



# BMJ Open STI epidemic re-emergence, socio-epidemiological clusters characterisation and HIV coinfection in Catalonia, Spain, during 2017–2019: a retrospective population-based cohort study

Alexis Sentís <sup>1,2,3</sup>, Marcos Montoro-Fernandez,<sup>1,4</sup> Evelin Lopez-Corbeto,<sup>1,4,5</sup> Laia Egea-Cortés,<sup>1,4</sup> Daniel K Nomah,<sup>1,6</sup> Yesika Díaz,<sup>1,4</sup> Patricia Garcia de Olalla,<sup>5,7</sup> Lilas Mercuriali,<sup>7</sup> Núria Borrell,<sup>8</sup> Juliana Reyes-Urueña <sup>1,4,5</sup>, Jordi Casabona,<sup>1,4,5,6</sup>  
The Catalan HIV and STI Surveillance Group

**To cite:** Sentís A, Montoro-Fernandez M, Lopez-Corbeto E, *et al.* STI epidemic re-emergence, socio-epidemiological clusters characterisation and HIV coinfection in Catalonia, Spain, during 2017–2019: a retrospective population-based cohort study. *BMJ Open* 2021;**11**:e052817. doi:10.1136/bmjopen-2021-052817

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-052817>).

Received 08 May 2021  
Accepted 12 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Alexis Sentís;  
[a.sentis@epiconcept.fr](mailto:a.sentis@epiconcept.fr)

## ABSTRACT

**Objectives** To describe the epidemiology of sexually transmitted infections (STIs), identify and characterise socio-epidemiological clusters and determine factors associated with HIV coinfection.

**Design** Retrospective population-based cohort.

**Setting** Catalonia, Spain.

**Participants** 42 283 confirmed syphilis, gonorrhoea, chlamydia and lymphogranuloma venereum cases, among 34 600 individuals, reported to the Catalan HIV/STI Registry in 2017–2019.

**Primary and secondary outcomes** Descriptive analysis of confirmed STI cases and incidence rates. Factors associated with HIV coinfection were determined using logistic regression. We identified and characterized socio-epidemiological STI clusters by Basic Health Area (BHA) using K-means clustering.

**Results** The incidence rate of STIs increased by 91.3% from 128.2 to 248.9 cases per 100 000 population between 2017 and 2019 ( $p < 0.001$ ), primarily driven by increase among women (132%) and individuals below 30 years old (125%). During 2017–2019, 50.1% of STIs were chlamydia and 31.6% gonorrhoea. Reinfections accounted for 10.8% of all cases and 6% of cases affected HIV-positive individuals. Factors associated with the greatest likelihood of HIV coinfection were male sex (adjusted OR (aOR) 23.69; 95% CI 16.67 to 35.13), age 30–39 years (versus <20 years, aOR 18.58; 95% CI 8.56 to 52.13), having 5–7 STI episodes (vs 1 episode, aOR 5.96; 95% CI 4.26 to 8.24) and living in urban areas (aOR 1.32; 95% CI 1.04 to 1.69). Living in the most deprived BHAs (aOR 0.60; 95% CI 0.50 to 0.72) was associated with the least likelihood of HIV coinfection. K-means clustering identified three distinct clusters, showing that young women in rural and more deprived areas were more affected by chlamydia, while men who have sex with men in urban and less deprived areas showed higher rates of STI incidence, multiple STI episodes and HIV coinfection.

**Conclusions** We recommend socio-epidemiological identification and characterisation of STI clusters and factors associated with HIV coinfection to identify at-risk

## Strengths and limitations of this study

- In this retrospective population-based cohort study, the use of data from the Catalan HIV/STI Registry allowed us to characterise the re-emergence of sexually transmitted infections (STIs), perform socio-epidemiological clustering and reveal factors associated with HIV coinfection.
- To the best of our knowledge, this is the first study to apply the k-means clustering methodology to identify and characterise distinct socio-epidemiological clusters of STI at a small health area level.
- A key limitation of this study is the high proportion of missing data around sociodemographic and lifestyle characteristics such as education level, sexual preference and country of birth. Nonetheless, our findings are consistent with previous analyses.

populations at a small health area level to design effective interventions.

## INTRODUCTION

The epidemic of sexually transmitted infections (STIs) continues to be a major concern and threat to global public health. Undiagnosed and untreated STIs can lead to a multitude of complications including HIV acquisition, long-term disabilities, infertility, adverse pregnancy outcomes and death.<sup>1 2</sup> Across Europe, incidence of STIs continue to be on the rise with confirmed cases reported in national surveillance systems increasing by 50% for gonorrhoea, 36% for syphilis, 68% for lymphogranuloma venereum (LGV) and 0.6% for chlamydia from 2014 to 2018.<sup>3–6</sup> This trend is reflected in Spain where new STI cases have been reported to increase 10-fold from 2000 to 2017, with 23 975 cases

of gonorrhoea, syphilis, chlamydia and LGV reported in 2017 alone.<sup>78</sup> During 2018 to 2019, the region of Catalonia in Spain recorded the highest incidence of STIs across the country, with a rise of 37% in the number of cases.<sup>9</sup> Incidence rates were highest among men who have sex with men (MSM), women and in young adults, particularly among young women who in recent years have shown a proportionally higher increase than men.<sup>79</sup> The surge in STI incidence rates may be explained by improvements in surveillance systems, introduction of new diagnostic methods with enhanced sensitivity, changes in sexual attitudes and behaviours, sociocultural shifts in society and the effects of tourism and globalisation.<sup>10</sup>

STIs and HIV infections are overlapping epidemics, which, besides from biological synergies, are largely driven by socioeconomic and other contextual factors acting as syndemics. Individuals affected by STIs are at increased risk of HIV infection and people living with HIV are more vulnerable to STIs.<sup>11 12</sup> Some studies have described social determinants of health, discrimination and inequalities as the main factors associated with the spatiotemporal clustering of STI cases.<sup>13 14</sup> While spatiotemporal clustering may be useful in grouping events or cases, other methodologies including k-means clustering allow grouping of different geographical units by common characteristics such as sociodemographic and epidemiological factors.<sup>15</sup> The socio-epidemiological characterisation of STIs, including association with HIV coinfection, and identification of distinct clusters are imperative to strengthen the integrated surveillance of STIs and HIV. Data from such an exercise could potentially increase the sensitivity, timeliness and representativeness of surveillance systems, and generate information to tailor public health strategies to tackle a continuously growing epidemic.

Therefore, we aimed to describe the epidemiology of STIs, identify and characterise socio-epidemiological clusters of STI and determine factors associated with HIV coinfection in Catalonia, Spain, during 2017–2019.

