)
Houston, TX	2.9e+05	78	3.0	(2.8, 3.1)	0.7	(0.7, 0.7)
Indianapolis, IN	3.6e+05	87	36.1	(34.5, 37.9)	11.2	(10.6, 11.7)
Kansas City, MO	4e+05	91	20.1	(19.2, 21.2)	5.6	(5.4, 5.9)
Los Angeles, CA	1.2e+06	94	16.5	(16.1, 16.9)	6.8	(6.6, 6.9)
Memphis, TN	2.6e+05	60	17.9	(17.0, 18.9)	8.7	(8.3, 9.2)
Milwaukee, WI	5.9e+05	99	49.9	(48.1, 51.9)	24.1	(23.2, 25.1)
Mobile, AL	6.8e+04	67	4.7	(4.3, 5.2)	0.5	(0.5, 0.6)
Nashville, TN	1.6e+05	72	14.4	(13.4, 15.6)	6.6	(6.1, 7.1)
New Haven, CT	1.6e+05	95	41.9	(38.9, 45.2)	14.4	(13.3, 15.6)
New Orleans, LA	4.6e+05	71	5.4	(5.2, 5.6)	3.4	(3.3, 3.5)
Omaha, NE	2.1e+05	95	19.1	(17.9, 20.3)	2.4	(2.2, 2.6)
Philadelphia, PA	2e+06	88	24.0	(23.5, 24.6)	10.1	(9.9, 10.3)
Pittsburgh, PA	6.7e+05	92	29.8	(28.7, 30.9)	11.3	(10.9, 11.7)
Providence, RI	2.4e+05	98	46.0	(43.3, 48.9)	16.9	(15.8, 18.0)
Raleigh, NC	3.3e+04	61	42.6	(37.1, 49.2)	30.5	(26.5, 35.4)
Richmond, VA	1.8e+05	71	39.6	(37.1, 42.4)	3.0	(2.8, 3.2)
Rochester, NY	3.3e+05	100	27.0	(25.5, 28.6)	9.8	(9.3, 10.5)
Sacramento, CA	9.6e+04	93	43.3	(39.4, 47.8)	20.3	(18.4, 22.5)
Saint Louis, MO	8.2e+05	89	19.5	(18.8, 20.2)	7.7	(7.4, 8.0)
Salt Lake City, UT	1.4e+05	98	56.9	(53.3, 61.1)	26.8	(25.0, 28.9)
San Antonio, TX	2.4e+05	93	3.0	(2.8, 3.1)	0.6	(0.6, 0.7)
Seattle, WA	3.7e+05	96	41.8	(39.7, 44.1)	13.8	(13.0, 14.5)
South Bend, IN	1e+05	96	11.3	(10.4, 12.4)	3.9	(3.6, 4.3)
Spokane, WA	1.2e+05	99	50.6	(46.5, 55.3)	9.8	(9.0, 10.8)
Syracuse, NY	2.1e+05	100	63.0	(58.6, 67.6)	35.8	(33.4, 38.7)
Tacoma, WA	1.1e+05	97	26.6	(24.6, 29.0)	6.4	(5.9, 7.0)
Trenton, NJ	1.3e+05	94	19.2	(17.6, 21.0)	6.7	(6.2, 7.4)
Washington, DC	4.9e+05	72	20.8	(20.0, 21.6)	7.2	(6.9, 7.5)
Winston-Salem, NC	7.7e+04	56	53.3	(48.6, 58.8)	24.4	(22.1, 27.0)
Worcester, MA	2e+05	100	37.7	(35.3, 40.5)	17.4	(16.2, 18.7)

Table S5. Cities: Demographic covariates and estimated reporting probabilities of each disease (showing median and 95% CI).