

Methods for combining data from multi-jurisdiction sentinel surveillance projects

Elizabeth Torrone¹, Eloisa Llata¹, Jaeyoung Hong¹, Preeti Pathela²

¹DSTDP, CDC, Atlanta, Georgia, United States, ²New York City Department of Health and Mental Hygiene, New York City, New York, United States

Objective

To identify best practices for combining public health data for multi-jurisdiction surveillance projects.

Introduction

Sentinel surveillance, where selected jurisdictions follow standardized protocols to collect and report enhanced public health data not available through other routine surveillance efforts, is a key part of national surveillance of sexually transmitted diseases (STDs). Although four STDs are nationally notifiable conditions (chlamydia, gonorrhea, syphilis and chancroid), the burden of these conditions (over 2.3 million cases were reported in 2017) limits the amount of detailed clinical and demographic data available for all cases. Sentinel surveillance in clinical settings serving at-risk populations, such as STD clinics, provides an opportunity to collect enhanced data elements on persons seeking STD-related services, such as sex of sex partners and anatomic site of infection.

However, there are challenges in combining data across jurisdictions as estimated effect measures may vary by jurisdiction (e.g., some may have higher observed burden of disease among certain populations) and the amount of data contributed by jurisdiction may vary; combined this could lead to biased estimates if heterogeneity is not taken into account.

Methods

Using data from the STD Surveillance Network (SSuN), a sentinel surveillance project implemented in 10 jurisdictions, we investigated the effect of using different statistical methods to combine data across jurisdictions. We evaluated 5 methodologies:

- “Fully stratified” where estimates were provided separately for each jurisdiction;
- “Aggregated” where numerators and denominators were summed across jurisdictions without weighting;
- “Mean estimate” where the mean of the jurisdiction-specific estimates was estimated;
- “Random effects” where jurisdiction-specific estimates were combined using an inverse variance weighted random effects model to adjust for heterogeneity between jurisdictions; and
- “Stratified random effects” where a possible effect modifier was identified and used to group jurisdictions prior to calculating the estimate from the random effects model.

Through SSuN, jurisdictions collect visit-level data on patients attending selected STD clinics and report clinical and demographic data. As an illustrative example, we estimated rectal gonorrhea positivity among gay, bisexual, and other men who have sex with men (MSM) attending participating clinics. Jurisdiction-specific positivity was estimated as the # of unique MSM testing positive at least once for rectal gonorrhea divided by all MSM tested 1 or more times for rectal gonorrhea in all of the clinics in the jurisdiction. The stratifying variable for the stratified random effects method was the percent of MSM screened in the jurisdiction’s clinics, as low screening coverage may reflect targeted testing of MSM likely to be infected which may inflate observed positivity. For each of the five methods, we estimated rectal gonorrhea positivity and the corresponding 95% confidence interval (CI).

Results

In 2017, 123,210 patients attended 30 STD clinics participating in the 10 SSuN jurisdictions, of which 31,052 (25.2%) were identified as MSM (jurisdiction-specific range: 8.8% to 70.0%). (Table 1) One jurisdiction (I) accounted for 39% of all MSM



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

included in the analysis while one jurisdiction (J) accounted for only 1.6% of MSM included. The proportion of MSM tested for rectal gonorrhea at least once varied by jurisdiction, ranging from 44.3% to 76.9%. The fully stratified method identified differences in rectal gonorrhea positivity across jurisdictions, with jurisdiction-specific positivity ranging from 9.9% to 24.1%. Aggregating across jurisdictions masked this heterogeneity and provided a single summary estimate of 15.2% (95% CI: 14.7, 15.7). Taking the mean across the jurisdiction-specific estimates also provided a summary estimate; however, the uncertainty of the estimate increased (15.8%, 95% CI: 13.3, 18.7). Accounting for the heterogeneity by using a random effects model resulted in an estimate of 15.5% (95% CI: 13.9, 17.2). After stratifying by a likely confounder (% of MSM screened); the random effects estimate among 3 jurisdictions with lower screening coverage (<60%) was 19.7% (95% CI: 14.6, 24.8) and among 7 jurisdictions with higher screening coverage (≥60%) was 14.3% (95% CI: 12.9, 15.7).

Conclusions

In a sentinel surveillance project implemented in 10 jurisdictions, there was substantial heterogeneity in the observed proportion of MSM testing positive for rectal gonorrhea in selected STD clinics. Although a stratified analysis captured the heterogeneity across jurisdictions, it may not be feasible to present fully stratified estimates for all analyses (e.g., surveillance reports likely provide metrics for multiple diseases). Additionally, it limits the ability to succinctly communicate key findings. Aggregating numerators and denominators across jurisdictions to calculate a single summary estimate masks this heterogeneity and biases estimates toward high volume jurisdictions. Taking the mean across jurisdictions ensures that high-volume jurisdictions do not bias the overall estimate; however, the mean may be biased by very high or very low positivity estimates in a few jurisdictions. Using a random effects model accounted for both varying sample sizes and differences in observed heterogeneity; although the summary estimate was similar to the aggregate in this example, the wider 95% CI more accurately reflects the uncertainty in the estimate. Finally, stratifying by a likely effect measure modifier (% of MSM screened) prior to estimating the measure from the random effects model captured key differences in jurisdictions while still providing a limited number of summary estimates. Analysts using data from multi-jurisdiction surveillance projects should fully investigate possible biases when combining estimates across jurisdictions. If there is observed heterogeneity across jurisdictions and it is not feasible to provide fully stratified estimates, analysts could consider using methods to account for heterogeneity and minimize bias due to differing sample sizes, such as stratified random effects models.

Acknowledgement

The authors thank the members of the SSuN Working Group: Bob Kohn, Robbie Madera, and Roxanne Kerani, as well as the SSuN jurisdictions that contributed data to this analysis.

Table 1. Trends in rectal gonorrhea among MSM tested in STD clinics participating in the STD Surveillance Network (SSuN), 2017

Jurisdiction	# of patients	% of patients identified as MSM	% of MSM patients screened for rectal gonorrhea	% MSM testing positive for rectal gonorrhea				
				Fully stratified	Aggregated across jurisdictions	Mean of jurisdiction estimates	Random effects model	Stratified random effects model
A	6,262	14.3%	44.3%	14.4%	n/a	n/a	n/a	19.7% (95% CI: 14.6, 24.8)
B	8,160	8.8%	46.4%	24.1%	n/a	n/a	n/a	
C	13,519	15.7%	58.5%	20.9%	n/a	n/a	n/a	
D	6,017	44.6%	65.6%	18.2%	n/a	n/a	n/a	14.3% (95% CI: 12.9, 15.7)
E	15,083	16.8%	65.8%	14.7%	n/a	n/a	n/a	
F	9,188	48.4%	67.8%	15.8%	n/a	n/a	n/a	



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISDS 2019 Conference Abstracts

G	10,657	30.1%	68.1%	13.5%	n/a	n/a	n/a	12.9, 15.7)
H	3,665	46.2%	70.5%	12.3%	n/a	n/a	n/a	
I	49,792	24.4%	71.7%	14.3%	n/a	n/a	n/a	
J	867	70.0%	76.9%	9.9%	n/a	n/a	n/a	
TOTAL	123,210	25.2%	67.5%	n/a	15.2% (95% CI: 14.7, 15.7)	15.8% (95% CI: 13.3, 18.7)	15.5% (95% CI: 13.9, 17.2)	n/a

MSM: Gay, bisexual and other men who have sex with men; CI: Confidence interval



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.