## METHODS

### Study design and data source

We conducted a retrospective population-based cohort analysis of all confirmed cases of the notifiable STIs, syphilis, gonorrhoea, chlamydia and LGV, in Catalonia between 1 January 2017 and 31 December 2019. Data were obtained from the Catalan HIV/STI Registry,<sup>16</sup> which uses information from the Epidemiological Repository of Catalonia (REC, in Catalan), an electronic database used by the Epidemiological Surveillance Network of Catalonia (XVEC, in Catalan). REC collects information from two sources: (1) the microbiological notification system (SNM, in Catalan) of confirmed cases from microbiological laboratories; and (2) the mandatory disease notification system (MDO, in Catalan) based on physician reporting of clinically suspected/probable and laboratory-confirmed cases as per established case definitions. Information collected through an epidemiological

questionnaire that records clinical, epidemiological and behavioural variables are included along with the mandatory notification in REC (online supplemental table S1). Case definitions for surveillance reporting are standardised according to the European Union definitions established by the European Centre for Disease Prevention and Control (online supplemental table S2).<sup>17 18</sup>

### Analysis variables

We extracted data around epidemiological, sociodemographic and clinical variables as detailed in online supplemental table S3. All individuals who had experienced at least one STI episode during the study period were linked, through the Spanish healthcare system personal identification code (CIP), to the Catalan HIV/STI Registry to identify HIV coinfections either before or after the recorded STI episode. In addition to the CIP, the Catalan HIV/STI Registry surveillance team performs checks of duplicates at least twice annually using a unique STI episode number (assigned to each notification and disease), name and date of birth. For our analysis, a deduplicated, HIV/STI-linked and anonymised version was provided.

A Basic Health Area (BHA; Àrea Bàsica de Salut (ABS), in Catalan) is a territorial unit of coverage served by a primary healthcare team. Each BHA typically serves a population of approximately 5000–25 000 people. The socioeconomic level of the BHAs was classified according to a deprivation index (calculated by the Agency of Health Quality and Assessment of Catalonia) which was attributed to each individual according to their residential address. The deprivation index is a composite measure based on indicators such as proportion of residents with low educational level, proportion of manual workers, proportion of residents with an annual income below a specified amount, and rate of premature mortality. Deprivation indices were categorised in quintiles, with the first quintile being the least deprived.<sup>19</sup>

Clinical variables that were extracted included reinfections, multiple STI episodes and coinfection with HIV. Reinfection was defined as an episode of the same STI detected after a defined period, which differed for each STI, following the previously recorded infection in the same individual during the study period. Multiple STI episodes were defined as total number of episodes of any STI reported for the individual during the study period (online supplemental table S3). As information regarding treatment response was not available, episodes occurring outside of the specific time frames for each STI were assumed not to be a persistent infection resulting from treatment failure.

### K-means clustering of STIs

We implemented k-means to define STI clusters by sociodemographic characteristics. Specifically, the k-means clustering methodology is an unsupervised machine learning approach that seeks to group heterogeneous units (in our case BHAs) into clusters based on similarities

in characteristics (variable values and categorical distribution among the BHA).<sup>15</sup> A clustering algorithm is a procedure for grouping a series of vectors according to a specific criterion, which could be distance or similarity. Proximity is defined in terms of a distance function and uses a k-means clustering method based on Euclidean distance to quantify similarities or differences between observations. This method is sensitive to outliers and requires both internal and external validation processes using different combinations of variables in the algorithms. The internal clustering validation considered an average Euclidean distance for each cluster and similarities between cases according to the correlation matrix of distances to determine the optimum number of clusters for which intracluster variation is minimum. Our external validation process was based on a description of the possible socio-epidemiological variables to determine the most appropriate clustering algorithm for our data set. Based on these validations, the researchers ended up with an exact number of clusters formed by defined variables which were included in the algorithm. In our case, the following variables were chosen to identify and build the final three socio-epidemiological clusters of STI by BHA: incidence rate by each STI, percentage of women, percentage of people with HIV coinfection, median age among all STI cases in each BHA and deprivation index of each BHA.

### Statistical analyses

We performed a descriptive analysis to summarise epidemiological, clinical, sociodemographic and geographical variables for the total confirmed STI cases and by STI clusters. Continuous variables were summarised as median and interquartile range (IQR), while categorical variables were reported as absolute frequencies and percentages. Annual incidence rates of STIs are described per 100 000 population for the total confirmed STI cases and by STI cluster, and calculated based on census information from the Statistical Institute of Catalonia (IDESCAT) (online supplemental table S4). Incidence trends were analysed using the  $\chi^2$  test for linear trend. For identifying and building clusters in k-means clustering, STI incidence rates were described per 1000 population due to the small population size per BHA.

We assessed risk factors associated with HIV coinfection among individuals diagnosed with STIs using multivariable logistic regression models to estimate odds ratios (ORs) and 95% confidence intervals (CIs). Individuals with more than one STI episode were counted once (first episode), and successive episodes in the same individual were grouped in a variable that considers the number of episodes, and included in the models. Sexual preference, country of birth and education level were excluded from the models because more than 50% of values were missing. We used backward stepwise elimination regression to include all analysed variables that showed statistical significance ( $p < 0.05$ ) by the Wald test in the final

multivariable logistic regression model. All analyses were performed using R Statistical Software V.3.6.1).

### Ethics approval statement

Data from the Catalan HIV/STI Registry, REC and all aggregated variables used in the study were handled according to international recommendations, the Helsinki Declaration revised by the World Medical Organization in Fortaleza in 2013 and Spanish Law 3/2018 on Data protection and Public Health 33/2011. Patient information was anonymised and deidentified prior to analysis and therefore no informed consent was required.

### Patient and public involvement

Patients were not directly involved in this study; only data from the nationally notifiable disease surveillance system were used.

## RESULTS

### STI epidemic and trends

Between 2017 and 2019, a total of 42 283 cases of STIs were reported among 34 600 individuals in Catalonia (table 1). Throughout the study period, half of all reported STIs were chlamydia (50.1%) and almost a third were gonorrhoea (31.6%). Reinfections accounted for 10.8% of all reported cases. Among the subjects affected by STIs, the events of gonorrhoea had the highest reinfection rate (15.7%), while chlamydia had the lowest occurrence of reinfection (6.7%).

The number of STI cases doubled from 9687 in 2017 to 18872 in 2019 (table 2). The incidence rate of STIs increased by 91.3% from 128.2 cases per 100 000 population in 2017 to 248.9 cases per 100 000 population in 2019. The annual incidence rate of STIs for the period 2017–2019 was 185.5 per 100 000 population. Incidence rates increased significantly ( $p < 0.001$ ) from 2017 to 2019 for all STI types except for syphilis cases which remained stable over the 3 years, with the highest increase in number of cases seen in chlamydia (188.8%) followed by gonorrhoea (63.8%) and LGV (56.1%). In 2017, chlamydia and gonorrhoea represented 36.8% and 36.0% of all reported STIs, respectively, but by 2019 chlamydia accounted for 55.1% of all cases. Gonorrhoea showed the second greatest increase from 2017 to 2019, with 47.5% occurring in individuals under 30 years of age throughout the study period. This increase in the number of confirmed STI cases from 2017 to 2019 was remarkably higher in women (132% vs 75% in men) and individuals below the age of 30 years (125% vs 68% in those  $\geq 30$  years). Indeed, women under 30 years of age presented the highest decrease in both number of cases, with an increase of 155.8% versus 93.6% in men below 30 years and in incidence rates with an increase of 154.1% (from 193.6 to 491.9 per 100 000 population) versus 93.6% in men under 30 years (from 202.9 to 384.0 per 100 000 population).

The vast majority of reported cases occurred in men for all STI types except chlamydia, of which 61.9%

**Table 1** Epidemiological characteristics of reported STI cases in Catalonia, Spain (2017–2019)

Characteristic	All STIs	Chlamydia	Gonorrhoea	Syphilis	LGV
Total cases, n (%)	42283 (100)	21202 (50.1)	13362 (31.6)	6975 (16.5)	744 (1.8)
<i>Sex, n (%)</i>					
Female	16676 (39.4)	13125 (61.9)	2667 (20.0)	875 (12.5)	9 (1.2)
Male	25607 (60.6)	8077 (38.1)	10695 (80.0)	6100 (87.5)	735 (98.8)
<i>Age, n (%)</i>					
<20 years	4438 (10.5)	3311 (15.6)	984 (7.4)	137 (2.0)	6 (0.8)
20–29 years	17691 (41.8)	10707 (50.5)	5361 (40.1)	1462 (21.0)	161 (21.6)
30–39 years	11102 (26.3)	4454 (21.0)	4116 (30.8)	2242 (32.1)	290 (39.0)
40–49 years	6092 (14.4)	2087 (9.8)	2032 (15.2)	1757 (25.2)	216 (29.0)
50–59 years	2037 (4.8)	530 (2.5)	658 (4.9)	789 (11.3)	60 (8.1)
>60 years	923 (2.2)	113 (0.5)	211 (1.6)	588 (8.4)	11 (1.5)
<i>Sexual preference, n (%)</i>					
MSM*	3270 (7.7)	785 (3.7)	1321 (9.9)	993 (14.2)	171 (23.0)
MSW	3149 (7.5)	1863 (8.8)	1040 (7.8)	243 (3.5)	3 (0.4)
WSW†	415 (1.0)	335 (1.6)	69 (0.5)	10 (0.1)	1 (0.1)
WSM	8189 (19.4)	7034 (33.2)	966 (7.2)	186 (2.7)	3 (0.4)
Missing (male)	19188 (45.4)	5429 (25.6)	8334 (62.4)	4864 (69.7)	561 (75.4)
Missing (female)	8072 (19.1)	5756 (27.2)	1632 (12.2)	679 (9.7)	5 (0.7)
STI reinfection, n (%)	4558 (10.8)	1418 (6.7)	2098 (15.7)	955 (13.7)	87 (11.7)
HIV coinfection, n (%)	2443 (5.8)	467 (2.2)	897 (6.7)	893 (12.8)	186 (25.0)
<i>Deprivation index‡</i>					
First quintile (least deprived)	10271 (24.3)	5185 (24.5)	3040 (22.8)	1757 (25.2)	289 (38.8)
Second quintile	7465 (17.7)	4328 (20.4)	2012 (15.1)	1037 (14.9)	88 (11.8)
Third quintile	4859 (11.5)	2763 (13.0)	1332 (10.0)	716 (10.3)	48 (6.5)
Fourth quintile	5703 (13.5)	3217 (15.2)	1578 (11.8)	827 (11.9)	81 (10.9)
Fifth quintile	7689 (18.2)	4319 (20.4)	2211 (16.6)	1079 (15.5)	80 (10.8)
Missing	6296 (14.9)	1390 (6.6)	3189 (23.9)	1559 (22.4)	158 (21.2)
<i>Health region of residence, n (%)</i>					
Barcelona	35215 (83.3)	17108 (80.7)	11566 (86.6)	5833 (83.6)	708 (95.2)
Other regions	7068 (16.7)	4094 (19.3)	1796 (13.4)	1142 (16.4)	36 (4.8)
<i>BHA setting</i>					
Rural	4193 (9.9)	2614 (12.3)	1039 (7.8)	516 (7.4)	24 (3.2)
Urban	29969 (70.9)	16347 (77.1)	8566 (64.1)	4516 (64.8)	540 (72.6)
Missing	8121 (19.2)	2241 (10.6)	3757 (28.1)	1943 (27.9)	180 (24.2)

\*Includes men who have sex with men, bisexual men and transgender men.

†Includes women who have sex with women, bisexual women and transgender women.

‡First quintile (31.52%), second quintile (40.09%), third quintile (46.27%), fourth quintile (53.98%) and fifth quintile (100%).

BHA, Basic Health Area; LGV, lymphogranuloma venereum; MSM, men who have sex with men; MSW, men who have sex with women; STI, sexually transmitted infection; WSM, women who have sex with men; WSW, women who have sex with women.

occurred in women (table 1). Among all STI cases, 78.6% were reported in individuals below 40 years of age. Chlamydia was reported most frequently among individuals below 30 years of age (66.1%), while syphilis occurred most in those above 30 years of age (77.1%). Among the 15 023 (35.5%) reported STI cases for which information regarding sexual preference was available, half (54.5%) were reported in women who have sex with men (WSM), 21.8% in MSM and 21.0% in MSW (table 1).

When examining the distribution of STI cases according to deprivation index, the highest proportion of cases was seen in less deprived areas, with 24.3% of all cases reported in the first quintile. In more deprived areas (fifth quintile), chlamydia (56%) and gonorrhoea (29%) occurred more frequently than syphilis (14%) and LGV (1%). Data around country of birth and education level were limited due to high rates of missing data (56.6% for country of birth and 76.5% for education level). Nevertheless,

**Table 2** Reported STI cases and incidence rates by year in Catalonia, Spain (2017–2019)

	2017		2018		2019		P trend
	Cases, n	Incidence rate per 100 000 population	Cases, n	Incidence rate per 100 000 population	Cases, n	Incidence rate per 100 000 population	
<b>Total</b>							
Total	9687	128.2	13 724	180.6	18 872	245.9	<0.001
<b>Sex</b>							
Female <sup>a</sup>	3362	87.4	5503	142.2	7811	200.0	<0.001
Male <sup>b</sup>	6325	170.5	8221	220.4	11 061	293.4	<0.001
<b>Age, years</b>							
<30 <sup>c</sup>	4607	198.4	7181	306.9	10 341	436.5	<0.001
≥30 <sup>d</sup>	5080	97.1	6543	124.4	8531	160.8	<0.001
<b>Sex and age</b>							
Male <30 years <sup>e</sup>	2411	202.9	3347	279.5	4668	384.0	<0.001
Male ≥30 years <sup>f</sup>	3914	155.2	4874	192.4	6393	250.3	<0.001
Female <30 years <sup>g</sup>	2196	193.6	3834	335.7	5673	491.9	<0.001
Female ≥30 years <sup>h</sup>	1166	43.0	1669	61.2	2138	77.7	<0.001
<b>STI type (total Catalonia)</b>							
Chlamydia	3562	47.1	7240	95.3	10 400	135.5	<0.001
Gonorrhoea	3492	46.2	4088	53.8	5782	75.3	<0.001
Syphilis	2430	32.2	2175	28.6	2370	30.9	0.1572
LGV	203	2.7	221	2.9	320	4.2	<0.001
<b>Health region of residence</b>							
Alt pirineu i Aran <sup>j</sup>	10	13.9	13	18.1	33	45.7	<0.001
Barcelona <sup>k</sup>	8205	165.0	11 475	229.5	15 535	307.8	<0.001
Camp de Tarragona <sup>l</sup>	294	49.1	491	81.3	870	142.2	<0.001
Catalunya Central <sup>m</sup>	358	69.4	484	93.1	583	110.7	<0.001
Girona <sup>n</sup>	564	65.7	893	103.2	1327	151.5	<0.001
Lleida <sup>o</sup>	212	58.9	247	68.5	404	111.5	<0.001
Terres de l'Ebre <sup>p</sup>	44	24.5	121	67.7	120	67.2	<0.001
<b>BHA setting<sup>q</sup></b>							
Rural	827	NA	1375	NA	1991	NA	NA
Urban	6475	NA	9868	NA	13 626	NA	NA
Missing	2385	NA	2481	NA	3255	NA	NA

Data source for denominators (a–q): the Statistical Institute of Catalonia (IDESCAT), online supplemental table S4. BHA, Basic Health Area; LGV, lymphogranuloma venereum; NA, not applicable; STI, sexually transmitted infection.

we report that among cases with available information, 72.4% were observed among individuals born in Spain and 85.0% in those with secondary or higher education (online supplemental table S5).

The incidence rate of STI cases was disproportionately higher in Barcelona (83.3%) compared with the other six regions combined (table 1). Barcelona reported the highest incidence rate of STIs, while Alt Pirineu i Aran recorded the lowest consistently throughout the study period (table 2). In 2019, the incidence rate of STIs was 307.8 cases per 100 000 population in Barcelona and 45.7 cases per 100 000 population in Alt Pirineu i Aran. Nevertheless, incidence rates of STIs increased significantly

from 2017 to 2019 in all regions, regardless of STI type. Similarly, the large majority of STI cases occurred in urban BHAs (70.9%) throughout the study period.

### Factors associated with HIV coinfection among individuals with STIs

In total, 6% of STI episodes affected HIV-positive individuals with a higher proportion of HIV coinfection observed with cases of syphilis and LGV (13% and 25%, respectively) and the lowest with cases of chlamydia (2%) (table 1). Factors associated with HIV coinfection among individuals with STIs in the multivariable analyses are shown in table 3. The likelihood of HIV coinfection was

**Table 3** Factors associated with HIV coinfection among individuals diagnosed with STIs in Catalonia, Spain (2017–2019)

Characteristic	Total, n (n=34 600)	HIV-positive, n (n=1376)	OR	95% CI	aOR	95% CI
<b>Sex</b>						
Female	14 938	29	1 (ref)		1 (ref)	
Male	19 662	1347	37.81	26.69 to 55.93	23.69	16.67 to 35.13
<b>Age group, years</b>						
<20	3696	5	1 (ref)		1 (ref)	
20–29	14 826	328	16.70	7.70 to 46.83	8.33	3.82 to 23.40
30–39	8704	595	54.17	25.05 to 151.57	18.58	8.56 to 52.13
40–49	4759	339	56.62	26.09 to 158.78	17.66	8.10 to 49.65
50–59	1748	89	39.60	17.80 to 112.58	13.06	5.84 to 37.24
>60	867	20	17.43	7.04 to 52.50	6.98	2.80 to 21.09
<b>Deprivation index*</b>						
First quintile (least deprived)	7679	501	1 (ref)		1 (ref)	
Second quintile	6098	210	0.51	0.43 to 0.60	0.70	0.59 to 0.83
Third quintile	4163	109	0.38	0.31 to 0.47	0.63	0.50 to 0.78
Fourth quintile	4663	186	0.60	0.50 to 0.71	0.83	0.69 to 1.00
Fifth quintile	6347	175	0.41	0.34 to 0.48	0.60	0.50 to 0.72
Missing	5650	195	0.51	0.43 to 0.60	0.51	0.39 to 0.67
<b>STI episodes (total), n</b>						
1	29 104	791	1 (ref)		1 (ref)	
2–4	5304	529	3.96	3.54 to 4.44	2.69	2.39 to 3.03
5–7	192	56	14.74	10.64 to 20.16	5.96	4.26 to 8.24
<b>BHA setting</b>						
Rural	3699	81	1 (ref)		1 (ref)	
Urban	23 812	1023	2	1.61 to 2.54	1.32	1.04 to 1.69

\*First quintile (31.52%), second quintile (40.09%), third quintile (46.27%), fourth quintile (53.98%) and fifth quintile (100%).  
aOR, adjusted OR; BHA, Basic Health Area; CI, Confidence Interval; OR, Odds ratio; STI, sexually transmitted infection.;

greater among males (adjusted OR (aOR) 23.69; 95% CI 16.67 to 35.13 compared with females) and in urban BHAs (aOR 1.32; 95% CI 1.04 to 1.69 compared with rural BHAs). All age groups from 20 years and above and having multiple STI episodes were also associated with greater odds of HIV coinfection. BHA deprivation indices beyond the first quintile were associated with lower likelihood of HIV coinfection among individuals with STIs.

#### Identification and characterisation of the socio-epidemiological clusters of STIs

Of the 373 Catalan BHAs, five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella, Roquetes-Canyelles and Viladecans 3) were excluded from the K-means clustering analysis because their delimitations and populations changed during the study period. In these five BHAs, 679 episodes were reported during the 3 years of the study period. This fact and having 5773 episodes with no information available about BHA of residence reduced the sample size for the cluster analysis from 42 283 to 35 831 STI cases. Of the 368 BHAs included in the analysis, we identified three distinct

clusters (table 4). Among the included BHAs, the incidence rate of STIs in 2017–2019 was 160.6 per 100 000 population per year.

Of the three clusters, the socio-epidemiological characteristics of STI-infected individuals in cluster A most closely resembled that of the total cases reported in the Catalan surveillance system that were included in the cluster analysis. Among the 109 BHAs in cluster A, median age was 31 years compared with 29 years among all reported STI cases, median deprivation index was 31.9 versus 39.8, the proportion of men was 67.4% versus 58.1% and HIV coinfection rate was 8.8% versus 6.1% (table 4).

Cluster B consisted of the largest number of BHAs (251) and had the highest deprivation index (44.9) of all three clusters and compared with the total. The incidence rate of STIs was lower in cluster B compared with that of the total reported cases included in the cluster analysis (136.3 vs 160.6 per 100 000 population), but represented the majority of all reported STI cases (55.7%). STI-infected individuals in cluster

**Table 4** Characteristics of socio-epidemiological STI clusters in Catalonia, Spain (2017–2019)

Characteristic	Cluster A	Cluster B	Cluster C	Total*
<i>Demographics</i>				
BHAs, n (%)	109 (29.6)	251 (68.2)	8 (2.2)	368 (100)
Median age, median years (IQR)	31 (18–60)	26 (17–58)	34 (20–58)	29 (17–59)
Median deprivation index (IQR)	31.9 (3.0–58.2)	44.9 (19.2–76.9)	25.6 (10.7–63.6)	39.8 (10.7–72.3)
Annual STI incidence rate (per 100 000 population)	162.0	136.3	721.0	160.6
<i>Reported STI cases, n (%)</i>				
Total	11 527 (32.2)	19 945 (55.7)	4 359 (12.2)	35 831 (100)
<i>Sex</i>				
Female	3 758 (32.6)	10 566 (53.0)	686 (15.7)	15 010 (41.9)
Male	7 769 (67.4)	9 379 (47.0)	3 673 (84.3)	20 821 (58.1)
<i>Country of birth</i>				
Spain	3 920 (34.0)	6 910 (34.7)	1 325 (30.4)	12 155 (33.9)
Outside Spain	1 279 (11.1)	2 819 (14.1)	435 (10.0)	4 533 (12.7)
Missing	6 328 (54.9)	10 216 (51.2)	2 599 (59.6)	19 143 (53.4)
<i>Sexual preference</i>				
MSM†	1 234 (10.7)	1 104 (5.5)	655 (15.0)	2 993 (8.4)
MSW	751 (6.5)	2 164 (10.9)	64 (1.5)	2 979 (8.3)
WSM‡	1 440 (12.5)	6 066 (30.4)	130 (3.0)	7 636 (21.3)
WSW	75 (0.7)	295 (1.5)	6 (0.1)	376 (1.1)
Missing (male)	5 784 (50.2)	6 111 (30.6)	2 954 (67.8)	14 849 (41.4)
Missing (female)	2 243 (19.5)	4 205 (21.1)	550 (12.6)	6 998 (19.5)
<i>STI type</i>				
Gonorrhoea	3 448 (29.9)	5 240 (26.3)	1 448 (33.2)	10 136 (28.3)
Chlamydia	5 739 (49.8)	12 314 (61.7)	1 649 (37.8)	19 702 (55.0)
Syphilis	2 117 (18.4)	2 263 (11.4)	1 027 (23.6)	5 407 (15.1)
LGV	223 (1.9)	128 (0.6)	235 (5.4)	586 (1.6)
Multiple (>1) STI episodes	1 600 (13.9)	1 524 (7.6)	1 048 (24.0)	4 172 (11.6)
HIV coinfection	1 011 (8.8)	495 (2.5)	686 (15.7)	2 192 (6.1)
<i>STI incidence rate categories of BHAs</i>				
First quintile (2.4 per 1000)	710 (6.2)	1 820 (9.1)	0	2 530 (7.1)
Second quintile (3.6 per 1000)	2 136 (18.5)	3 359 (16.8)	0	5 495 (15.3)
Third quintile (5.2 per 1000)	1 377 (12.0)	6 588 (33.0)	0	7 965 (22.2)
Fourth quintile (9.8 per 1000)	5 688 (49.4)	7 508 (37.6)	0	13 196 (36.8)
Fifth quintile (42.8 per 1000)	1 616 (14.0)	670 (3.4)	4 359 (100)	6 645 (18.6)
<i>BHA setting</i>				
Rural	797 (6.9)	3 158 (15.8)	0	3 955 (11.0)
Urban	9 461 (82.1)	15 787 (79.2)	4 359 (100)	29 607 (82.6)
Missing	1 269 (11.0)	1 000 (5.0)	0	2 269 (6.3)

\*Of the 373 Catalan BHAs, five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella, Roquetes-Canyelles and Viladecans 3) were excluded from the K-means clustering analysis because their delimitations and populations changed during the study period.

†Includes men who have sex with men, bisexual men and transgender men.

‡Includes women who have sex with women, bisexual women and transgender women.

BHA, Basic Health Area; IQR, Interquartile range; LGV, lymphogranuloma venereum; MSM, men who have sex with men; MSW, men who have sex with women; STI, sexually transmitted infection; WSM, women who have sex with men; WSW, women who have sex with women.

B were the youngest among all groups (26 years) and were predominantly women (53.0%), whereas men represented the majority in all other groups. Compared with the total, cluster B consisted of more rural BHAs (15.8% vs 11.0%) and had a higher proportion of heterosexual men and women (approximately

12% higher) and chlamydia cases (61.7% vs 55.0%). Rates of multiple STI episodes and HIV coinfection in cluster B were the lowest of all three clusters and compared with the total (table 4).

Cluster C consisted of only eight BHAs and had the lowest deprivation index (25.6) among all groups (table 4). The

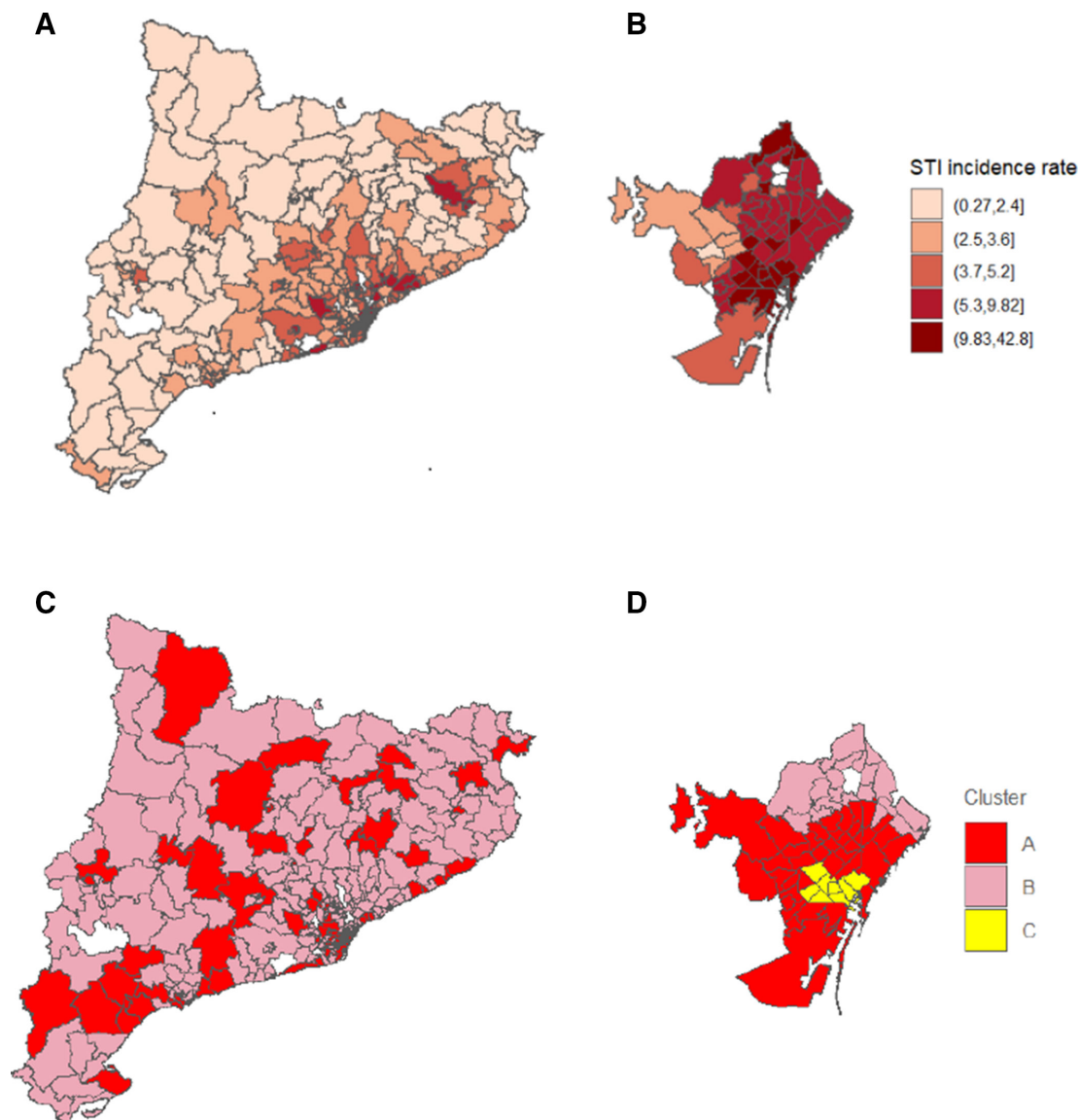
incidence rate of STIs in cluster C was the highest among all groups (721.0 per 100000 population), with all cases reported in urban BHAs. STI-infected individuals in cluster C were the oldest among all groups (34 years) and had the highest proportion of MSM. Similar to other clusters and the total reported cases, chlamydia remained the most common STI type; however, cluster C was characterised by higher rates of gonorrhoea (33.2%), syphilis (23.6%), LGV (5.4%), multiple STI episodes (24.0%) and HIV coinfection (15.7%).

Almost 60% of STI cases in cluster B occurred in BHAs in the three lowest quintiles of STI incidence rates, while more than 60% in cluster A occurred in areas of high STI incidence

rates (fourth and fifth quintiles). All 4359 STI cases in cluster C were reported in BHAs in the highest quintile of STI incidence rate. This correlated well with the fact the number of STI cases per BHA was higher in clusters A and C (105.8 and 544.9 cases per BHA, respectively) than in the total (97.4 cases per BHA), which indicates higher proportion of high incidence rates (table 4 and figure 1).

## DISCUSSION

Our findings revealed that the incidence of STIs in Catalonia almost doubled from 2017 to 2019, primarily driven



**Figure 1** Incidence rates (per 1000 population) and socio-epidemiological clusters of STIs by BHA during 2017–2019. (A) STI incidence rates in Catalonia; (B) STI incidence rates in Barcelona city\*; (C) STI socio-epidemiological clusters in Catalonia and (D) STI socio-epidemiological clusters in Barcelona city\*. \*Health Regions were used as a bigger unit of analysis than BHA. The municipality of Barcelona is shown to enhance the visualisation of cluster C. From a total of 373 Catalan BHA, five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella, Roquetes-Canyelles and Viladecans 3) were excluded from the K-means clustering analysis because their delimitations and populations changed during the study period. BHA, Basic Health Area; STIs, sexually transmitted infections.



by the increase in cases among young adults (under 30 years) and in cases of chlamydia (particularly in women) and gonorrhoea. In 2017–2019, the majority of STI cases occurred also in individuals below the age of 30 years, and those living in urban and less deprived areas, with most cases reported in Barcelona. The identification and characterisation of socio-epidemiological clusters of STI showed that young women living in rural and more deprived areas were more likely to be affected by chlamydia. Furthermore, MSM living in urban and less deprived areas showed, more frequently than other population groups, higher STI incidence rates, more multiple STI episodes and higher percentages of HIV coinfection. Similarly, the factors associated with HIV coinfection were being men, older than 20 years old, living in urban and less deprived areas and having multiple STI episodes.

After a long period of continuous reduction of STI incidence in Western countries, which coincided with the beginning and hardest times of the HIV epidemic from the 1980s to 2010s, many countries including the USA and European countries are recently reporting an ongoing re-emergence of STIs.<sup>3–6</sup> The rise in STI cases has been partially attributed to enhancement of surveillance systems and the introduction of improved diagnostic tools in recent years.<sup>10</sup> Other contributing factors, described mostly among MSM, include the use of HIV pre-exposure prophylaxis, the use of recreational drugs for sex, substance and/or alcohol abuse and widespread use of the internet and other technologies to seek sexual partners.<sup>20–22</sup>

Chlamydia has been reported more frequently in WSM, while syphilis, gonorrhoea and LGV were more common in MSW and MSM.<sup>3–9</sup> Similarly, in our study, we found different epidemiological characteristics for each STI type. Chlamydia was more common in women, mostly in WSM, with a large majority occurring in individuals below the age of 30 years. Gonorrhoea, syphilis and LGV were substantially more frequent in men, specifically among MSM, and showed higher percentages of reinfections and HIV coinfections than chlamydia. Most STI cases were observed in Spanish-born individuals and among those with secondary or higher education levels, although these findings should be interpreted with caution because of the high proportion of missing data for sexual preference and education level. Our findings are consistent with earlier studies and reports of STIs in Catalonia in 2007–2015,<sup>12</sup> 2012–2017<sup>23</sup> and 2018–2019,<sup>9</sup> showing a proportionally higher increase in young adults, mostly women, especially for chlamydia but also for gonorrhoea. Our findings are also consistent with that of a previous study among residents of Barcelona showing that STIs are becoming more prevalent in individuals with favourable socioeconomic status and education levels.<sup>24</sup>

Consistent with previous data,<sup>12</sup> we found that male sex, age above 20 years (particularly 30–60 years), living in urban or less deprived areas, and having multiple STI episodes were associated with an increased risk of HIV coinfection. STIs and HIV have been described as

synergic infections and should be viewed as a syndemic.<sup>25</sup>

The WHO and other public health agencies have emphasised the importance of integrating surveillance of STIs, HIV and even viral hepatitis, and strengthening understanding of determinants of these infections by linking biological and behavioural surveillance, to enhance the identification and characterisation of populations at increased risk of infection.<sup>10 25</sup> Sociodemographic and socioeconomic are increasingly being established as more important risk factors of STI acquisition than individual behaviours, particularly among women from disadvantaged groups.<sup>26 27</sup>

The k-means clustering methodology is a machine learning approach that has proven its utility and potential in classifying and grouping health-related outcomes. It has been used in the field of bipolar disorder to define cluster-based disease severity using heterogeneous variables such as sociodemographic, clinical, cognitive, vital signs and laboratory parameters.<sup>28</sup> More recently, its potential to monitor and group SARS-CoV-2 prevalence by magnitude and trends (higher, medium and lower) at a regional level in Italy has been described.<sup>29</sup> Identification of these ‘clusters of characteristics’ may be useful, in their specific context, to better detect and characterise case profiles by site or geographical area, which could ultimately lead to better-designed interventions to improve health outcomes. In a recent study of STI risk among MSMs, hierarchical cluster analysis, another machine learning methodology, identified factors other than behaviour, such as sexual networks and risk perception, that influence the vulnerability to STIs and HIV infections.<sup>30</sup> To the best of our knowledge, this current study is the first to apply the k-means clustering methodology to identify and characterise socio-epidemiological clusters of STI.

A key limitation of this study is the high proportion of missing data around sociodemographic and lifestyle characteristics, a common phenomenon in population-based epidemiological studies where questionnaires are used. This may have potentially introduced information bias or inaccurate representation of the true situation when describing high-risk populations. Although not formally assessed, we classify these missing data as missing completely at random due to time constraints in completion of the epidemiological questionnaires by surveillance officers and healthcare professionals who notified the diseases to the surveillance systems. Nonetheless, our findings are similar to those reported in previous analyses.<sup>3–6 9 12 24</sup> The age category above 60 years old may contribute to residual confounding although the risk is minimal because it is the age group with the smallest sample size and the range is larger than for other age categories. Categorisation of the deprivation indices by quintiles could have diluted the findings if deprivation was a strong confounder or unevenly distributed, although we do not believe either event to be the case in our analysis.

A strength of our study is the inclusion of ecological variables of socioeconomic status which are highly

relevant and pertinent for describing groups at increased risk of STIs. We believe that the most valuable outcome of our study is that it shows the utility of complementing traditional epidemiological analyses with new methodologies, in this case, a machine learning approach, to combine heterogeneous data sources. This would allow identification and characterisation of target populations at increased risk of STIs to design more efficient measures to prevent and control STIs and HIV infection at a small health area level.

In conclusion, consistent with other European countries, our study found that STIs increased at an alarming rate during 2017 to 2019 in Catalonia, Spain, and continues to be a worrisome public health concern. The STI epidemic is both an issue of the health sector, and it also poses a threat to the broader global development framework and agenda. While declines in HIV infection has been observed in the last decade in Catalonia, as in many other regions in Europe, primarily due to the success of wider and earlier use of antiretroviral therapies, STI rates have been increasing dramatically, both among the MSM population, and also in heterosexual women and young adults. We found that young women living in rural and deprived areas were more likely to be affected by chlamydia, while MSM living in urban and less deprived areas had higher overall STI incidence rates, multiple STI episodes and greater likelihood of HIV coinfection. Preventative strategies must consider these populations priority targets and take into account structural social determinants identified as crucial in our analysis. Our findings suggest that monitoring the STI epidemic in accordance with determinants of health and designing intervention programmes targeted at the local context would be of paramount importance rather than using national or regional prevalence as the key monitoring variable.

#### Author affiliations

<sup>1</sup>Centre of Epidemiological Studies of Sexually Transmitted Disease and AIDS in Catalonia (CEEISCAT), Department of Health, Generalitat of Catalonia, Badalona, Spain

<sup>2</sup>Pompeu Fabra University (UPF), Barcelona, Spain

<sup>3</sup>Epidemiology Department, Epicconcept, Paris, France

<sup>4</sup>Fundació Institut d'Investigació Germans Trias i Pujol (IGTP), Badalona, Spain

<sup>5</sup>Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain

<sup>6</sup>Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>7</sup>Epidemiology Service, Public Health Agency of Barcelona, Barcelona, Spain

<sup>8</sup>Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona, Agency of Public Health of Catalonia, Generalitat of Catalonia, Tarragona, Spain

**Acknowledgements** We thank all healthcare professionals working in STI/HIV surveillance, prevention and control in Catalonia who enable case detection, diagnosis and treatment, as well as the notification and information gathering for the epidemiological questionnaires. We thank Stefanie Chuah for her valuable support in editing the manuscript.

**Collaborators** The Catalan HIV and STI Surveillance Group: A Sentís, E López, V Gonzalez, R Lugo, MP Bonamusa, J Reyes, J Casabona (Centre d'Estudis Epidemiològics sobre les Infeccions de Transmissió Sexual i Sida de Catalunya); P Garcia de Olalla, Lilas Mercurial, E Masdeu, M Ros, C Rius (Servei d'Epidemiologia

de l'Agència de Salut Pública de Barcelona); M Company, M Danés, N Camps (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública a Girona); RM Vileu, G Ferrús, N Borrell, S Minguell (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública a Tarragona); J Ferràs (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública a Terres de l'Ebre); I Parrón (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública al Barcelonès Nord i Maresme); I Mòdol, A Martínez, P Godoy (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública a Lleida); MA Tarrès, J Pérez, M Boldú, I Barrabeig (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública a Barcelona Sud); E Donate, L Clotet, MR Sala (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública al Vallès Occidental i Vallès Oriental); M Carol, V Guadalupe-Fernández (Servei de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública a Catalunya Central) and J Mendiros, P Ciruela, G Carmona, R Mansilla, JL Martínez, S Hernández (Subdirecció General de Vigilància Epidemiològica i Resposta a Emergències de Salut Pública, Agència de Salut Pública de Catalunya).

**Contributors** AS conceptualised and designed the study. MM-F cleaned the database, MM-F, LE-C and YD performed the statistical and cluster analysis. AS, EL-C and DKN reviewed scientific literature. AS, JMR-U and JC drafted the manuscript and AS, EL-C, PGdO, LM, NB, JMR-U and JC interpreted the results. All authors critically reviewed the manuscript and approved the final version to be published. AS is the author acting as guarantor.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Map disclaimer** The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Alexis Sentís <http://orcid.org/0000-0002-8132-5645>

Juliana Reyes-Urueña <http://orcid.org/0000-0002-3122-6518>

#### REFERENCES

- Rowley J, Vander Hoorn S, Korenromp E, *et al*. Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. *Bull World Health Organ* 2019;97:548–62.
- WHO Department of Reproductive Health and Research. Report on global sexually transmitted infection surveillance, 2018. Available: <https://apps.who.int/iris/bitstream/handle/10665/277258/9789241565691-eng.pdf?ua=1> [Accessed 8 Jan 2021].
- European Centre for Disease Prevention and Control. Chlamydia infection. annual epidemiological report for 2018. Stockholm: ECDC, 2020. Available: <https://www.ecdc.europa.eu/sites/default/files/documents/AER-for-2018-STI-chlamydia.pdf> [Accessed 7 Aug 2021].

- 4 European Centre for Disease Prevention and Control. Gonorrhoea. annual epidemiological report for 2018. Stockholm: ECDC, 2020. Available: <https://www.ecdc.europa.eu/sites/default/files/documents/gonorrhoea-annual-epidemiological-report-2018.pdf> [Accessed 7 Aug 2021].
- 5 European Centre for Disease Prevention and Control. Syphilis. annual epidemiological report for 2018. Stockholm: ECDC, 2020. Available: <https://www.ecdc.europa.eu/sites/default/files/documents/syphilis-aer-2018.pdf> [Accessed 7 Aug 2021].
- 6 European Centre for Disease Prevention and Control. Lymphogranuloma venereum. annual epidemiological report for 2018. Stockholm: ECDC, 2020. Available: <https://www.ecdc.europa.eu/sites/default/files/documents/lymphogranuloma-venereum-aer-2018.pdf> [Accessed 7 Aug 2021].
- 7 Àrea de vigilància del VIH y conductas de riesgo. Vigilancia epidemiològica de las infecciones de transmisión sexual, 2017. Madrid: Centro Nacional de Epidemiología, Instituto de Salud Carlos III/Plan Nacional sobre el Sida - Dirección General de Salud Pública, 2019. Available: [https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/Vigilancia ITS\\_1995\\_2017\\_def.pdf](https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/Vigilancia ITS_1995_2017_def.pdf) [Accessed 27 Mar 2021].
- 8 Àrea de vigilància del VIH y conductas de riesgo. Vigilancia epidemiològica de las infecciones de transmisión sexual, 1995-2010. Madrid: Centro Nacional de Epidemiología, Instituto de Salud Carlos III/Plan Nacional sobre el Sida - Dirección General de Salud Pública, 2012. Available: [https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/VigilanciaITS1995\\_2010.pdf](https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/VigilanciaITS1995_2010.pdf) [Accessed 27 Mar 2021].
- 9 Centre d'Estudis Epidemiològics sobre les Infeccions de Transmissió Sexual i Sida de Catalunya (CEEISCAT). Vigilància epidemiològica de les Infeccions de Transmissió Sexual (ITS) a Catalunya. Informe anual 2019. Badalona, Spain, 2020. Available: [https://canalsalut.gencat.cat/web/.content/\\_A-Z/S/sida/enllasos/anual ITS.pdf](https://canalsalut.gencat.cat/web/.content/_A-Z/S/sida/enllasos/anual ITS.pdf) [Accessed 27 Mar 2021].
- 10 Taylor MM, Wi TE. Transforming and integrating STI surveillance to enhance global advocacy and investment in STI control. *J Int AIDS Soc* 2019;22 Suppl 6:25361.
- 11 Centers for Disease Control and Prevention. HIV/AIDS & STDs. Available: <https://www.cdc.gov/std/hiv/default.htm> [Accessed 22 Feb 2021].
- 12 Sentís A, Martin-Sanchez M, Arando M, *et al*. Sexually transmitted infections in young people and factors associated with HIV coinfection: an observational study in a large City. *BMJ Open* 2019;9:e027245.
- 13 Le Polain De Waroux O, Harris RJ, Hughes G, *et al*. The epidemiology of gonorrhoea in London: a Bayesian spatial modelling approach. *Epidemiol Infect* 2014;142:211-20.
- 14 Gesink D, Wang S, Norwood T, *et al*. Spatial epidemiology of the syphilis epidemic in Toronto, Canada. *Sex Transm Dis* 2014;41:637-48.
- 15 Jain AK. Data clustering: 50 years beyond k-means. *Pattern Recognit Lett* 2010;31:651-66.
- 16 Centre d'Estudis Epidemiològics sobre les Infeccions de Transmissió Sexual i Sida de Catalunya (CEEISCAT). SIVES 2015 - Sistema Integrat de Vigilància Epidemiològica de la SIDA/VIH/ITS a Catalunya. Barcelona, Spain, 2015. Available: [https://scientiasalut.gencat.cat/bitstream/handle/11351/3418/informe\\_SIVES\\_2015\\_informe\\_epidemiologic\\_CEEISCAT\\_2015.pdf.pdf?sequence=1&isAllowed=y](https://scientiasalut.gencat.cat/bitstream/handle/11351/3418/informe_SIVES_2015_informe_epidemiologic_CEEISCAT_2015.pdf.pdf?sequence=1&isAllowed=y) [Accessed 27 Mar 2021].
- 17 Notifiable diseases and epidemic outbreaks. Department of health, public health agency of Catalonia, Generalitat de Catalunya, 2016. Available: <https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/malalties-de-declaracio-obligatoria-i-brots-epidemics/> [Accessed 20 Jun 2021].
- 18 European Centre for Disease Prevention and Control. Eu case definitions, 2018. Available: <https://www.ecdc.europa.eu/en/surveillance-and-disease-data/eu-case-definitions> [Accessed 7 Aug 2020].
- 19 Agency for Health Quality and Assessment of Catalonia. Nou indicador socioeconòmic per al finançament de les ABS. Observatori del Sistema de Salut de Catalunya, 2017. Available: [http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador\\_socioeconomic\\_2015/](http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador_socioeconomic_2015/) [Accessed 6 Aug 2020].
- 20 Folch C, Esteve A, Zaragoza K, *et al*. Correlates of intensive alcohol and drug use in men who have sex with men in Catalonia, Spain. *Eur J Public Health* 2010;20:139-45.
- 21 Saberi P, Neilands TB, Lally MA, *et al*. The association between use of online social networks to find sex partners and sexually transmitted infection diagnosis among young men who have sex with men and transgender women living with HIV. *J Int Assoc Provid AIDS Care* 2019;18:232595821986732-11.
- 22 Jansen K, Steffen G, Potthoff A, *et al*. STI in times of PrEP: high prevalence of Chlamydia, gonorrhoea, and Mycoplasma at different anatomic sites in men who have sex with men in Germany. *BMC Infect Dis* 2020;20:110.
- 23 Vives N, Garcia de Olalla P, González V, *et al*. Recent trends in sexually transmitted infections among adolescents, Catalonia, Spain, 2012-2017. *Int J STD AIDS* 2020;31:1047-54.
- 24 Martí-Pastor M, García de Olalla P, Barberá M-J, *et al*. Epidemiology of infections by HIV, syphilis, gonorrhoea and lymphogranuloma venereum in Barcelona City: a population-based incidence study. *BMC Public Health* 2015;15:1015.
- 25 Murti M, Wong J, Whelan M, *et al*. The need for integrated public health surveillance to address sexually transmitted and blood-borne syndemics. *Can Commun Dis Rep* 2019;45:63-6.
- 26 Hill AV, De Genna NM, Perez-Patron MJ, *et al*. Identifying Syndemics for sexually transmitted infections among young adults in the United States: a latent class analysis. *J Adolesc Health* 2019;64:319-26.
- 27 Jagadeesh S, Combe M, Couppié P. Mapping priority neighborhoods: a novel approach to cluster identification in HIV/AIDS population. *Res Sq* 2020.
- 28 Fuente-Tomas Ldela, Arranz B, Safont G, *et al*. Classification of patients with bipolar disorder using k-means clustering. *PLoS One* 2019;14:e0210314.
- 29 Maugeri A, Barchitta M, Agodi A. A clustering approach to classify Italian regions and provinces based on prevalence and trend of sars-cov-2 cases. *Int J Environ Res Public Health* 2020;17:5286-14.
- 30 Blondeel K, Dias S, Furegato M, *et al*. Sexual behaviour patterns and STI risk: results of a cluster analysis among men who have sex with men in Portugal. *BMJ Open* 2021;11:33290.